



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

**A STUDY TO DEVELOP CLINICAL DECISION RULES FOR
THE EMERGENCY DEPARTMENT USE OF RADIOGRAPHY
IN ACUTE ANKLE INJURIES**

by

IAN GILMOUR STIELL

**Thesis submitted to
the School of Graduate Studies and Research
in partial fulfilment of the requirements for the
M.Sc. degree in Epidemiology**

UNIVERSITY OF OTTAWA



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-70501-9

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

The Problem

The majority of ankle injury patients presenting to an emergency department undergo a radiographic examination even though the prevalence of fracture in such patients is less than 15%. This inefficient use of x-rays adds unnecessarily to the cost of health care. Emergency physicians are in need of reliable and sensitive clinical decision rules to help them be more selective in their use of radiography.

Background Information

Methodological standards for the development and validation of clinical prediction or decision rules have been well described. Previous studies to assess clinical predictors of the need for radiography in acute ankle injuries have been flawed by a number of methodological weaknesses. Consequently there are no widely accepted guidelines to assist clinicians.

Goals and Objectives

This study was performed in two phases. The goal of **Phase 0** was to determine the current use of radiography in ankle injury patients in two large university hospitals. The objectives were to study the clinical and demographic features of these patients and to document the referral fraction and yield of radiography.

The goal of **Phase I** was to develop clinical decision rules that are sensitive for detecting significant fractures in adult ankle injury patients. The objectives were to perform standardized clinical assessments on ankle injury patients, test the interobserver reliability of clinical predictors, use multivariate techniques to develop decision making models, and to assess the classification performance of these decision rules.

Methods

Phase 0 involved a retrospective chart audit of five months' use of radiography for adult ankle injury patients seen at two university hospital emergency departments.

Phase I involved the assessment of thirty-two standardized clinical findings by emergency staff physicians on 750 patients, prior to x-ray. A pilot stage was used to refine data collection techniques. One hundred patients were assessed by two physicians to allow the calculation of the kappa coefficient of interobserver agreement. Those variables found to be most reliable and associated with significant fractures by univariate analysis were then analyzed by logistic regression and recursive partitioning techniques to develop the decision rules. Performance of the rules was assessed by means of classification tables and receiver operating characteristic curves.

Results

Phase 0 data revealed that 1,290 ankle injury patients presented with acute ankle injuries during the five month study period. 94.3% were referred for at least one x-ray series and the overall yield for significant fracture was 13.2%.

In **Phase I**, 155 patients were enrolled in a pilot stage and another 750 in the main study. The variables with the highest interobserver agreement and their kappa values were: ability to bear weight (0.83), bone tenderness at the base of the fifth metatarsal (0.78), at the posterior edge of lateral malleolus (0.75), at tip of the medial malleolus (0.66), at the posterior edge or tip of lateral malleolus combined (0.76), and at the posterior edge or tip of medial malleolus combined (0.66).

All 70 significant fractures found in the 689 ankle x-ray series performed were identified among people who had pain near the malleoli and who either: 1) were age 55 or greater, or 2) had localized bone tenderness of the posterior edge or tip of either malleolus, or 3) were unable to bear weight both immediately after the injury and in the emergency department. Combining these four predictors yielded a decision rule that was 100% sensitive and 40.1% specific for detecting ankle fractures, and would allow a reduction of 36.0% of ankle x-ray series ordered. Likewise, all 32 significant fractures

of the midfoot on the 230 foot x-ray series performed were found among patients with pain in the midfoot and bone tenderness at either the base of the fifth metatarsal, the cuboid, or the navicular.

Conclusion

This study confirmed that the majority of ankle injury patients are currently being referred for radiography but that the yield of the x-rays for fracture is relatively low. Highly sensitive decision rules for the use of ankle and foot x-ray series were developed and should now be validated and tested in the field. These rules may permit clinicians to confidently reduce the number of x-rays ordered in adult ankle injury patients and thereby reduce unnecessary costs to the health care system.

ACKNOWLEDGEMENTS

This thesis would not have been possible without a great deal of support at home. My wife Kathy shouldered extra burdens and responsibilities and showed wonderful understanding during difficult times. My sons, Andrew and Michael, eased the process with their patience and cooperation. Thank-you.

I am grateful to my thesis supervisors, Dr. Ian McDowell and Dr. Rama Nair, who kindly provided their time and expertise to oversee this project. Their insights and guidance were invaluable.

My clinical co-investigators, Dr. Gary Greenberg and Dr. Doug McKnight, provided essential collaboration and common sense throughout. The study would have been impossible without the good-humoured cooperation of the staff emergency physicians at the Ottawa Civic and Ottawa General Hospitals. Data collection was largely carried out by Dr. Hany Aeta, Teresa Cacciotti RN, and Pamela Sheehan RN. Irene Harris expertly assisted with documents.

Thank-you to Jack Boulet for computer software advice and to the students and staff of the Department of Epidemiology for feedback and comments. Dr. Nick Birkett offered advice and a fresh perspective on the final draft of this thesis.

The Ministry of Health of Ontario provided grant support (#04090N) for the study and the Ottawa Civic Hospital and the Emergency Department provided protected time.

TABLE OF CONTENTS

- ABSTRACT i
- ACKNOWLEDGEMENTS iv
- 1. INTRODUCTION 1
 - 1.1 Statement of the Problem 1
 - 1.2 Contribution of the Study 2
 - 1.3 Organization of the Study 2
- 2. REVIEW OF LITERATURE 3
 - 2.1 Methodological Standards 3
 - 2.1.1 Definition of Outcome 3
 - 2.1.2 Definition of Predictor Variables 4
 - 2.1.3 Reliability of Predictor Variables 4
 - 2.1.4 Selection of Subjects 4
 - 2.1.5 Mathematical Techniques 4
 - 2.1.6 Sensibility 5
 - 2.1.7 Accuracy 5
 - 2.1.8 Prospective Validation 6
 - 2.1.9 Effect on Patient care 6
 - 2.2 Previous Studies 7
 - 2.2.1 Methodological Weaknesses 7
 - 2.2.2 Specific Studies 8
 - 2.3 Relevant Clinical Variables 10
- 3. GOALS AND OBJECTIVES 12
 - 3.1 Phase 0: Baseline Assessment 12
 - 3.1.1 Goal 12
 - 3.1.2 Objectives 12
 - 3.2 Phase I: Development of the Rules 13
 - 3.2.1 Goal 13
 - 3.2.2 Objectives 13
- 4. METHODS - PHASE 0: BASELINE ASSESSMENT 14
 - 4.1 Patient Population 14
 - 4.1.1 Inclusion Criteria 14
 - 4.1.2 Exclusion Criteria 14
 - 4.2 Data Collection 16
 - 4.3 Data Analysis 16
- 5. METHODS - PHASE I: DEVELOPMENT OF THE RULE 18
 - 5.1 Patient Population 18
 - 5.1.1 Inclusion and Exclusion Criteria 18

5.1.2 Sample Size	18
5.1.3 Patient Selection	21
5.1.4 Ethics	21
5.2 Standardized Patient Assessment	22
5.2.1 Selection of Variables	22
5.2.2 Pilot Stage	22
5.2.3 Variables from History	23
5.2.4 Variables from Physical Examination	23
5.2.5 Patient Assessment	24
5.2.6 Physicians' Predicted Probability	24
5.3 Assessment of Interobserver Reliability	25
5.4 Outcome Measures	25
5.5 Data Analysis	26
5.5.1 Interobserver Agreement	26
5.5.2 Univariate Analysis	27
5.5.3 Multivariate Techniques	27
5.5.4 Logistic Regression	27
5.5.5 Recursive Partitioning	29
6. RESULTS - PHASE 0: BASELINE ASSESSMENT	31
6.1 Characteristics of Patients	31
6.2 Use of Radiography	32
7. RESULTS - PHASE I: DEVELOPMENT OF THE RULE	38
7.1 Descriptive Statistics	38
7.2 Physicians' Predictions and Attitudes	41
7.3 Interobserver Agreement	46
7.4 Derivation of Ankle X-ray Series Rule	51
7.4.1 Univariate Analysis	51
7.4.2 Multivariate Analysis	59
7.4.3 Logistic Regression Analysis	60
7.4.4 Recursive Partitioning Analysis	61
7.5 Classification Performance	65
7.6 Derivation of Foot X-ray Series Rule	68
7.6.1 Derivation	68
7.6.2 Classification Performance	72
8. DISCUSSION	74
8.1 Phase 0: Baseline Use of Radiography	74
8.2 Phase I: Development of the Rules	75
8.2.1 Interobserver Agreement	75
8.2.2 Clinical Decision Rules	76
8.2.3 Methodological Issues	78
8.2.4 Limitations of the Study	79
8.3 Future Research	81
8.3.1 Phase II: Prospective Validation	81

8.3.2 Phase III: Field Trial	82
8.4 Conclusion	83
REFERENCES	85
LIST OF APPENDICES	i

LIST OF TABLES

Table 1. Clinical predictor variables based on history as assessed in previous studies	11
Table 2. Clinical predictor variables based on physical examination as assessed in previous studies	11
Table 3. Sample size required to estimate various levels of sensitivity and specificity with 95% confidence \pm 0.05 for zone I injuries	19
Table 4. 95% confidence intervals for various levels of sensitivity and specificity for zone II injuries	20
Table 5. 95% confidence intervals for various kappa values measured in 100 patients	21
Table 6. Characteristics of ankle injury patients seen over five month period at Ottawa Civic and Ottawa General Hospital emergency departments (Phase 0)	33
Table 7. Use of radiography in ankle injury patients seen over five month period at Ottawa Civic and Ottawa General Hospital emergency departments (Phase 0)	36
Table 8. Ankle injury patients seen and use of radiography by experience of attending physician in Phase 0	37
Table 9. Characteristics of patients included in Phase I, development of the decision rules	39
Table 10. Use of radiography for 750 ankle injury patients in Phase I	40
Table 11. Classification performance of physicians' predicted probability of ankle series fracture with 0.5 probability threshold for ordering x-rays	44
Table 12. Classification performance of physicians' predicted probability of ankle series fracture with 0.1 probability threshold for ordering x-rays	44
Table 13. Characteristics of patients assessed for interobserver agreement in Phase I	47
Table 14. Univariate association of continuous and nominal variables with significant fracture on ankle x-ray series	52
Table 15. Univariate association of ordinal variables with significant fracture on ankle x-ray series	54
Table 16. Variables used in multivariate analyses	60
Table 17. Variables in logistic regression model	62
Table 18. Classification performance of logistic regression model using predicted fracture probability of 0.5 as threshold for ordering an ankle x-ray series	62
Table 19. Classification performance of logistic regression model using predicted fracture probability of 0.015 as threshold for ordering an ankle x-ray series	62
Table 20. Classification performance of clinical decision rule for ankle x-ray series	65

Table 21. Univariate association of variables with significant midfoot fracture on foot x-ray series	69
Table 22. Classification performance of clinical decision rule for foot x-ray series	69
Table 23. Classification performance of ankle series and foot series decision rules combined	72
Table 24. Comparison of use of radiography in Phase I to that possible if decision rules had been implemented	73

LIST OF FIGURES AND ILLUSTRATIONS

Figure 1. Anatomical zones included in the study	15
Figure 2. Distribution of patients and ankle series fracture by age group in Phase 0	34
Figure 3. Distribution of patients and ankle series fracture by month in Phase 0	35
Figure 4. Distribution of patients and fracture by physicians' predicted probability	42
Figure 5. Distribution of patients and fracture by physicians' comfort with no x-ray	43
Figure 6. ROC curve for physicians' predicted probability of ankle series fracture	45
Figure 7. Interobserver agreement (kappa values) of general physical findings	48
Figure 8. Interobserver agreement (kappa values) of swelling	49
Figure 9. Interobserver agreement (kappa values) of bone tenderness	50
Figure 10. Distribution of patients and ankle series fracture by age in Phase I	53
Figure 11. Relative risk (95% CI) of ankle series fracture for variables from history	55
Figure 12. Relative risk (95% CI) of ankle series fracture for general findings	56
Figure 13. Relative risk (95% CI) of ankle series fracture for location of swelling	57
Figure 14. Relative risk (95% CI) of ankle series fracture for bone tenderness	58
Figure 15. Recursive partitioning of predictor variables for ankle series fracture	63
Figure 16. Clinical decision rule for ankle x-ray series	64
Figure 17. ROC curves comparing models derived by LR and RP	66
Figure 18. ROC curves comparing RP model and physicians' predicted probability	67
Figure 19. Relative risk (95% CI) for predictor variables of foot series fracture	70
Figure 20. Clinical decision rule for foot x-ray series	71

1. INTRODUCTION

1.1 Statement of the Problem

Radiographic assessment of acute ankle injury patients is performed primarily to exclude the presence of a clinically significant fracture that may alter patient management. The prevalence of significant fracture (further defined in section 5.4) is less than 15% among patients commonly presenting to emergency departments with an acute ankle injury.^{1 2 3 4 5 6 7} Despite this relatively low proportion of fractures, the vast majority of such patients are routinely referred for an x-ray examination.^{5 8} An ankle x-ray series is typically the second most commonly performed musculoskeletal examination in the emergency department after a cervical spine series.⁹

Such a conservative approach to patient management by emergency physicians leads to many unnecessary radiologic studies. This results in increased radiation exposure and waiting times for the patient and additional costs to the health care system.^{10 11 12}

The Ministry of Health of Ontario reports that more than 180,000 ankle x-ray series are performed annually in the province of Ontario.¹³ Based on this experience, we estimate that some 6 million ankle radiographs are taken each year in Canada and the United States. The annual billings to the Ontario Health Insurance Plan (OHIP) for a high volume procedure like ankle radiography exceeds that of a low volume, high technology procedure such as coronary catheterization (\$4,369,667.50 versus \$4,102,933.20 in 1989).¹³ Even a modest reduction in the proportion of ankle injury patients having an x-ray would lead to large savings in health care dollars.

There are no widely accepted guidelines for the use of radiography in ankle injuries equivalent to those successfully introduced for skull x-rays.^{14 15} The few studies that have addressed this issue have provided contradictory results and have a number of methodological shortcomings in terms of reliability, validity, sensibility, and effectiveness.¹⁶ None of these studies to develop guidelines has effectively addressed the issue of the reliability of physical findings.

Without the support of recognized guidelines, emergency physicians tend to follow the expedient route of ordering a radiograph for most ankle injury patients.¹⁷ This

practice is fostered by the nature of emergency medicine: high case volumes with brief physician-patient contact and very little follow-up.¹⁸ Patients appear to expect an x-ray and physicians are fearful of lawsuit should they miss a fracture.¹⁹

1.2 Contribution of the Study

We feel that emergency medicine requires sensible guidelines for the use of radiography in extremity trauma. Emergency physicians and patients, however, are unlikely to be satisfied with a recommendation that correctly classifies most patients yet misses a few fractures.²⁰ Hence, our aim in this study was to develop decision rules that were 100% sensitive for detecting clinically significant fractures on each of the ankle and foot x-ray series commonly ordered for ankle injury patients. This, we hope, will permit physicians to be more efficient in their use of radiography in the emergency department.

1.3 Organization of the Study

The author was assisted by a clinical committee of two experienced emergency physicians, one from each of the study hospitals. The committee discussed the choice of variables, the methods of patient assessment, and the final selection of variables for the decision rules.

The author was further assisted by three research assistants, one medical resident and two nurses. The research assistants aided with data extraction from the patient charts in Phase 0, collection of data sheets and radiography reports in Phase I, and data entry.

The author was solely responsible for development of the study protocol, design of the data sheets, establishment of the computerized database, verification of the completed data sheets, all statistical analyses, and all written reports, tables, and figures.

This study was supported by a grant (#04090N) from the Emergency Health Services Division of the Ontario Ministry of Health. This grant provided funding for the work of the two nurse research assistants.

2. REVIEW OF LITERATURE

2.1 Methodological Standards

There has been considerable interest, recently, in the area of clinical prediction or decision rules. These rules attempt to reduce the uncertainty of clinical decision making by standardizing the interpretation of clinical data. Rules have been developed to estimate the likelihood of a particular diagnosis or outcome in various clinical scenarios including: myocardial ischemia or infarction in emergency department chest pain patients,^{21 22 23} pneumonia in patients with acute respiratory illness,^{24 25} bacteremia in hospitalized patients,²⁶ and relapse in acute asthmatics.²⁷

Concomitant with the reporting of these prediction rules has been an interest in the methodological standards for their development and validation. Wasson and colleagues critically appraised the methods of 33 published reports of clinical prediction rules and suggested principles to be followed in designing such studies.²⁸ Feinstein assessed hundreds of clinical indexes and rating scales and then devoted an entire book to the standards of constructing and evaluating such indexes.²⁹ The following discussion of the methodological standards for clinical prediction rules depends heavily on the work of Wasson and Feinstein.

2.1.1 Definition of Outcome

The outcome or diagnosis to be predicted must be clearly defined and ideally should be based on biological criteria e.g. biochemical or radiological findings. Outcomes that depend on behavioural criteria, such as the decision to admit a patient, may not be reproducible in other settings.

Assessment of outcome is susceptible to bias. Bias, a systematic error in the conduct or design of a study, may be classified in many ways but two general overall categories are selection bias and observation bias.³⁰ To avoid observation bias, the investigator who assesses outcome should be blinded to the predictor variables. Selection bias will be discussed in section 2.1.4.

2.1.2 Definition of Predictor Variables

The clinical findings to be used as predictors in a decision rule must be clearly defined and the method of assessment must be well standardized. Use of a protocol and a data collection sheet is recommended. Again, to avoid observation bias, the assessment of the predictor variables must be done without knowledge of the outcome. Predictor variables can be more accurately assessed, with less bias, when the data is collected prospectively.

2.1.3 Reliability of Predictor Variables

Clinical prediction rules are highly dependent on clinical findings derived from physical examination. Unless these clinical findings are reliable, the resultant rule will not be dependable. Reliability refers to the consistency or reproducibility of the findings by the same clinician (intraobserver reliability) or by different clinicians (interobserver reliability).^{31 32 33} The assessment of clinical predictor variables should be shown to have good agreement beyond chance.

2.1.4 Selection of Subjects

Selection bias may occur if there are systematic differences in characteristics between patients included in the study and those not included. The method of selection of subjects must be described.

One must know about the population and setting used to develop a clinical prediction rule in order to decide whether the findings are applicable to other populations. Ideally the patients used to derive the rule (derivation set) should encompass a wide spectrum of clinical and demographic characteristics and thereby increase the generalizability of the results. The medical setting and the referral pattern should be described.

2.1.5 Mathematical Techniques

Many techniques are available for deriving prediction rules and those used should be identified and described. The most basic approach is that of two-by-two crosstabulation of the predictor variables and the outcome. The large number of

predictors and the complexity of their relationship usually requires the use of multivariate statistical methods. Commonly used approaches to multivariate analysis include logistic regression analysis, discriminant function analysis, and recursive partitioning analysis. The rationale for selecting one of these techniques must be clearly stated.

2.1.6 Sensibility

Clinical prediction rules should exhibit sensibility, a term used by Feinstein to refer to a composite of many attributes other than reliability and accuracy. These include purpose, overt format, content validity, and ease of use. The purpose and the application of the rule must be clearly stated and must be relevant. The overt format of the rule should be easily comprehensible and should reflect the individual components. The rule should demonstrate content validity, suitability in the selection and aggregation of component variables. Finally, the rule must be easy to use in the intended clinical application; there must be neither too many variables nor excessive complexity in their interpretation. The use of continuous variables in regression equations may be too awkward for clinicians; suitable cutpoints may be more appropriate in the final rule.

2.1.7 Accuracy

The investigator should demonstrate that the newly derived prediction rule accurately classifies patients for their risk of the specified outcome. The proportion of patients who are correctly classified can be expressed in terms of the true positive rate (sensitivity) and the true negative rate (specificity). The former is important when the price of misclassification is the patient receiving inadequate therapy (e.g. not obtaining a cast for a fracture). The latter is important when misclassification leads to the burden of unnecessary treatment or diagnostic tests for the patient (e.g. undergoing needless coronary angiography for misdiagnosed myocardial ischemia). Therefore, the investigator should indicate the relative cost of misclassifying the outcome as falsely negative or falsely positive.

Other useful performance characteristics of a decision rule are the positive predictive value (probability that the person has the outcome, given a positive prediction by the rule) and the negative predictive value (probability that the person does not have

the outcome, given a negative prediction by the rule). The predictive value of a decision rule depends not only on sensitivity and specificity but also on the prevalence of the outcome in the population being assessed.

The simplest method of demonstrating the accuracy of a new prediction rule is to prepare a two by two classification table of the predicted versus the actual outcomes, in the derivation set population. This permits the calculation of the sensitivity, specificity, and overall accuracy of the rule, as well as the potential impact on patients (e.g. diagnostic tests saved or needless admissions prevented). More sophisticated statistical cross-validation techniques, such as the "jackknife" and "bootstrap" methods, can be used to validate the rule on the original study population from which the rule was derived.

2.1.8 Prospective Validation

The misclassification rate tends to be higher when prediction rules are tested on a different population (the validation set) than in the original derivation set.³⁴ Consequently, a prediction rule can be considered more robust if it stands up to prospective validation in a second study in the same location. Such a study would also afford an opportunity for the investigator to assess the reliability or reproducibility of the rule.

The most stringent test of accuracy is to study the misclassification rate in a new population in one or more different clinical settings. This demonstrates that the new clinical prediction rule is not dependent on characteristics specific to the derivation set and that it may be considered generalizable.

2.1.9 Effect on Patient care

The true value of a prediction rule lies in the ability to improve patient care, such as a decrease in unnecessary diagnostic tests or admissions to expensive coronary care units. Despite being methodologically sound, a rule may have little clinical use if it is not practical or fails to gain acceptance by physicians and patients. For example, the impact of a rule for the use of radiography may be lessened by patient demands for x-rays or by physician fear of a malpractice suit. Wasson found that very few prediction rules had undergone a field trial, wherein the effect on patient care was assessed.²⁸ Such

studies may be difficult to undertake but must be done to genuinely demonstrate the effectiveness of clinical prediction rules.³⁵

2.2 Previous Studies

A number of studies have attempted to assess the clinical variables that might be useful in predicting the need for radiography in acute ankle injuries. While most work has been done in the last decade, papers addressing this problem can be identified from as early as 1960.³⁶ The majority of these studies have been carried out in Great Britain, perhaps because the National Health Service has traditionally emphasized cost-effectiveness more than North American health care systems.

2.2.1 Methodological Weaknesses

The results of previous studies have been inconsistent and contradictory and have left physicians uncomfortable with the concept of not routinely ordering a radiograph.¹⁸ These investigations have not led to widely accepted guidelines to govern the ordering of x-rays for ankle injury patients. Many methodological weaknesses can be found in these previous studies and will be briefly outlined in the following paragraphs.

Many previous studies have neglected to clearly describe their outcomes in terms of defining a significant fracture or indicating what anatomical structures were to be included.^{1 2 3 7 8 37 38} Another common problem has been the lack of standardized collection techniques or well defined predictor variables.^{1 2 36 37 38 39} Two studies were flawed by use of a retrospective chart audit to evaluate the clinical variables.^{6 38} No study has attempted to assess the reliability or interobserver agreement of predictor findings.

Most studies suffered from a small sample size, encompassing less than 250 ankle injury patients, of whom less than 15% had fractures.^{1 2 3 8 36 37 38 40 41 42} This both severely limited the validity of multivariate analysis and the generalizability of the findings. In many instances, investigators did not adequately describe their statistical methods and did not attempt to describe the accuracy of their conclusions, even for the derivation set.^{1 2 4 6 36 37 2.}

Previous guidelines or rules for ankle injury radiography have lacked sensibility for clinicians. None has offered 100% sensitivity for significant fractures. One guideline suggested variables that did not make clinical sense, incorporating such illogical predictor variables as the posterior tibial pulse, the achilles tendon, and the colour of the ankle.⁷ Some guidelines required the computation of scores and referral to tables of probabilities or use of handheld computers.^{7 42} Others required the measurement of swelling with a tape or callipers.^{4 41} We feel that busy emergency physicians are unlikely to embrace any protocol that requires more than a "bare hands" approach to patient assessment.

No ankle guidelines have been prospectively validated, with the exception of two studies on the general use of radiography in extremity injuries.^{8 42} Neither of these studies had large numbers of ankle injury patients nor gave specific rules for ankle injuries. Finally, only the study by Brand et al. attempted to demonstrate an impact on actual patient care and only for the general use of radiography in all extremity injuries.⁸

2.2.2 Specific Studies

Most previous studies have significant methodological problems, as described above, and will not be considered in further detail. The author has chosen the four strongest studies for closer scrutiny of the methods and results.

Brooks et al. studied 241 patients, aged 12 to 65 years, at the accident department of the Royal Infirmary in Edinburgh.² Assessments were performed, using a data sheet, by a variety of physicians, including residents and general practitioners. Five factors, including pain, swelling, and bruising were rated on a four point scale and maximal tenderness at one of three sites was indicated. Specific definitions of inclusion criteria and outcome measures were not given. Statistical methods were not described and sensitivity and specificity were not provided. The authors conclude that x-rays are unnecessary in inversion ankle injuries if there is not severe tenderness over the lateral malleolus. This guideline, however, would have missed 12 fractures (total fractures not given) in their own study population, suggesting inadequate sensitivity.

Vargish et al. examined 150 patients, including infants, over six months in an emergency department in Iowa City.³ Clear definition of the outcome measure in terms of clinically significant fractures was not given. Whereas some 24 predictor variables

were assessed, we are given no information about the methods and the personnel involved in the assessments. The data was analyzed using contingency tables to determine the association of predictor variables with fracture. By inspection rather than multivariate analysis, the authors chose two of nine significantly associated variables for their guidelines. They conclude that patients without tenderness below the lateral malleolus and "who could bear some weight" do not require an ankle x-ray series. Sensitivity and specificity are not given and cannot be calculated, with certainty, from the data provided. The guidelines suggested, however, appear to miss 13 of the 19 fractures, a very poor classification rate.

Dunlop and colleagues performed a more robust study on 500 ankle injury patients, also at the Edinburgh Royal Infirmary.⁵ Outcomes were independently assessed and classified either as important fractures or as injuries requiring only symptomatic treatment. Fifteen predictor variables were assessed by "the casualty officer on duty", using a data collection sheet. The reliability of these predictor variables was not determined. Logistic regression identified three variables that contributed significantly to the identification of fractures: tenderness localized to the lateral malleolus, inability to bear weight, and age over 60 years. This model would have identified 95% of 60 clinically significant fractures. Unfortunately, standardized and reproducible definitions of localized tenderness and ability to bear weight were not given. Furthermore, 21 fractures of the base of the fifth metatarsal were not considered by this approach. Finally, the authors did not discuss the importance of prospectively validating their findings.

Brand et al. produced a more methodologically sound study which developed protocols for radiography in all extremity injuries.⁸ While this investigation involved 1,712 patients in the development and validation of the protocol, the derivation set included only 248 ankle injuries. Significant outcome fractures were not defined for the ankle region. Nine predictor variables were assessed by clinical assistants (non-physicians) for all extremity injuries. The reliability of the findings was not determined. Recursive partitioning techniques developed protocols for various body regions. Fractures distal to the knee could be predicted with 88.0% sensitivity by identifying patients with gross signs or point tenderness. Unfortunately, this does not permit the

clinician to specifically assess the ankle and we cannot estimate the sensitivity for this subgroup. Furthermore, the major predictor, point tenderness, was not defined sufficiently to be clinically useful. This study did prospectively validate the findings and did demonstrate a 16% reduction in the ordering of lower limb radiography. Use of a poorly described control group, however, limited the validity of the latter conclusion.

2.3 Relevant Clinical Variables

At least 29 different clinical variables based on the history and physical examination have been assessed in previous studies as predictors of fracture in ankle injury patients (Tables 1 and 2). Very few of these have been found to be consistently useful, perhaps reflecting the methodological weakness of the majority of the studies. Tables 1 and 2 indicate the relative predictive value of the various variables in the four studies described in the above sections.

The weak statistical reporting in these studies does not permit a quantitative analysis of the sensitivity or odds ratios of individual predictor variables. Consequently, a qualitative rating has been used to indicate the relative value of the predictors within each study. This process, therefore, was more useful for identifying potential predictors for the current study rather than for attempting a meta-analysis of previous findings.

Age was the only variable based on history found to be useful in more than one of these studies. Ability to bear weight was helpful in three studies but the association with fracture was inconsistent. Furthermore, ability to bear weight was never precisely defined nor assessed for interobserver agreement. Localized tenderness was consistently useful but was not well defined as to the exact location of the tenderness. Several of the weaker studies found ecchymosis and swelling to be highly predictive, contrary to the findings presented in Tables 1 and 2. Numerous other findings were found to be slightly useful in at least one study.

No consistent picture of the usefulness or reliability of these predictor variables emerges from review of the previous investigations. The author felt that a broad selection of variables for the present study was justified, in view of these inconsistent findings. More details about variable selection are given in section 5.2.1.

Table 1. Clinical predictor variables based on history as assessed in previous studies

Variable	Vargish ³	Brand ⁸	Dunlop ⁵	Brooks ²
Age	++	+	++	O
Sex	O	-	O	-
Time from injury	O	-	O	-
"Cracking" sound	+	-	-	-
Mechanism of injury	+	-	-	O
Previous ankle fracture	-	-	O	-
Bear weight immediately	+	-	-	-
Left or right foot	+	-	-	-
Manner of arrival	O	-	-	-
Immediate onset pain	O	-	-	-
Pain at rest	-	+	-	O
Type of activity at injury	O	-	-	-
Medical risk bone injury	-	+	-	-
Impaired sensation	-	+	-	-

+++ most useful ++ moderately useful + slightly useful
 O not useful - not assessed

Table 2. Clinical predictor variables based on physical examination as assessed in previous studies

Variable	Vargish ³	Brand ⁸	Dunlop ⁵	Brooks ²
Ecchymosis	O	O	O	O
Range of motion	O	+	-	-
Drawer sign	-	-	-	O
Ability to bear weight	+++	+	++	-
Swelling location	O	-	O	O
Swelling degree	O	+	O	O
Tenderness - point	+++	+++	+++	+++
Tenderness - bone/soft	++	+++	+++	+++
Bone conduction	+	-	-	-
Impaired mentation	-	+	-	-
Pain with motion	O	-	-	O
Open wound	-	+	-	-
Gross deformity	+	+++	-	-
Neurovascular deficit	O	-	-	-
Palpable mass	-	+	-	-

+++ most useful ++ moderately useful + slightly useful
 O not useful - not assessed

3. GOALS AND OBJECTIVES

The work of this thesis comprises two distinct phases. Phase 0 is a review of medical records to determine the current use of radiography for ankle injury patients: the proportion of patients referred for an x-ray series and the yield of the x-rays for fracture. This is followed by Phase I, a cross-sectional survey, which develops clinical decision rules to guide emergency physicians in their use of radiography. Phase I alters physician behaviour, hence Phase 0 must come first in order to accurately reflect the true x-ray ordering practice in emergency departments.

The Discussion will describe two further phases that might complete this work in the future: Phase II would prospectively validate the rules; Phase III would implement the rules and assess the actual impact on patient care.

3.1 Phase 0: Baseline Assessment

3.1.1 Goal

The goal of Phase 0 of this study was to determine the current use of radiography in ankle injury patients in two major hospital emergency departments. This information will then be used to describe the potential impact of the clinical decision rules for reducing the numbers of unnecessary radiographs. This data may also serve as a control for a future effectiveness study (Phase III), to be described in the Discussion section.

3.1.2 Objectives

To collect baseline information about the current use of radiography, at two hospitals, for emergency department patients with ankle injuries:

- a) demographic and clinical characteristics of patients
- b) proportion referred for ankle and foot x-ray series (referral fraction)
- c) proportion of x-rays positive for significant fracture (yield)
- d) costs to OHIP of radiography
- e) waiting times for patients with and without radiography.

3.2 Phase I: Development of the Rules

3.2.1 Goal

The goal of Phase I was to develop clinical decision rules that are 100% sensitive for detecting clinically significant fractures in ankle injury patients.

3.2.2 Objectives

- a) To develop and pretest standardized assessment methods for ankle injury patients
- b) To perform standardized clinical assessments on patients with ankle injuries and determine which variables are most predictive of a fracture
- c) To test the interobserver reliability of the assessment of these clinical variables
- d) To use multivariate techniques to develop highly sensitive clinical decision rules for guiding the use of radiography in acute ankle injuries
- e) To assess the classification performance of the derived decision rules

4. METHODS - PHASE 0: BASELINE ASSESSMENT

4.1 Patient Population

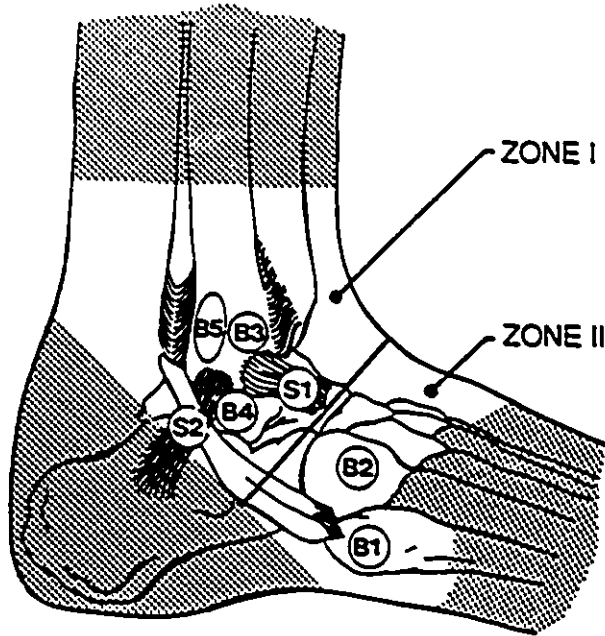
4.1.1 Inclusion Criteria

All patients presenting to the Ottawa Civic and Ottawa General Hospital emergency departments, with any pain or tenderness of the ankle after sustaining acute blunt injuries (including twisting injuries, falls from a height, direct blows, and vehicle accidents) were eligible. "Ankle" was broadly defined to include the area generally involved in common twisting injuries and was subdivided into two zones representing the malleolar region and the midfoot respectively. These zones, as depicted in Figure 1, correspond to the areas that generally require assessment by a standard ankle x-ray series (zone I or the malleolar region) or a standard foot x-ray series (zone II or the midfoot). We defined the zones to include the following anatomical structures and their overlying soft tissues: a) zone I (the malleolar region): distal 6 cm of tibia, distal 6 cm of fibula, and talus b) zone II (the midfoot): navicular, cuboid, cunieforms, the anterior process of the calcaneus, and base of the fifth metatarsal. The body and the tuberosities of the calcaneus were not included in this definition.^{43 44}

4.1.2 Exclusion Criteria

- a) Patients under the age of 18 years. Young patients with immature skeletal development may have different clinical characteristics than adults
- b) Pregnant women, due to concern about exposure to radiation and harm to the fetus
- c) Isolated injuries of the skin. Superficial lacerations, superficial abrasions, puncture wounds, and burns rarely require radiography
- d) Patients referred from outside the hospital and who have already had x-rays. Patients with known fractures would bias the clinical assessment
- e) Patients whose ankle injury occurred more than 10 days previously are not considered to have an acute injury

LATERAL VIEW



MEDIAL VIEW

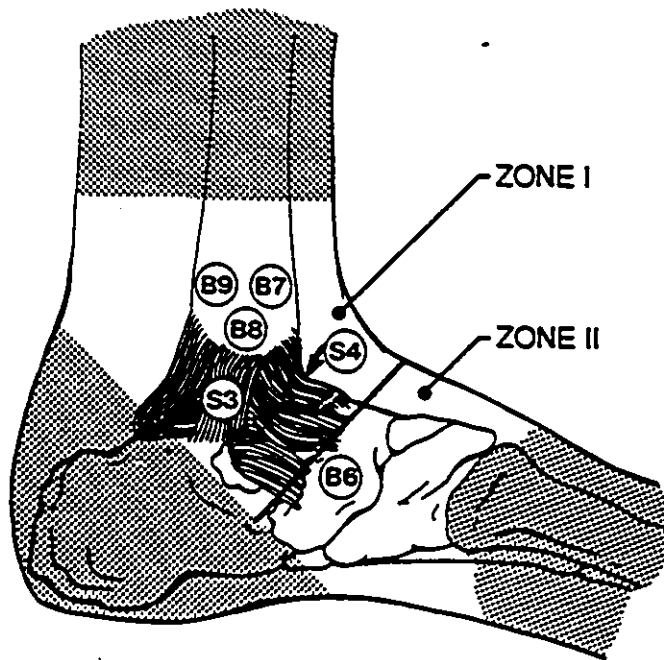


Figure 1. Anatomical zones included in the study

f) Patients who return for reassessment of the same ankle injury have already had radiography and would therefore bias the assessment. These patients have likely already been entered into the study

4.2 Data Collection

We reviewed the records of treatment for all ankle injury patients seen at the Ottawa Civic (OCH) and Ottawa General Hospital (OGH) emergency departments over a five month period, February 1 through June 30, 1990. Slightly different methods were used for this retrospective chart review at each hospital. At OGH, our research assistant screened all records of treatment for ankle injury patients. At OCH, the emergency records were pre-screened for ankle or foot problems by the billing nurse. The research assistant then assessed the selected charts for eligible cases. This approach was augmented by a review of the daily log of patient visits. Data abstraction was performed by the three research assistants and some of the charts were selected and verified by the investigator.

The following data was collected and entered into a computerized database, using the SPSS Data Entry II programme:⁴⁵

- a) Patient demographic information: unique number, age, sex, date of visit
- b) Radiography use and results: ankle / foot / ankle and foot / none, corresponding OHIP codes, result of x-ray, and total time spent in emergency department from registration until discharge or referral to orthopedic service (available at OGH only)
- c) Physician information: staff / resident / intern (to detect differential patterns of radiography use)

4.3 Data Analysis

Descriptive statistics were tabulated for the following information:

- a) Demographic and clinical profile of patients: age, gender, date seen, mechanism of injury, prevalence and type of fractures
- b) Use of radiography:

- i) proportion of patients referred for x-ray
 - ii) mean number of x-ray series per patient
 - iii) proportion of x-rays positive for significant fracture
 - iv) patterns of x-ray use
 - v) cost of x-rays performed based on the estimated 1991 OHIP schedule of benefits for physicians services. The 1991 OHIP schedule has not yet been published but the rates have been estimated to equal the 1988 schedule⁴⁶ plus 8% (the recently approved fee increase)
 - vi) mean time spent in emergency department
- c) Physician use of radiography by status

5. METHODS - PHASE I: DEVELOPMENT OF THE RULE

5.1 Patient Population

5.1.1 Inclusion and Exclusion Criteria

The criteria were those used for Phase 0.

5.1.2 Sample Size

750 patients were assessed, in total, at the two hospitals during this phase and 100 of these were studied for interobserver agreement. Since no hypothesis was tested, the sample size was primarily based on the precision of the sensitivity and specificity of the derived decision rule for malleolar region (zone I) injuries. Zone II (midfoot) injuries are less common and less important clinically, hence we felt that sample size considerations for these cases should be secondary to those of zone I. 95% confidence interval estimates for zone I will be presented.

Other considerations were the precision of the estimates of interobserver variability and the logistic regression coefficients. The sample size had to accommodate both the large number of clinical variables (32) being assessed and the large number of physicians (21) making the assessments. A sample size of 750 clearly exceeds the ten subjects per variable recommended as a rule of thumb for multivariate analyses.

a) **Sensitivity and specificity for zone I injuries:** The goal of the study was to develop a decision rule that was 100% sensitive for detecting fractures with the highest possible specificity. We wished to achieve this goal with 95% confidence that the estimated parameter was within ± 0.05 of the true value. Sensitivity is calculated from the number of ankle series (zone I) fracture cases, which was estimated, from the Phase 0 findings, to be 10% of the total number of ankle injury patients seen. Specificity is calculated from the number of non-fracture cases, which was estimated to be 90% of the patients. The following formula was used to calculate sample size for fracture and non-fracture cases, based on varying levels of sensitivity and specificity:⁴⁷

$$N = 4 Z_{\alpha}^2 P (1 - P) / W^2$$

N = total number of patients required

Z_{α} = 1.96 for 95% confidence level

P = expected proportion (sensitivity or specificity)

W = total width of confidence interval

Table 3 shows the sample sizes required for various levels of sensitivity and specificity. As no confidence interval can be calculated for the desired sensitivity of 100%, the final sample size was conservatively based on the number of fracture cases required for a sensitivity level of 95%. Therefore, given the requirement of 73 ankle series fracture cases and a fracture prevalence of 10%, the total sample size was estimated to be 730 ankle injury patients. This would clearly surpass the number of patients required for a specificity of 90% or less. Considering that a total of 250 eligible patients are seen monthly at the two hospitals, the minimum accrual period would be three months. To be conservative, six months were allowed to reach the target of 730 ankle injury patients.

Table 3. Sample size required to estimate various levels of sensitivity and specificity with 95% confidence \pm 0.05 for zone I injuries

Fracture Patients Required		Non-fracture Patients Required	
Sensitivity	Sample Size	Specificity	Sample Size
0.95	73	0.50	384
0.90	138	0.40 0.60	369
0.85	196	0.30 0.70	323
0.80	246	0.20 0.80	246
		0.10 0.90	138

b) **Sensitivity and specificity for zone II injuries:** One can estimate 95% confidence intervals, for various levels of sensitivity and specificity for zone II injuries, given a total sample size of 730. From phase 0, we would expect 30% (220) of the

patients to have zone II pain or tenderness and 4% (30) to have midfoot fractures. The 95% confidence intervals were calculated, based on the normal approximation of proportions:⁴⁸

$$P \pm 1.96 \sqrt{P(1-P) / N}$$

Table 4 indicates that the estimate of 95% sensitivity for detecting 30 zone II fractures would be within ± 0.08 of the true value. The least precise estimate of specificity for 190 non-fracture zone II injuries would be within ± 0.07 of the true value.

Table 4. 95% confidence intervals for various levels of sensitivity and specificity for zone II injuries

30 Fracture Patients		190 Non-fracture Patients	
Sensitivity	95% C.I.	Specificity	95% C.I.
0.95	± 0.08	0.50	± 0.07
0.90	± 0.11	0.40 0.60	± 0.07
0.85	± 0.13	0.30 0.70	± 0.07
0.80	± 0.14	0.20 0.80	± 0.06
		0.10 0.90	± 0.04

c) **Interobserver agreement:** The number of patients required for the measurement of interobserver agreement was dictated primarily by considerations of feasibility. The clinical committee felt that it would be logistically impossible to have all 730 patients assessed by two physicians and that 100 would be a manageable number. Therefore, expected 95% confidence intervals, for various kappa levels, were calculated for a sample size of 100. These confidence intervals were estimated by the normal approximation of proportions. Table 5 indicates that the least precise estimate, for a kappa of 0.50, would be within ± 0.10 of the true value. A more accurate method of determining sample size for the kappa statistic may have resulted in broader confidence intervals but was felt to be too complex for the scope of this thesis.⁴⁹

Table 5. 95% confidence intervals for various kappa values measured in 100 patients

Kappa Value		95% Confidence Interval
0.90	0.10	± 0.06
0.80	0.20	± 0.08
0.70	0.30	± 0.09
0.60	0.40	± 0.10
0.50		± 0.10

5.1.3 Patient Selection

Consecutive ankle injury patients were entered into the study if they met the inclusion and exclusion criteria and if one of the 21 designated assessor physicians (described in section 5.2.5) was available. One of the assessor physicians was on duty in the ambulatory area of the two emergency departments approximately 90% of the time. For the duration of the study, interns and residents were asked not to see ankle injury patients. The decision to enter a patient into the study was made by the assessor physician based on the criteria described in sections 4.1.1 and 4.1.2., i.e. any pain or tenderness in zone I or zone II after blunt trauma. Regular review of all patient charts revealed that very few eligible patients were not entered into the study by the assessor physicians.

5.1.4 Ethics

The research ethics committees of the participating hospitals agreed that informed consent need not be obtained for this study. Normal patient management would not be altered and patients would not be subjected to additional risk.

5.2 Standardized Patient Assessment

5.2.1 Selection of Variables

Patients underwent a standardized assessment of clinical variables that had been defined by the clinical committee of emergency medicine specialists. These variables were felt to be potentially useful in predicting whether or not ankle injury patients were likely to have a fracture and would therefore require x-ray examination. This judgement was based on the experience of the committee and on the results of previous studies, described above.

We decided that several variables, found not to be useful in most previous studies, were unlikely to be of value in predicting which patients require an x-ray. Consequently, the following factors were not assessed: right or left foot, manner of arrival at hospital, time of onset of pain, pain at rest, activity at time of injury, use of analgesics after injury, medical risk of fracture (cancer, chronic renal failure, Paget's disease), history of impaired sensation in lower limbs, altered level of consciousness, wound overlying bone tenderness, bone conduction, pain with motion, vascular deficit, neurological deficit. The presence of bone deformity or crepitation obviously mandates an x-ray and was felt to be an unnecessary variable in this study.

5.2.2 Pilot Stage

The data collection sheets (Appendix A) and the assessment techniques were evaluated and revised during a one month pilot period that involved 155 patients, prior to the actual study. This pilot period also allowed additional time for the training and education of the assessor physicians (described in section 5.2.5).

This stage prompted a number of alterations and refinements to the data collection sheets (Appendix B). Omitted were questions dealing with recent use of analgesics, history of impaired sensation in the lower limbs, and evidence of an altered level of consciousness. Questions about swelling, range of motion, and ability to bear weight were changed from a "yes/no" response to a four point scale of "none/minimal/moderate/marked", in order to gain more information about these findings. More specific instructions about assessing ability to bear weight were given to

the treating physicians. Two points of bone tenderness on the malleoli were dropped and two in the midfoot were added. The anatomical diagrams were subdivided into two zones to allow the physicians the option of ordering ankle or foot x-ray series, or both. Finally, the physicians were asked to predict the probability of a fracture in both zones and to indicate their theoretical degree of comfort in not ordering any x-rays for each patient.

5.2.3 Variables from History

The following information, based on the patient's history, was gathered by the assessing physician:

- a) Age
- b) Sex
- c) Ability to bear weight on ankle immediately after injury (yes and continued current activity / yes but with difficulty / no)
- d) Popping or cracking sound at time of injury
- e) Mechanism of injury (twisting / fall from a height / direct blow / vehicle accident / other)
- f) Direction of injury (inversion / eversion / plantar flexion / unknown / not applicable)
- g) Time from injury to assessment in hours
- h) Application of ice in previous 6 hours
- i) Ankle fracture within previous 12 months

5.2.4 Variables from Physical Examination

The following information, based on the physical examination, was gathered by the assessing physician:

- a) Ecchymosis (yes / no)
- b) Visible swelling (none / minimal / moderate / marked):
 - i) Medial malleolus
 - ii) Lateral malleolus
 - iii) Anterior talofibular ligament
 - iv) Anterior aspect of ankle
- c) Restriction of active plantar-dorsiflexion (none / minimal / moderate / marked)

- d) Bone tenderness (yes / no):
- | | |
|------------------------------------|--------------------------------------|
| i) Proximal fibula (B0) | ii) Base of fifth metatarsal (B1) |
| iii) Cuboid (B2) | iv) Lateral malleolus anterior (B3) |
| v) Lateral malleolus inferior (B4) | vi) Lateral malleolus posterior (B5) |
| vii) Navicular (B6) | viii) Medial malleolus anterior (B7) |
| ix) Medial malleolus tip (B8) | x) Medial malleolus posterior (B9) |
- e) Soft tissue tenderness:
- | | |
|---------------------------------------|-----------------------------------|
| i) Anterior talofibular ligament (S1) | ii) Calcaneofibular ligament (S2) |
| iii) Deltoid ligament (S3) | iv) Anterior aspect ankle (S4) |
- f) Anterior drawer sign (positive / negative) "Positive" was defined as greater anterior displacement of the foot on the tibia compared to the uninjured ankle
- g) Ability to bear weight now by questioning (yes / no)
- h) Degree of difficulty in walking demonstrated in emergency department; must be able to bear weight for a minimum of four steps (none / minimal / moderate / marked / unable / refused / not applicable)

5.2.5 Patient Assessment

All eligible patients were entered into the main study when one of the 21 designated assessor physicians was on duty. The assessor physicians were all fulltime certified emergency staff physicians (i.e. not interns, residents or general practitioners) at the two hospitals. We trained these physicians, by means of a one hour lecture and a practical demonstration, to assess the clinical variables in a uniform manner. The physicians examined the patients, recorded their findings on the data collection sheets, and then sent all patients for an x-ray examination.

5.2.6 Physicians' Predicted Probability

To help evaluate the potential usefulness of the clinical decision rule, each physician assessor estimated the likelihood that the patient had a significant fracture in zone I and zone II. This was done before the patient was x-rayed and estimated probability to the closest ten percentile. The purpose of this question was to allow comparison between the judgement of experienced physicians and the final decision rules.

Physicians were also asked to indicate their theoretical degree of comfort in not ordering any x-rays for each patient, on a five point Likert scale (very comfortable to very uncomfortable). This would provide another indication of the potential to reduce use of radiography.

5.3 Assessment of Interobserver Reliability

One hundred of the patients were assessed for the 23 physical examination variables by a second emergency physician who was blinded to the results of the first assessment. This was carried out, on a convenience basis, whenever two fulltime emergency staff physicians were available together in the department.

5.4 Outcome Measures

All patients in the study had at least one x-ray series performed. Patients with tenderness in zone I (malleolar region) underwent a standard ankle x-ray series examination and those with tenderness in zone II (midfoot) had a standard foot x-ray series examination. Patients with tenderness in both zones, therefore, underwent both an ankle and a foot series. Phase 0 data indicated that only 15% of ankle injury patients routinely underwent both an ankle and a foot x-ray series. Consequently, we felt that the ordering of both series in all patients would be excessive and unacceptable to physicians. We felt that the risk of missing a significant fracture within one of the zones, given the absence of any pain or tenderness in that zone, was remote.

The x-rays were interpreted by independent, qualified radiologists who were blinded to the contents of the data collection sheets. They classified the presence of traumatic lesion in the zone I and zone II structures as follows:

- a) No Significant Fracture, defined as no fracture or avulsion fractures of less than or equal to three millimetres across
- b) Clinically Significant Fracture, defined as all other fractures

This definition of fracture was agreed upon by members of the orthopedic departments. The concept of significant fracture reflects clinical management in that

avulsion fractures of three mm or less are not treated with plaster immobilization in the study institutions. In order to be conservative and consistent with the 3 mm definition, we classified fractures of the base of the fifth metatarsal as significant, even though these are rarely treated with a cast.

In summary, therefore, clinical decision rules were developed separately for the outcome measures in each zone i.e. clinically significant fractures in zone I as demonstrated on ankle x-ray series and in zone II as demonstrated on foot x-ray series.

5.5 Data Analysis

The research assistants added information from the records of treatment and radiography reports to the data collection sheets. All data sheets and radiography reports were reviewed by the author prior to data entry into the computerized database. Analyses were performed on the University of Ottawa mainframe computer and an IBM compatible personal computer using SPSS statistical software packages.⁵⁰

5.5.1 Interobserver Agreement

The interobserver agreement for each variable was measured by calculating the kappa coefficient, the proportion of potential agreement beyond chance that was achieved.^{51 52 53} The basic, unweighted kappa calculations were used for simplicity. Even though weighted kappa values may be more appropriate for ordinal data, we found that the weighted and unweighted values would be approximately the same because the number of cases more than one category different was less than five percent of the total. 95% confidence intervals were calculated, where sigma represents the standard error of kappa, as:⁵²

$$K \pm 1.96 \sigma$$

Appendix C presents a sample calculation of kappa, applied to the interobserver agreement on the presence or absence of moderate to marked swelling over the medial malleolus.⁵⁴ A kappa value of +1.0 represents perfect agreement and values greater

than 0 represent observed agreement greater than that due to chance. The following guidelines for interpreting the strength of agreement from kappa values has been suggested: less than 0.40 poor to fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial, and greater than 0.80 almost perfect agreement.⁵¹

5.5.2 Univariate Analysis

Univariate analyses determined, for each variable, the strength of association with fractures in the ankle and foot x-ray series, respectively. All tests assumed independent samples and a P value of less than 0.05 was considered significant. No adjustment for multiple testing was made as the purpose of the univariate analyses was the screening of variables and not the testing of a hypothesis. The appropriate univariate technique was chosen according to the type of data: for nominal data, the chi-square test with continuity correction; for ordinal variables, the Mann-Whitney U test; and, for continuous variables, the unpaired 2-tailed t-test, using pooled or separate variance estimates as appropriate. Swelling and restriction of range of motion were assessed using an ordinal scale of "none, minimal, moderate, marked". Analysis of these variables was done twice: first as ordinal data and secondly as data that had been dichotomized as "none-minimal" or "moderate-marked".

5.5.3 Multivariate Techniques

Multivariate techniques were used to analyze those variables found to be both reliable ($\text{kappa} > 0.6$) and strongly associated with a significant fracture ($P < 0.05$). The objective was to find the models with the best combination of predictor variables for the ankle and the foot x-ray series, separately. We sought models that would be 100% sensitive for detecting significant fractures while achieving the maximum possible specificity.

5.5.4 Logistic Regression

Logistic regression analysis was considered to be more appropriate than several alternative techniques to develop a model that could predict the binary outcome of fracture or no fracture from the independent predictor variables. Logistic regression is

used for analyses with a binary dependent variable because the assumptions necessary for hypothesis testing in multiple linear regression analysis are violated: the distribution of errors cannot be assumed to be normal.⁵⁵ Another multivariate technique, discriminant function analysis, requires that many of the independent variables be continuous for optimal performance. In this study, however, most of the predictor variables were nominal.

In the logistic regression model the risk of developing a fracture may be expressed as a function of the predictor variables x_1, x_2, \dots, x_r :⁵⁶

$$\text{Probability of fracture} = \frac{1}{1 + e^{-(a + b_1x_1 + \dots + b_r x_r)}}$$

The logistic model can be rewritten in terms of the odds of having a fracture, as the log odds or logit:

$$\text{Logit} = \text{Log} \frac{\text{Prob}(\text{Fracture})}{\text{Prob}(\text{NoFracture})} = a + b_1x_1 + \dots + b_r x_r$$

Therefore, the simple odds can be calculated as:

$$\frac{\text{Prob}(\text{Fracture})}{\text{Prob}(\text{No Fracture})} = e^{a + b_1x_1 + \dots + b_r x_r}$$

Model building proceeded with forward stepwise variable selection until no variables meet the entry (0.05) or removal (0.10) criteria for the significance level of the likelihood-ratio test.⁵⁷ In order to provide a more simple model for clinicians, cutpoints were sought for continuous variables. Such variables were categorized by entering several different clinically significant cut-off values into the logistic regression equation to determine which had the best predictive value. For example, increased age was found to be significantly associated with fracture by univariate analysis. Age was entered into the logistic regression, initially using age forty-five as a cutpoint, and then at five year intervals for successive analyses.

The logistic regression software programme provided a classification table of predicted versus observed fracture status for the cases in each model. The predicted status was based on a fracture probability threshold of 0.5. Such a threshold provided very high overall accuracy and specificity but was relatively insensitive for fractures. This appeared to be a result of the low prevalence of fractures.

In order to improve the sensitivity of the logistic regression models for fractures, the author sought alternate probability thresholds. This was done by manually reviewing the casewise listings of the predicted probability of fracture and the actual status for fracture for each case. A lower probability threshold could be chosen that provided 100% sensitivity for fractures. The classification was then retabulated, with the new threshold value, to give the accuracy and specificity of the model.

Further classifying the cases at different probability threshold levels provided a series of sensitivity and specificity values, which were used to construct a receiver operating characteristic (ROC) curve.⁵⁸ The ROC curve is a graphic method for indicating the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1 - specificity) of a test or diagnostic manoeuvre. One can compare two tests by assessing their areas under the ROC curve: the most discriminating tests have the largest area. Our goal, however, was to develop models with maximum specificity, given 100% sensitivity, and not necessarily with maximum discrimination or accuracy. Consequently, areas under the ROC curve were not calculated.

5.5.5 Recursive Partitioning

A recursive partitioning technique was used as an alternative to logistic regression, in the expectation of achieving better classification of the fracture cases. Logistic regression, while providing good overall accuracy, could only offer 100% sensitivity for fractures after manipulation of the probability threshold. The resultant models had relatively low specificity. The recursive partitioning process could be guided to correctly classify one outcome group at the expense of the other. In this study, identification of all the fracture cases was the priority of the model building, even at the expense of misclassifying more of the nonfracture cases.

Recursive partitioning, as described by Friedman and Ciampi, is able to summarize complex data by identifying and describing homogeneous strata.^{59 60 61} These subpopulations are defined by values of certain clinical variables and are homogeneous for a specified outcome, such as the risk of having a fracture. The resultant models may be easier to interpret for clinicians in that they are based on the simple presence or absence of predictor variables, rather than regression coefficients.

The recursive partitioning approach successively divides the population into subpopulations of increasing homogeneity until no further splitting can be done. The splits are based on the predictor variables so that the subpopulations can be described in direct clinical terms. The first split is based on the variable with the largest significant chi-square statistic and the population is then subdivided according to the values of that variable. The procedure is repeated for each subpopulation until no variable produces a significant chi-square.

Several statistical software packages are available for performing recursive partitioning,^{62 63 64} but the author chose to perform the technique using simple repetitive cross-tabulation analyses. The eligible predictor variable with highest chi-square value, with one degree of freedom, was chosen for the first partitioning. All cases positive for this variable were put aside as being at high risk for fracture. All cases negative for this variables then underwent repeat cross-tabulation analysis. Again, those positive for the predictor with the highest chi-square value were put aside as high risk. This process was repeated until all fractures cases had been identified for the high risk group. Those remaining were considered to be at low risk for fracture. Those predictor variables responsible for the partitioning steps were integrated into a decision rule model.

The performance of this model was assessed by developing a two by two classification table of predicted versus actual fractures. As for the logistic regression model, sensitivity and specificity could be calculated. Construction of an ROC curve from this simplistic model required further manipulation of the data. The predictor variables of the model were forced into a logistic regression analysis to estimate fracture probabilities for the cases. Again, creating classification tables for various probability thresholds allowed calculation of a series of sensitivity and specificity values, which could then be used to create an ROC curve.

6. RESULTS - PHASE 0: BASELINE ASSESSMENT

The information gathered in this phase describes the general demographic and clinical features of patients commonly presenting to emergency departments with acute ankle injuries. This data allows us to confirm that typical ankle injury patients were used in the rule development in Phase I. Phase 0 data will also provide a baseline of physicians' radiography ordering practices that can be used as a historical control for a future impact study, Phase III.

We used several different data collection approaches in order to ensure completeness of the sampling process. At the Ottawa General Hospital (OGH), all emergency records were reviewed manually to select eligible cases. At the Ottawa Civic Hospital (OCH), a manually performed audit determined that some 10% of cases had been missed during the primary chart identification by the emergency department billing nurse. Consequently, a research assistant manually reviewed the daily patient visit log and reviewed all cases that might be eligible and were not already in the Phase 0 database. This process reassured us that virtually all eligible ankle injury patients at both hospitals were included.

6.1 Characteristics of Patients

Table 6 demonstrates the demographic and clinical features of 1,290 ankle injury patients seen between February 1 and June 30, 1990 at the two study hospitals. The patient profiles are very similar at the two hospitals and further remarks will refer to the overall descriptive statistics. Patients were young, on average, but the age range extended to 92 years. Males and females were equally represented and the majority had experienced a twisting mechanism of injury.

The overall prevalence of significant fracture was 14.5%: 10.7% in the malleolar region and 3.8% in the midfoot. A further 4.3% experienced clinically insignificant avulsion fractures. Figure 2 reveals the striking rise in prevalence of malleolar region fracture with increased age, from 5.0% to 27.9%. 34.1% of the fractures were found in patients aged 55 or greater, a group that represented only 13.4% of all patients. More

patients were seen in May and June but the prevalence of malleolar fracture was much higher in February (Figure 3).

6.2 Use of Radiography

Use of radiography for ankle injury patients was similar at the two hospitals, with referral fractions of 96.2% at OCH and 92.3% at OGH (Table 7). Overall, 81.3% of patients were referred for an ankle series and 28.9% for a foot series. 15.8% of the patients had both an ankle and a foot series leading to an overall mean of 1.1 x-ray series per patient. The yield of radiography, the proportion of x-ray series positive for a significant fracture, was 13.2% for the ankle series and 13.1% for the foot series. In other words, 86.8% (1,235) of the 1,422 radiographs performed were negative for a significant fracture.

The costs to the Ontario Health Insurance Plan of five months of radiographs for ankle injury patients was estimated to exceed \$45,000.00. The mean time in the emergency department at OGH was 35.8 minutes greater if a patient was referred for an x-ray series.

The use of radiography by different groups of physicians, as shown in Table 8, is very similar whether the treating physician is an intern, a resident, a parttime emergency physician, or a fulltime emergency physician. This data is presented for descriptive purposes rather than hypothesis testing; hence, no formal tests of significance have been performed.

Table 6. Characteristics of ankle injury patients seen over five month period at Ottawa Civic and Ottawa General Hospital emergency departments (Phase 0)

Characteristic	OCH (N = 657)	OGH (N = 633)	Total (N=1290)
Mean age in years (\pm SD) range	36.8 \pm 15.9 18-88	35.4 \pm 14.5 18-92	36.1 \pm 15.2 18-92
Male (%)	353 (53.7)	341 (54.0)	694 (53.8)
Mechanism of injury (%)			
Twisting	555 (84.5)	543 (85.8)	1098 (85.1)
Direct Blow	64 (9.7)	38 (6.0)	102 (7.9)
Fall from a height	20 (3.0)	16 (2.5)	36 (2.8)
Vehicle Accident	8 (1.2)	6 (0.9)	14 (1.1)
Other	10 (1.5)	30 (4.7)	40 (3.1)
Significant fractures (%)	103 (15.7)	84 (13.3)	187 (14.5)
Malleolar region (%) *	73 (11.1)	65 (10.3)	138 (10.7)
lateral malleolus	67	56	123
medial malleolus	19	14	33
posterior malleolus	11	7	18
talus	1	2	3
Midfoot (%) *	30 (4.6)	19 (3.0)	49 (3.8)
base 5th metatarsal	26	14	40
navicular	2	4	6
anterior process calcaneus	4	5	9
cuboid	2	2	4
cunieforms	1	1	2
Avulsion fractures (%) *	32 (4.9)	24 (3.8)	56 (4.3)
lateral malleolus	11	8	19
talus	9	4	13
medial malleolus	5	2	7
base 5th metatarsal	4	3	7
cuboid	3	4	7
navicular	4	1	5
anterior process calcaneus	1	4	5
posterior malleolus	2	-	2
cunieforms	1	1	2

* Patients may have fractures in more than one location

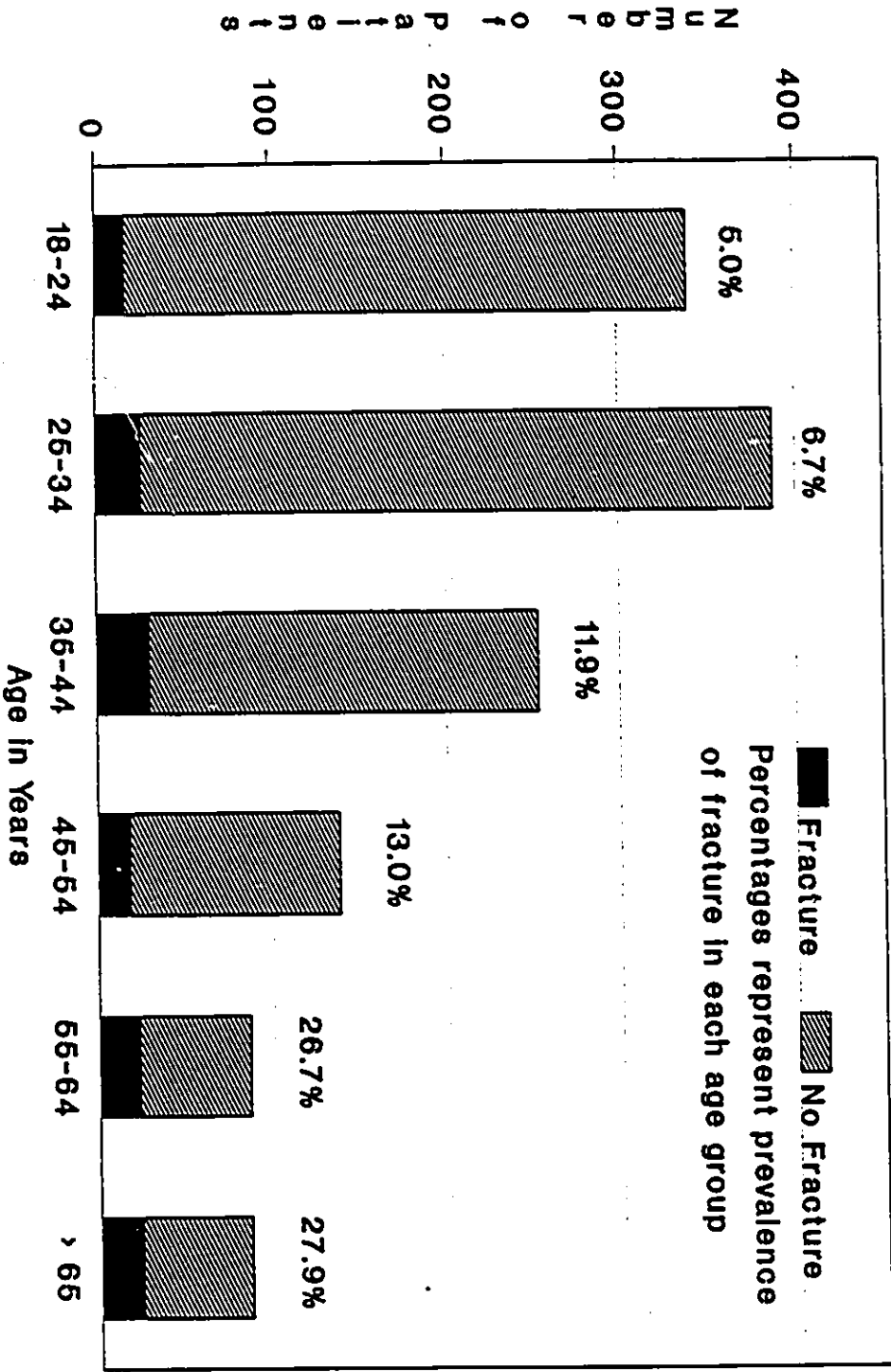


Figure 2. Distribution of patients and ankle series fracture by age group in Phase 0

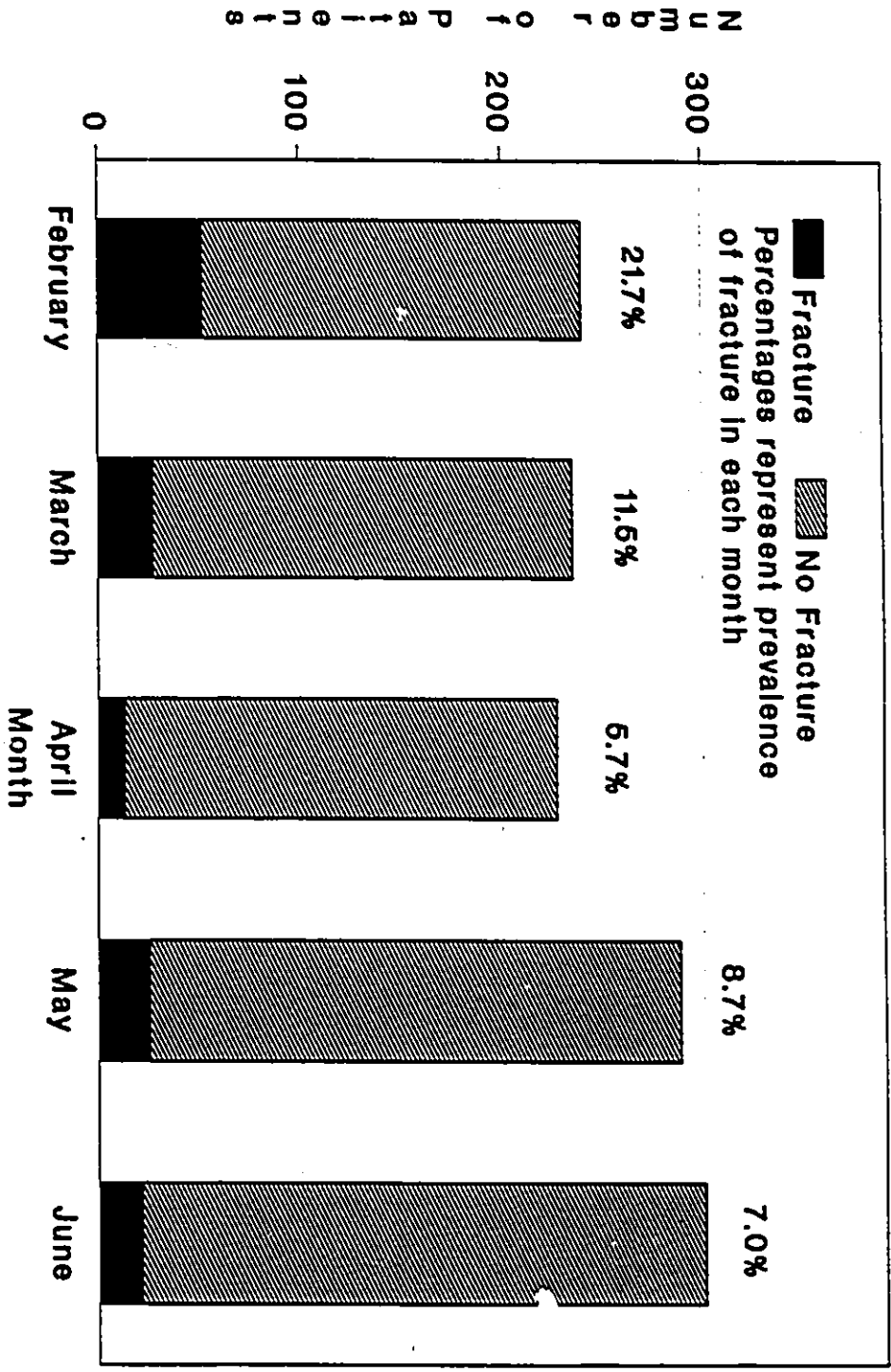


Figure 3. Distribution of patients and ankle series fracture by month in Phase 0

Table 7. Use of radiography in ankle injury patients seen over five month period at Ottawa Civic and Ottawa General Hospital emergency departments (Phase 0)

Radiography	OCH (N = 657)	OGH (N = 633)	Total (N = 1290)
Patients referred for x-ray (%)			
Ankle series	544 (82.8)	505 (79.8)	1049 (81.3)
Foot series	204 (31.1)	169 (26.7)	373 (28.9)
Either series	632 (96.2)	584 (92.3)	1216 (94.3)
Total x-ray series performed	748	674	1422
Mean per patient	1.14	1.07	1.10
X-rays positive for significant fracture (%)			
Ankle series	73 (13.4)	65 (12.9)	138 (13.2)
Foot series	30 (14.7)	19 (11.2)	49 (13.1)
Both series combined	103 (13.8)	84 (12.5)	187 (13.2)
Pattern of x-ray series ordered (%)			
None	25 (3.8)	49 (7.7)	74 (5.7)
Ankle-only	428 (65.1)	415 (65.6)	843 (65.3)
Foot only	88 (13.4)	79 (12.5)	167 (12.9)
Ankle and foot	116 (17.7)	90 (14.2)	206 (16.0)
Cost of x-ray series performed (\$)			
Ankle (@ \$33.30) *	14,252.40	13,819.50	28,071.90
Foot (@ \$33.30) *	2,930.40	2,630.70	5,561.10
Ankle and foot (@ \$55.40) *	6,426.40	4,986.00	11,412.40
Total	23,609.20	21,436.20	45,045.40
Mean time in ED (minutes)			
Overall		110.3	
X-ray		113.1	
No x-ray		77.3	
Difference		35.8	

* Estimated 1991 OHIP fee schedule, based on 1988 fee schedule x 1.08%

Table 8. Ankle injury patients seen and use of radiography by experience of attending physician in Phase 0

Radiography	Fulltime Emergency Physician	Parttime Emergency Physician	Resident Housestaff	Intern Housestaff
Number of patients seen	667	194	127	264
Patients referred for x-ray (%)				
Ankle series	81.2	81.4	84.3	78.4
Foot series	29.1	23.4	29.9	30.7
Either series	94.9	93.0	95.3	92.8
Mean series per patient	1.11	1.05	1.13	1.09
X-rays positive for fracture (%)				
Ankle series	13.2	11.4	15.9	14.0
Foot series	11.9	6.5	2.6	17.3
Both series combined	12.8	10.3	12.4	14.9
Pattern of series ordered (%)				
None	5.1	6.2	4.7	7.2
Ankle only	65.8	70.1	65.4	62.1
Foot only	13.0	12.4	11.8	14.8
Ankle and foot	16.0	11.3	18.1	15.9

7. RESULTS - PHASE I: DEVELOPMENT OF THE RULE

7.1 Descriptive Statistics

Seven hundred and fifty ankle injury patients were entered into Phase I of the study from July to December, 1990: 513 (68.4%) at OCH and 237 (31.6%) at OGH. Another 377 eligible patients, 97 (25.7%) at OCH and 280 (74.3%) at OGH, were not recruited into the study. The study patients were found to range in age from 18 to 92 years and to be evenly distributed by gender (Table 9). The great majority of injuries (84%) were of a twisting nature. The patients were found to have 70 (9.3%) significant malleolar fractures and 32 (4.3%) significant midfoot fractures. Another 5.7% experienced insignificant avulsion fractures. The mean number of x-ray series per patient was 1.2 and the yield of the x-rays for significant fracture was 10.2% for the ankle series and 13.9% for the foot series (Table 10).

There was a slightly lower prevalence of malleolar region fractures in Phase I (9.3%) than in Phase 0 (10.7%). Otherwise, the demographic and clinical data suggest that there were no important differences between the Phase 0 and Phase I patients (Tables 6 and 9). Hence, despite the difference in recruitment rates at OCH and OGH, the patients used to develop the rules in Phase I can be considered typical of ankle injury patients commonly seen at the two institutions .

Table 9. Characteristics of patients included in Phase I, development of the decision rules

Characteristic	N = 750	
Mean age in years (\pm SD) range	35.1 \pm 14.9	18 - 92
Male (%)	389 (51.9)	
Mechanism of injury (%) (N = 747)		
Twisting	627 (83.9)	
Direct blow	52 (7.0)	
Fall from a height	35 (4.7)	
Vehicle accident	15 (2.0)	
Other	18 (2.4)	
Significant fractures (%)	102 (13.6)	
Malleolar region (%)	70 (9.3)	N = 70
lateral malleolus		41 (58.6)
medial malleolus		7 (10.0)
posterior malleolus		2 (2.9)
bimalleolar		12 (17.1)
trimalleolar		8 (11.4)
talus		0
Midfoot (%)	32 (4.3)	N = 32
base 5th metatarsal		28 (87.5)
navicular		2 (6.3)
anterior process calcaneus		2 (6.3)
cuboid		0
cunieforms		0
Avulsion fractures (%) *	43 (5.7)	N = 43
lateral malleolus		18 (41.9)
talus		13 (30.2)
cuboid		7 (16.3)
navicular		5 (11.6)
anterior process calcaneus		5 (11.6)
medial malleolus		1 (2.3)

* Patients may have fractures in more than one location

Table 10. Use of radiography for 750 ankle injury patients in Phase I

Radiography	Number
Patients referred for x-ray (%)	
Ankle series	689 (91.9)
Foot series	230 (30.7)
Either series	750 (100)
Total x-ray series performed	919
Mean per patient	1.2
X-rays positive for significant fracture (%)	
Ankle series	70 (10.2)
Foot series	32 (13.9)
Both series combined	102 (11.1)

7.2 Physicians' Predictions and Attitudes

We found that physicians clearly expected a low yield from many of the x-rays that they requested. In 57.8% of cases where an ankle x-ray series was ordered ($N = 682$), the physician felt there was only a 0 to 0.1 probability of a significant fracture (Figure 4). Likewise, they expected a 0 to 0.1 probability of fracture in 47.3% of the foot x-ray series ordered. Furthermore, the physicians stated that they theoretically would have been either comfortable or very comfortable in ordering no x-rays for 45.9% of the 655 patients for whom the question was answered (Figure 5). There was, however, a low but important prevalence of fractures (2.0%) among patients for whom the physicians indicated that they would have been either very comfortable or comfortable in ordering no x-rays.

Table 11 shows that these experienced physician assessors would have missed 29% of fractures on the ankle series if the threshold for ordering an x-ray had been their own predicted probability for fracture of at least 0.5. Only with the threshold reduced to at least 0.1 probability for fracture would they have performed reasonably well, although even then they would have missed two of the fractures (Table 12).

Figure 6 shows the receiver operating characteristic (ROC) curve for the physicians' predicted probability as a model for detecting ankle series fractures. While the shape of the curve indicates good discrimination between the fracture and non-fracture cases, we can see that 100% sensitivity is not achieved until the specificity is 0%.

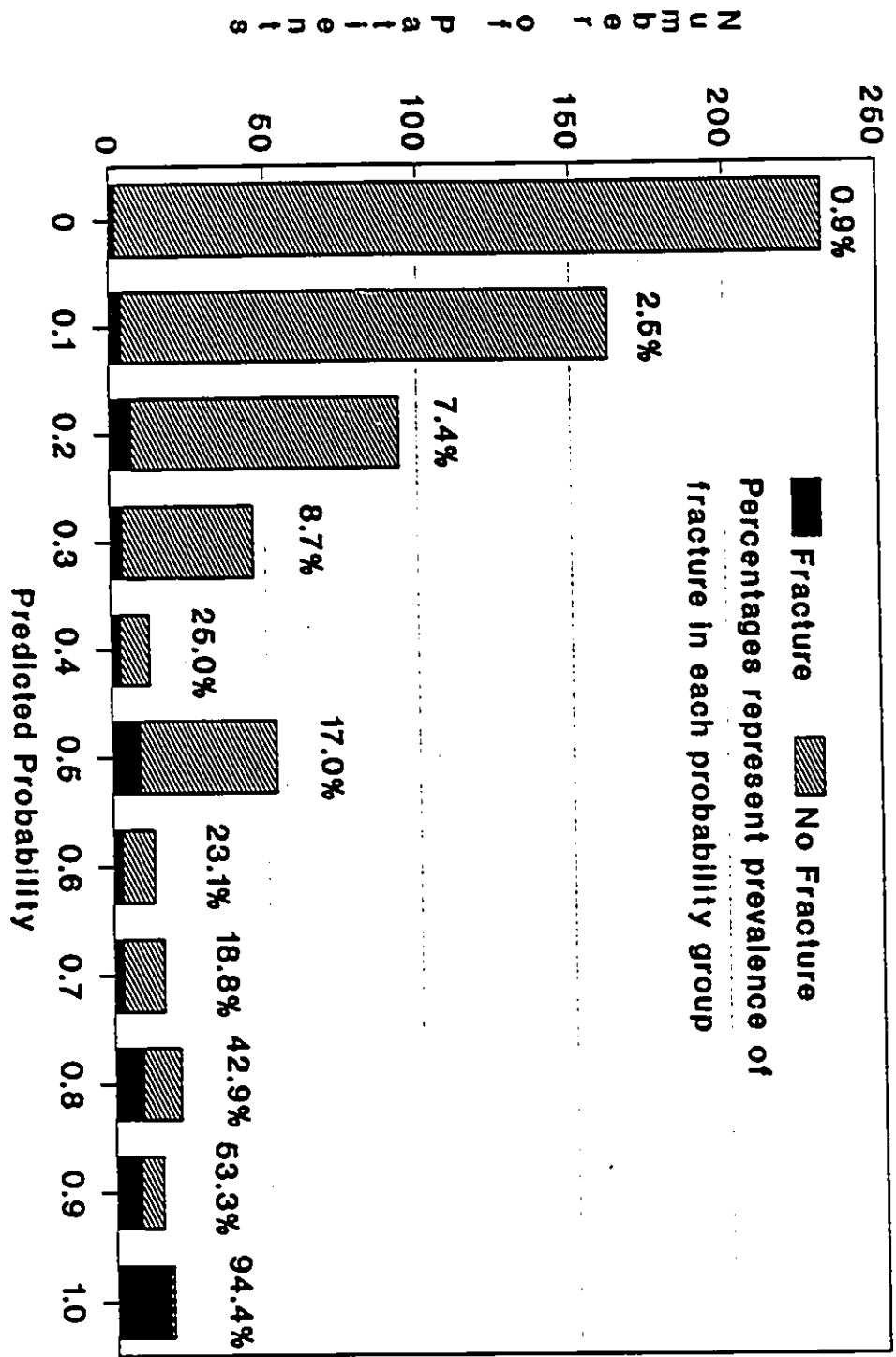


Figure 4. Distribution of patients and ankle series fractures by physicians' predicted probability of fracture

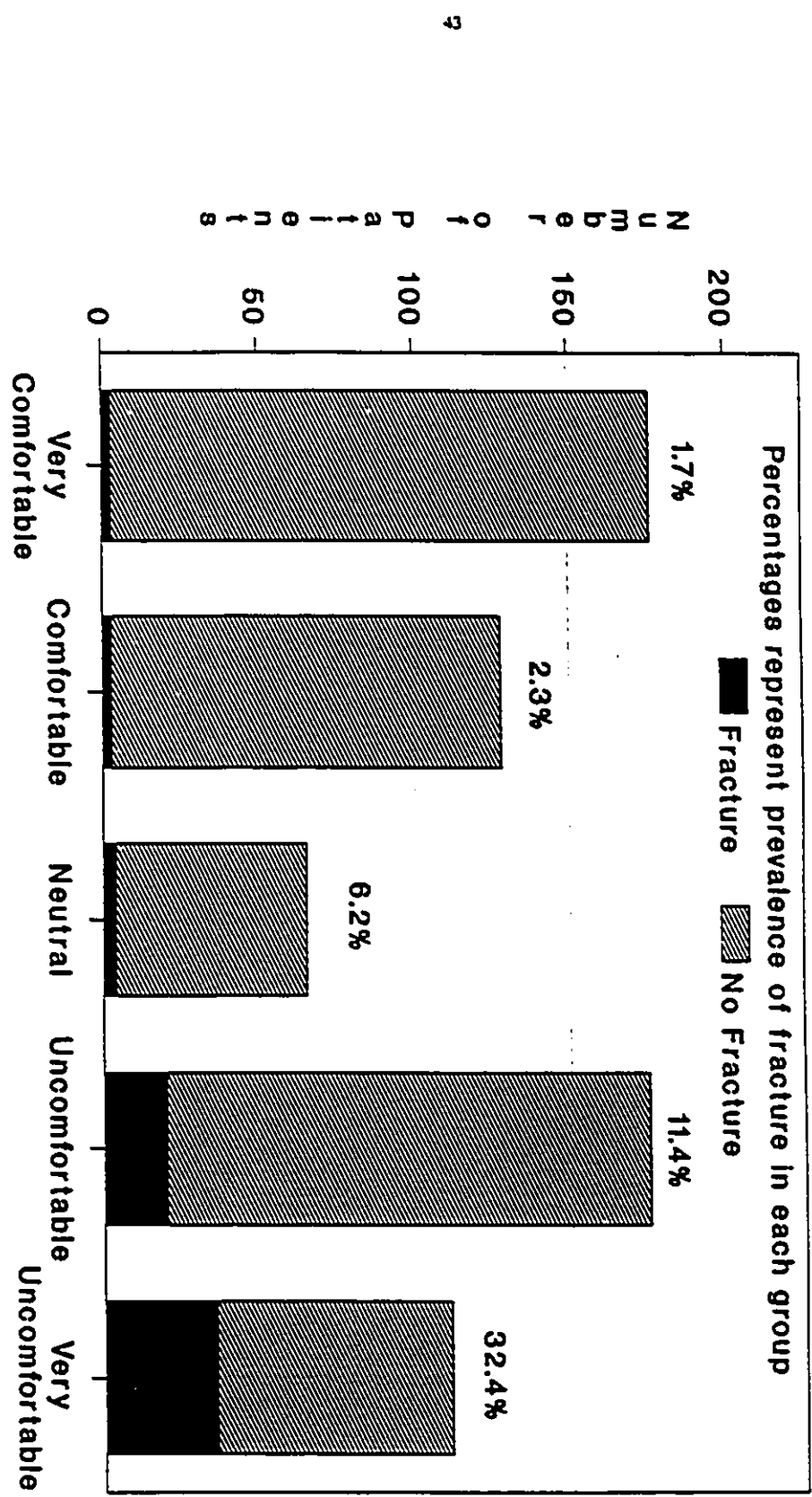


Figure 5. Distribution of patients and ankle series fractures by physicians' theoretical comfort with no x-ray

Table 11. Classification performance of physicians' predicted probability of ankle series fracture with 0.5 probability threshold for ordering x-rays

Predicted Fracture	Actual Fracture		
	Yes	No	
Yes	49	87	
No	20	526	
Total			682
Sensitivity			71.0%
Specificity			85.8%
Accuracy			84.3%

Table 12. Classification performance of physicians' predicted probability of ankle series fracture with 0.1 probability threshold for ordering x-rays

Predicted Fracture	Actual Fracture		
	Yes	No	
Yes	67	383	
No	2	230	
Total			682
Sensitivity			97.1%
Specificity			37.5%
Accuracy			43.5%

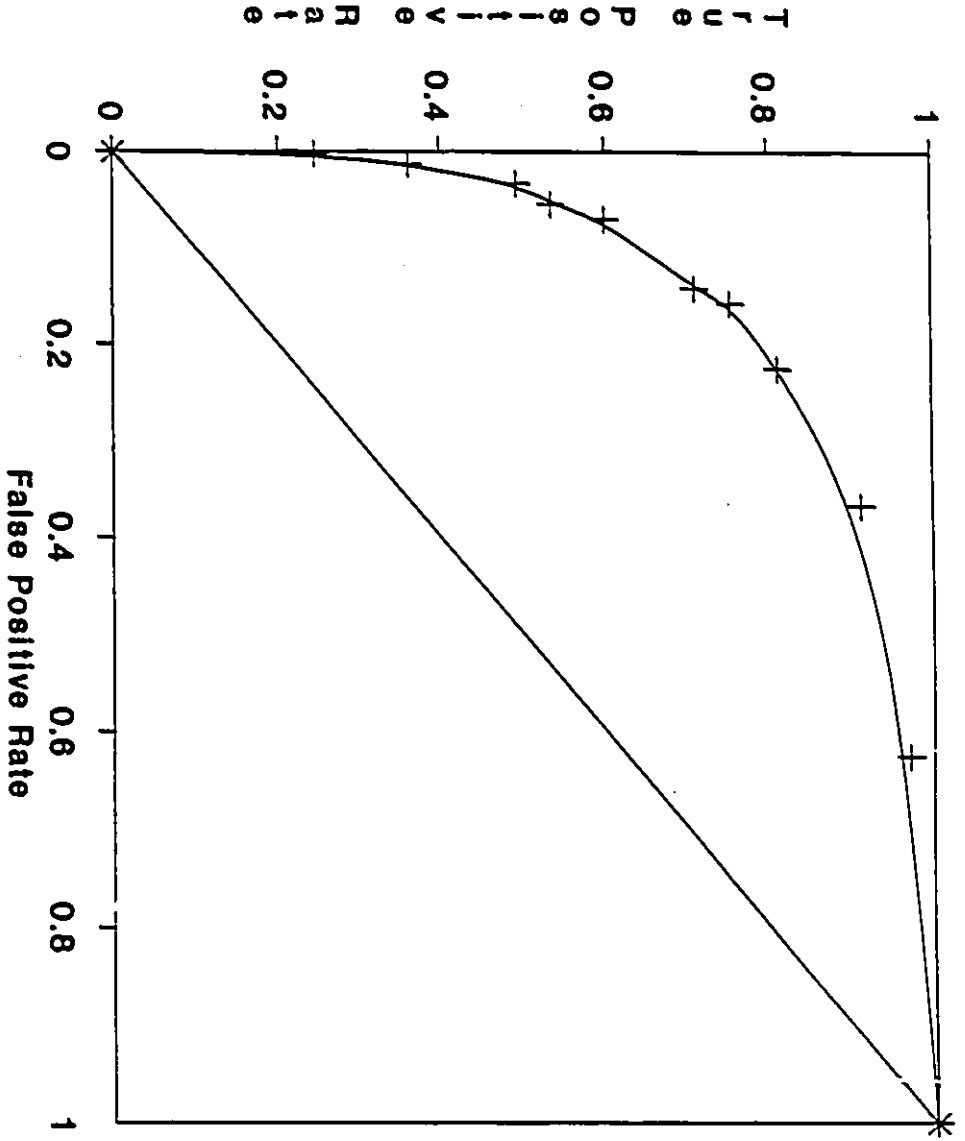


Figure 6. R O C curve for physicians' predicted probability of ankle series fracture

7.3 Interobserver Agreement

One hundred of the study patients were independently assessed by two physicians. The demographic data, type of injury, and incidence of fracture are given in Table 13 and suggest that, despite some differences, these patients were comparable to the larger group used to derive the decision rule (Table 9). There was a good spectrum of fracture and nonfracture cases.

General physical findings, with the exception of ability to bear weight, did not show good agreement (Figure 7). The reliability of judging restriction of range of motion was improved by dichotomizing the four-point scale but offered, at best, only fair agreement. Neither ecchymosis nor soft tissue tenderness at any of four areas were found to have substantial agreement and the anterior drawer sign fared poorly. In contrast, ability to bear weight in the emergency department, when dichotomized, proved to be the most reliable variable, with a kappa value of 0.83, reflecting nearly perfect agreement.

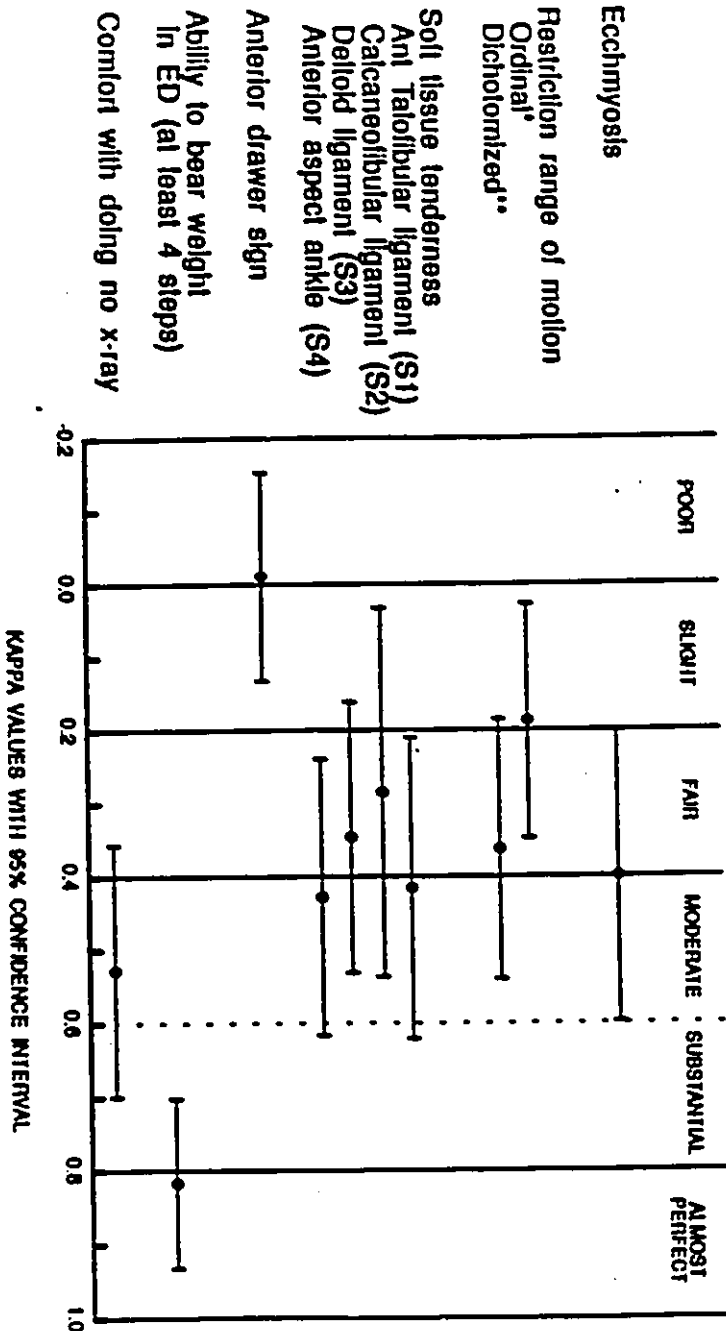
Judgement of the degree of swelling at four locations, on an ordinal scale, generally demonstrated only slight to moderate agreement (Figure 8). Dichotomizing the scale improved results significantly for swelling over the lateral and medial malleolus.

Other than ability to bear weight, the best agreement (kappa values greater than 0.6) was found for localized bone tenderness (Figure 9). Substantial interobserver agreement was obtained for tenderness at the base of the fifth metatarsal, the anterior and posterior edges of the lateral malleolus, and the inferior tip of the medial malleolus. Furthermore, the best agreement was seen when bone areas were combined, as for the posterior edge and inferior tip of each of the malleoli.

The data used to calculate the interobserver agreement for the various findings are given in Appendix D.

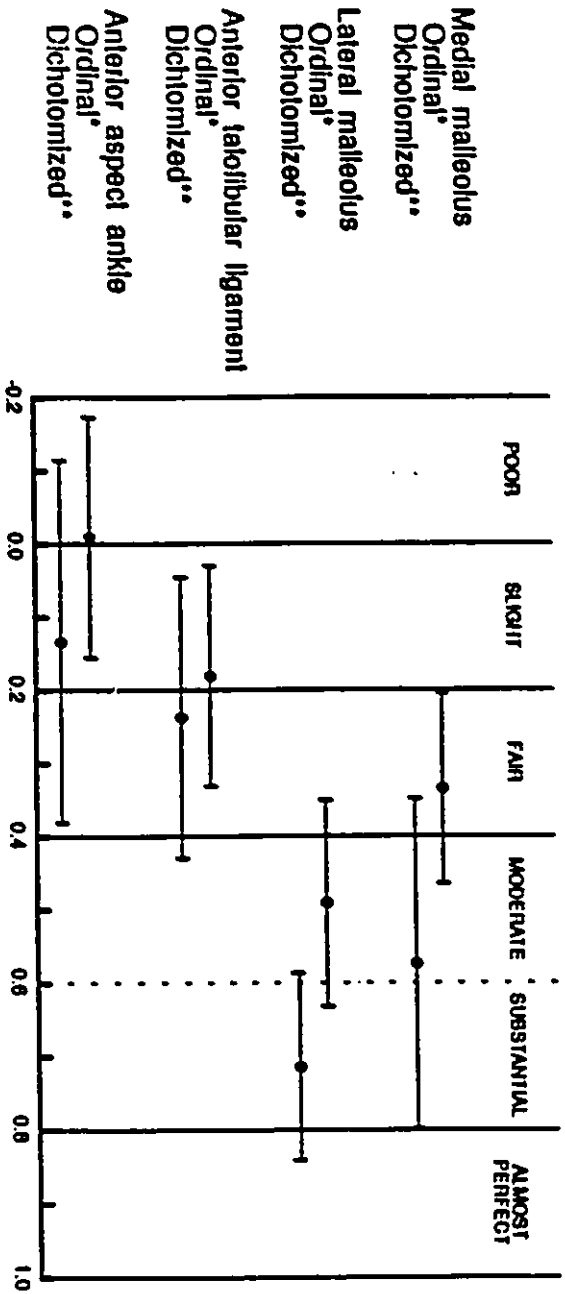
Table 13. Characteristics of patients assessed for interobserver agreement in Phase I

Characteristic	N = 100
Mean age in years (\pm SD) range	37.3 \pm 15.1 18 - 83
Male	51
Mechanism of injury	
twisting	91
direct blow	4
fall from a height	2
vehicle accident	2
other	1
Fractures	18
Malleolar region	17
Midfoot	1



*Ordinal: none/minimal/moderate/marked
 **Dichotomized: none/minimal/moderate/marked

Figure 7. Interobserver agreement (kappa values) of general physical findings in 100 ankle injury patients



*Ordinal: none/minimal/moderate/marked
 **Dichotomized: none-minimal/moderate-marked

Figure 8. Interobserver agreement (kappa values) of swelling in 100 ankle injury patients

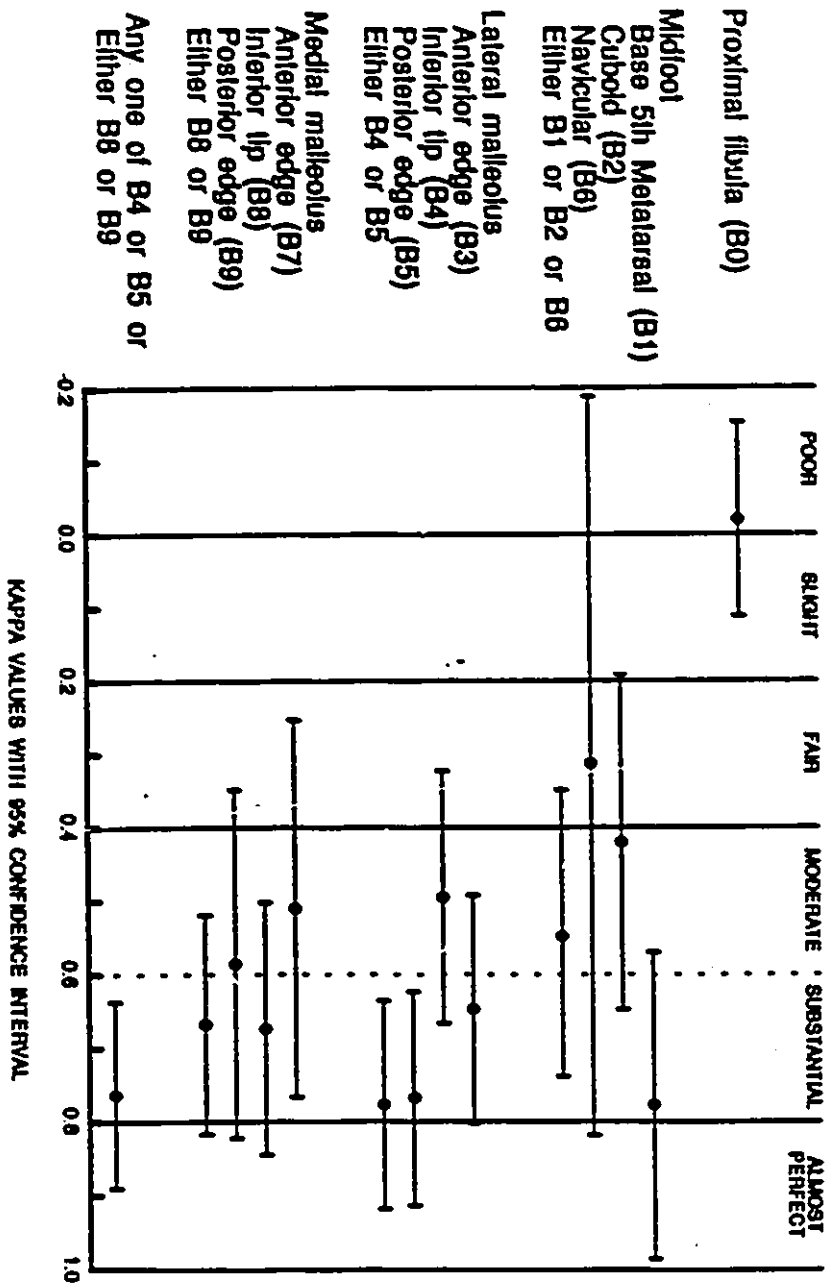


Figure 9. Interobserver agreement (kappa values) of bone tenderness in 100 ankle injury patients

7.4 Derivation of Ankle X-ray Series Rule

7.4.1 Univariate Analysis

Table 14 shows that 14 nominal variables were strongly associated ($P < 0.001$) and one was weakly associated ($P < 0.05$) with significant fracture of the ankle x-ray series. These included several ordinal variables that had been dichotomized for the analysis: ability to bear weight immediately, ability to continue current activity, range of motion, and degree of swelling (in four areas).

Two continuous variables based on the history, age and time from injury, were also associated with fracture. Figure 10 demonstrates the strong association between increased age and prevalence of fracture, as noted in Phase 0. These two continuous variables were dichotomized at various cutpoints and their association tested for significance by chi-square analysis. The cutpoints yielding the most significant associations were used for the multivariate analysis.

Two other variables based on history also had sufficient significance ($P < 0.05$) to be used in the multivariate analysis: ability to bear weight immediately and ability to continue current activity. We assumed that variables based on the history would have adequate interobserver reliability for further use in model development.

Association of ordinal data with ankle series fracture was generally high (Table 15). However, poor reliability restricted the use of these data to the dichotomized form. This was done by reducing a four point scale (none / minimal / moderate / marked) to a two point scale (none-minimal / moderate-marked).

Figures 11 to 14 show the relative risk, with 95% confidence intervals, for significant ankle series fracture for each of the predictor variables.³⁰

Table 14. Univariate association of continuous and nominal variables with significant fracture on ankle x-ray series

Variable	Significant Fracture (N = 70)	Other Cases (N = 619)
History (%)		
Mean age (\pm SD)	43.1 \pm 18.4	34.1 \pm 14.1*
Female	61	46
Mean time in hours from injury (\pm SD)	9.1 \pm 13.3	20.7 \pm 33.7*
"Cracking" sound heard	29	24
Twisting mechanism	19	15
Ankle fracture previous 12 months	2	1
Unable to bear weight immediately#	62	21****
Able to continue current activity#	3	25****
General findings (%)		
Ecchymosis	32	22
Range of motion moderately limited#	73	37****
Anterior drawer sign	14	2****
Unable to bear weight in ED (four steps)	80	30****
Moderate-marked swelling (%)		
Medial malleolus#	40	10****
Lateral malleolus#	84	42****
Anterior talofibular ligament#	56	44
Anterior aspect of ankle#	29	11****
Soft tissue tenderness (%)		
Anterior talofibular ligament	69	80
Calcaneofibular ligament	19	17
Deltoid ligament	25	20
Anterior aspect of ankle	31	26
Bone tenderness (%)		
Proximal fibula	7	2**
Lateral malleolus -anterior edge	90	50****
-inferior tip	62	39****
-posterior edge	69	26****
Medial malleolus -anterior edge	32	7****
-inferior tip	38	13****
-posterior edge	34	5****

* P < 0.001 by t-test

** P < 0.05, *** P < 0.001, **** P < 0.0001 by continuity adjusted chi-square

Dichotomized from original ordinal data

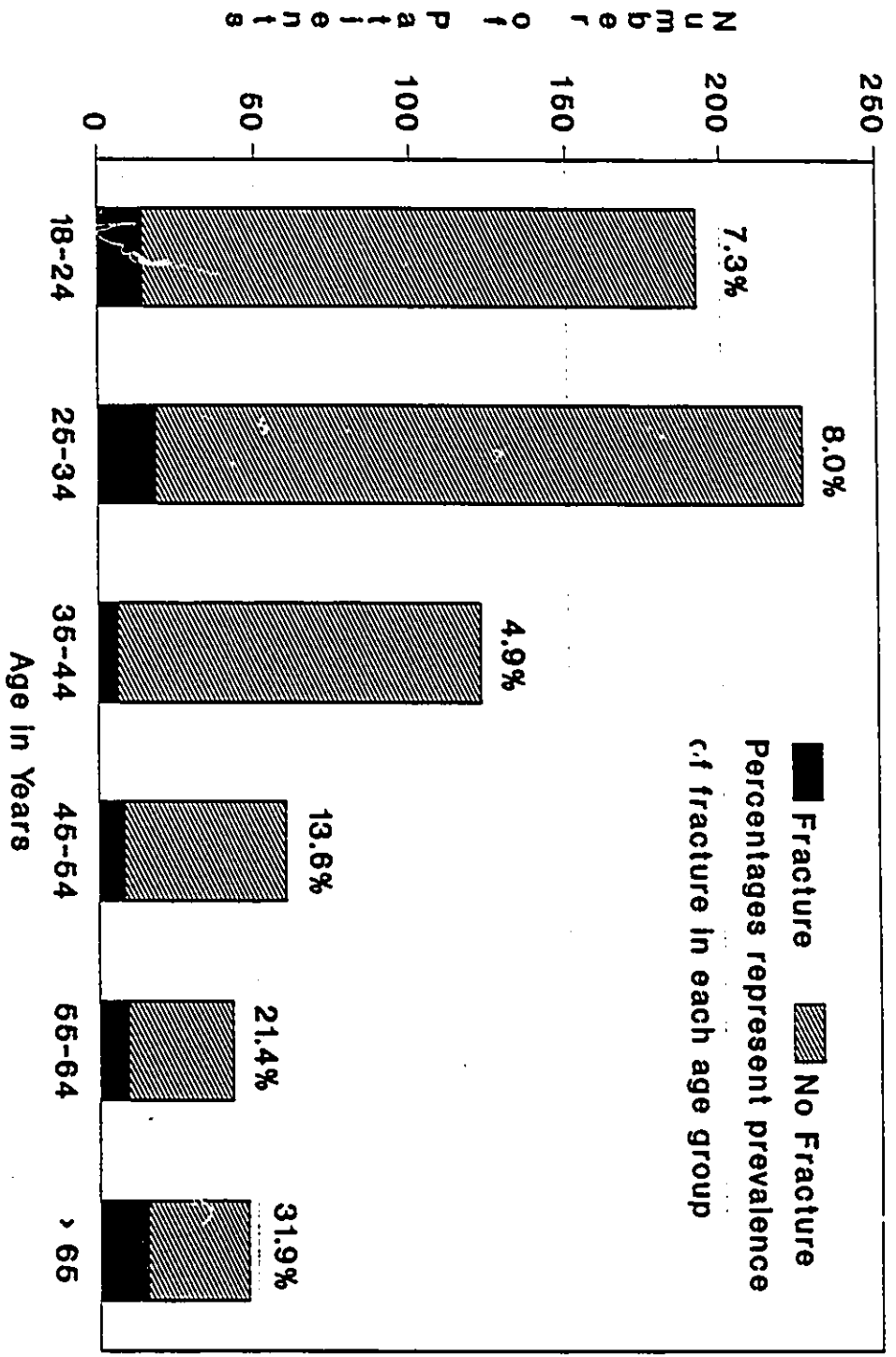


Figure 10. Distribution of patients and ankle series fractures by age group in Phase I

Table 15. Univariate association of ordinal variables with significant fracture on ankle x-ray series

Variable*	P Value**
Ability to bear weight immediately (3)	<0.0001
Range of motion (4)	<0.0001
Swelling (4)	
medial malleolus	<0.0001
lateral malleolus	<0.0001
anterior talofibular ligament	0.01
anterior aspect of ankle	0.0001
Difficulty bearing weight in ED (5)	<0.0001
Predicted probability of fracture (11)	<0.0001
Theoretical comfort with no x-ray (5)	<0.0001

* No. of points on ordinal scale

** Mann-Whitney U test

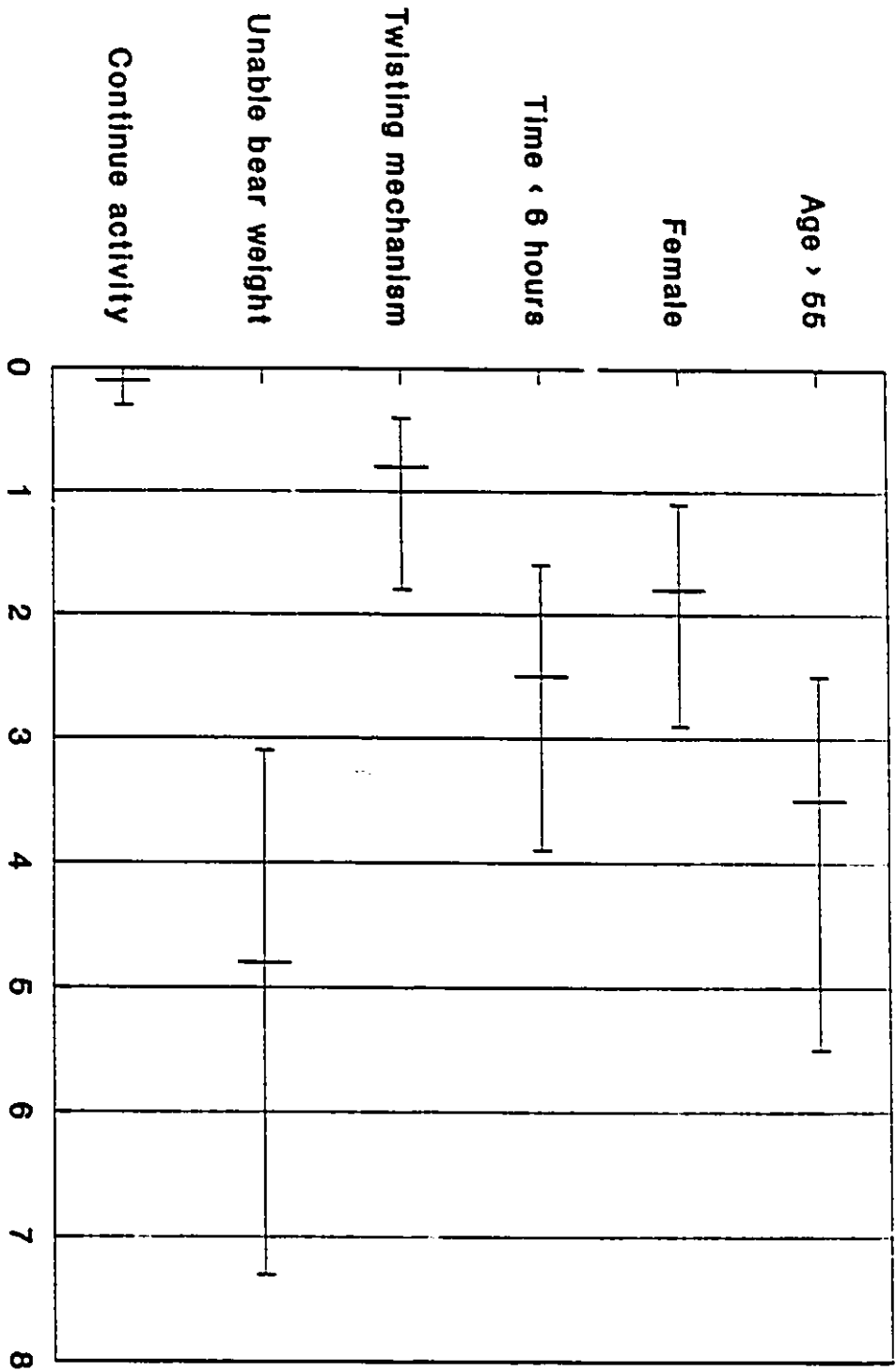


Figure 11. Relative risk with 95% confidence intervals of ankle series fracture for variables from history

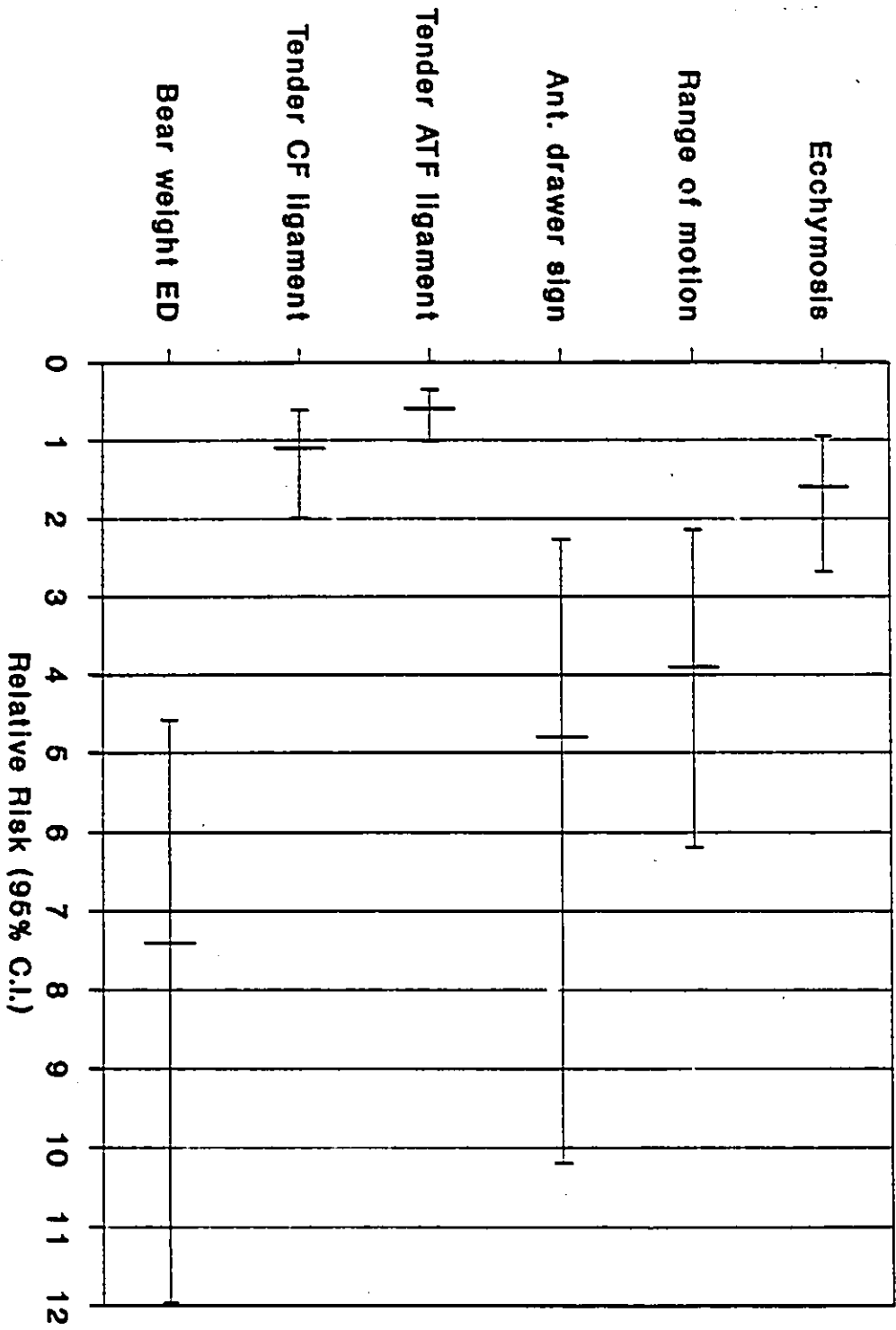


Figure 12. Relative risk with 95% confidence intervals of ankle series fracture for general findings

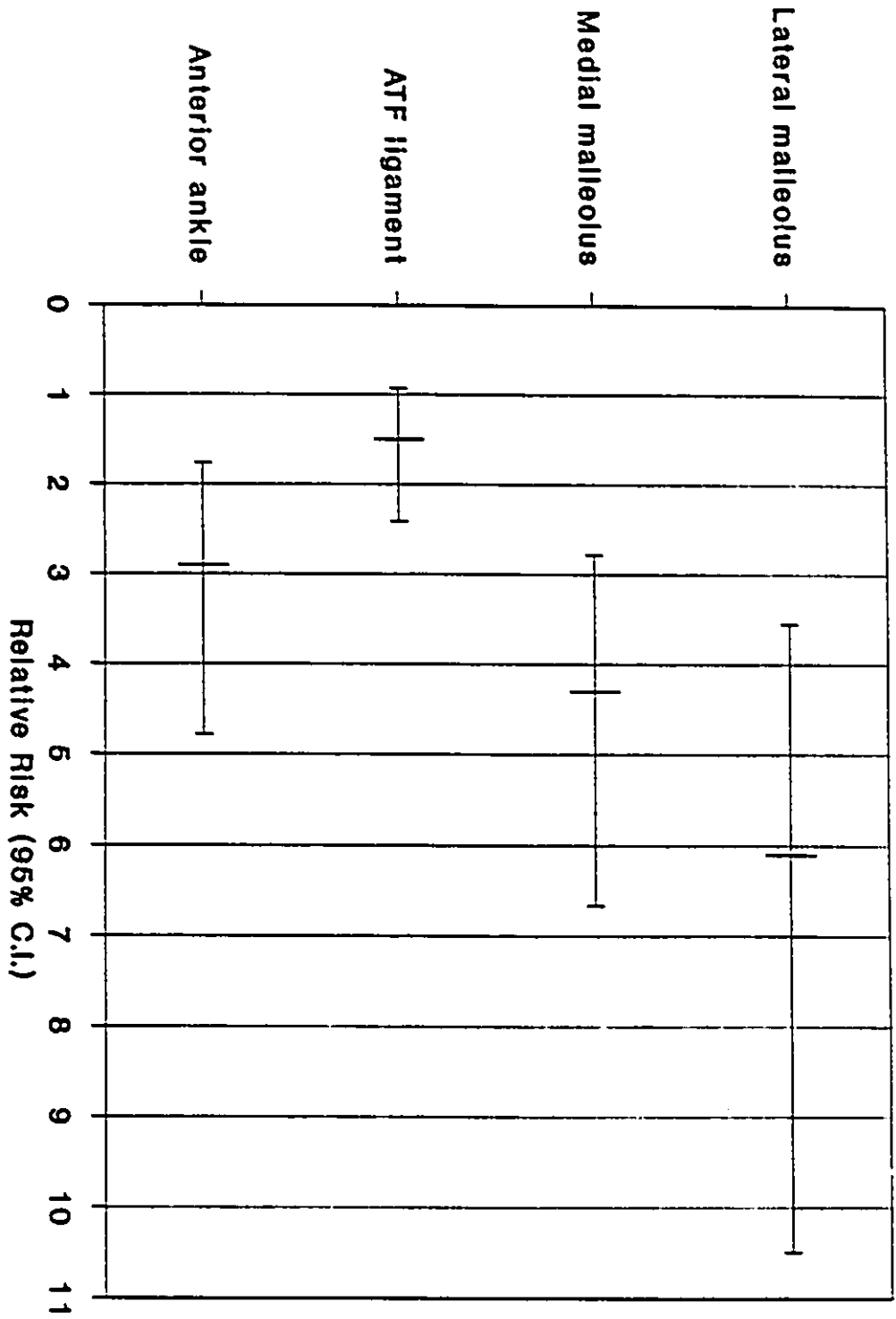


Figure 13. Relative risk with 95% confidence intervals of ankle series fracture for location of swelling

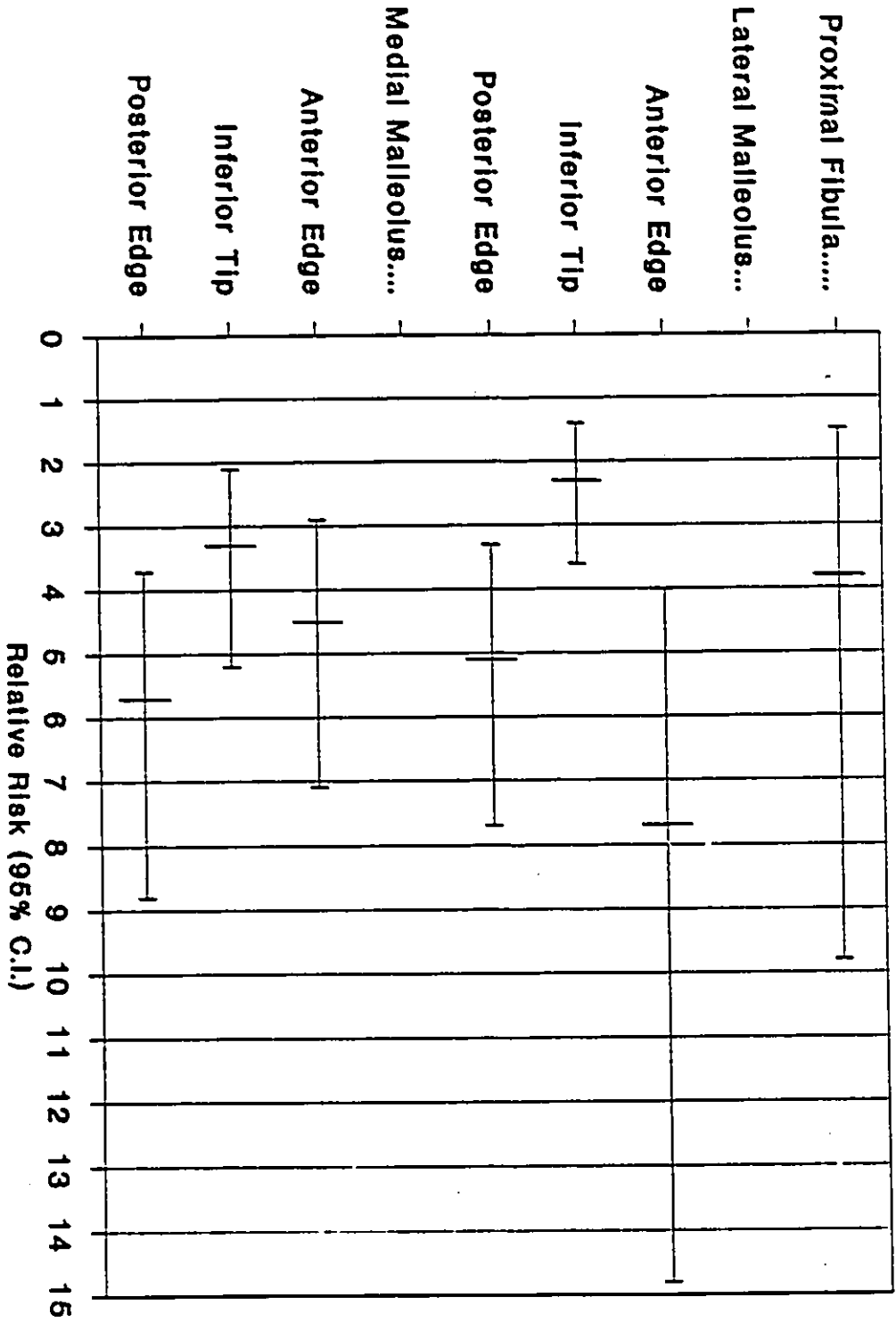


Figure 14. Relative risk with 95% confidence intervals of ankle series fracture for areas of bone tenderness

7.4.2 Multivariate Analysis

Those variables based on physical examination were used for multivariate analysis if significantly associated with ankle x-ray series fracture ($P < 0.05$) and if their interobserver agreement was high (kappa greater than 0.6). The following variables, while statistically significant, did not show sufficient interobserver agreement to be included in the model: ecchymosis, range of motion, anterior drawer sign, swelling of medial malleolus, swelling of anterior aspect ankle, and bone tenderness of either proximal fibula, inferior tip lateral malleolus, anterior edge medial malleolus, or posterior edge of medial malleolus.

We found that several combinations of variables had acceptable reliability (Figure 9) and were highly associated with fracture and were therefore entered for further analysis: inability to bear weight both immediately and in emergency department, swelling of lateral malleolus in first 6 hours, bone tenderness of either inferior tip or posterior edge of lateral malleolus, and bone tenderness of either inferior tip or posterior edge of medial malleolus. Table 16 shows the variables used in the multivariate analyses as well as their chi-square values (with one degree of freedom), which were the basis for the recursive partitioning process.

Table 16. Variables used in multivariate analyses

Variable	Chi-square*	Variable	Chi-square*
Age		Bone tenderness	
45 or greater	25.5	lateral malleolus	
55 or greater	29.5	-anterior edge	37.9
65 or greater	23.6	-posterior edge	51.1
		-inferior or posterior	39.2
Swelling		medial malleolus	
lateral malleolus	42.7	-inferior tip	26.6
lateral malleolus < 6 hr	47.9	-inferior or posterior	34.5
Inability to bear weight			
immediately	51.5		
continue current activity	16.1		
in ED	65.7		
both immediately and ED	68.3		

* Continuity adjusted chi-square with one degree of freedom

7.4.3 Logistic Regression Analysis

Logistic regression (LR) analysis used 674 of a possible 689 cases due to missing data in 15 cases: age in one case and bone tenderness in 14 cases. Stepwise forward selection yielded a model consisting of six variables (Table 17). The odds ratios for each variable, adjusted for the other variables, are given along with the 95% confidence intervals. One can calculate the probability of having a fracture for a patient with a specific set of predictor variables. For example, a patient negative for all of the variables would have a probability of significant fracture of $1 / (1 + e^{5.3}) = 0.005$. Alternately, a patient who was positive for all of the variables would have a fracture probability of $1 / (1 + e^{-2.0}) = 0.881$.

A decision rule using this model would have correctly classified 92% of patients yet would have missed 68% of the fractures on the ankle x-ray series, based on the preset predicted fracture probability of 0.5 (Table 18). Clearly, a sensitivity of only 32.4% would be unacceptable. Inspection of a listing of cases revealed that the sensitivity could

have been improved to 100% only by reducing the threshold for ordering an x-ray to a 0.015 predicted probability of fracture (Table 19). We felt that such a model was unsatisfactory due to the low specificity (29.0%) and the complexity of incorporating such a large number of variables (six) into a decision rule.

7.4.4 Recursive Partitioning Analysis

Chi-square recursive partitioning (RP) techniques successfully yielded a model (Figure 15) that was 100% sensitive to ankle series fractures with fewer variables (four) than the LR model. The final variables in this model were: 1) age 55 or greater, 2) inability to bear weight both immediately and four steps in the emergency department, 3) bone tenderness at posterior edge (distal 6 cm) or inferior tip of the lateral malleolus, and 4) bone tenderness at posterior edge or inferior tip of the medial malleolus. The final variables have been grouped into a clinical decision rule that would permit physicians to categorize ankle injury patients into high and low risk groups for fracture (Figure 16).

This model was derived by choosing the partitions based on the variable with the highest chi-square value at each step. Other models could be developed by manipulating the order of the partitioning process, and usually involved the variables swelling over the lateral malleolus or bone tenderness at the anterior edge of the lateral malleolus. None of these other solutions to the problem of achieving 100% sensitivity offered comparable specificity or as few variables as the original model.

Table 17. Variables in logistic regression model

Variable	Coefficient	Significance	Odds Ratio	95% C.I.
Age 55 or greater	1.8	<0.0001	6.18	2.9 - 12.5
Unable to bear weight immediately & in ED	1.3	<0.0001	3.78	2.0 - 6.9
Swelling lateral malleolus < 6 hours	0.7	0.04	1.93	1.1 - 3.8
Bone tenderness				
lateral malleolus anterior	1.2	0.008	3.40	1.3 - 8.2
lateral inferior/posterior	1.2	0.004	3.45	1.4 - 7.7
medial inferior/posterior	1.1	0.0006	3.07	1.6 - 5.7
Constant	-5.3	<0.0001		

Table 18. Classification performance of logistic regression model using predicted fracture probability of 0.5 as threshold for ordering an ankle x-ray series

Predicted Fracture	Actual Fracture		
	Yes	No	
Yes	22	7	
No	46	599	
Total			674
Sensitivity			32.4%
Specificity			98.8%
Accuracy			92.1%

Table 19. Classification performance of logistic regression model using predicted fracture probability of 0.015 as threshold for ordering an ankle x-ray series

Predicted Fracture	Actual Fracture		
	Yes	No	
Yes	68	430	
No	0	176	
Total			674
Sensitivity			100%
Specificity			29.0%
Accuracy			36.2%

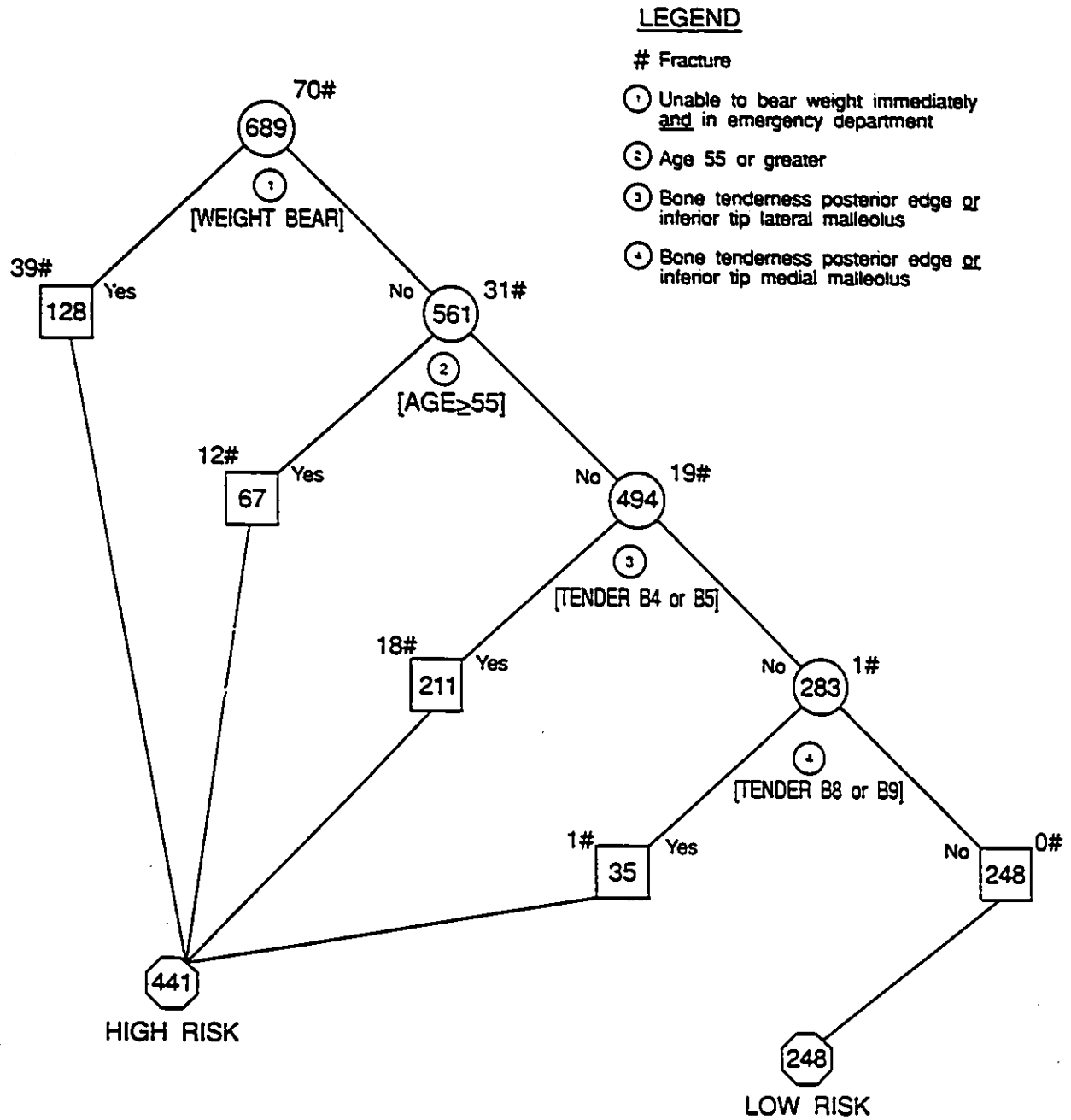


Figure 15. Recursive partitioning of predictor variables for ankle series fracture

**1) AN ANKLE X-RAY SERIES IS ONLY NECESSARY IF:
THERE IS PAIN NEAR THE MALLEOLI AND ONE OR MORE OF THESE
FINDINGS:**

- a) AGE 55 OR GREATER
- b) UNABLE TO BEAR WEIGHT
(both immediately and in ED - 4 steps)
- c) BONE TENDERNESS
(at the posterior edge or tip
of either malleolus)

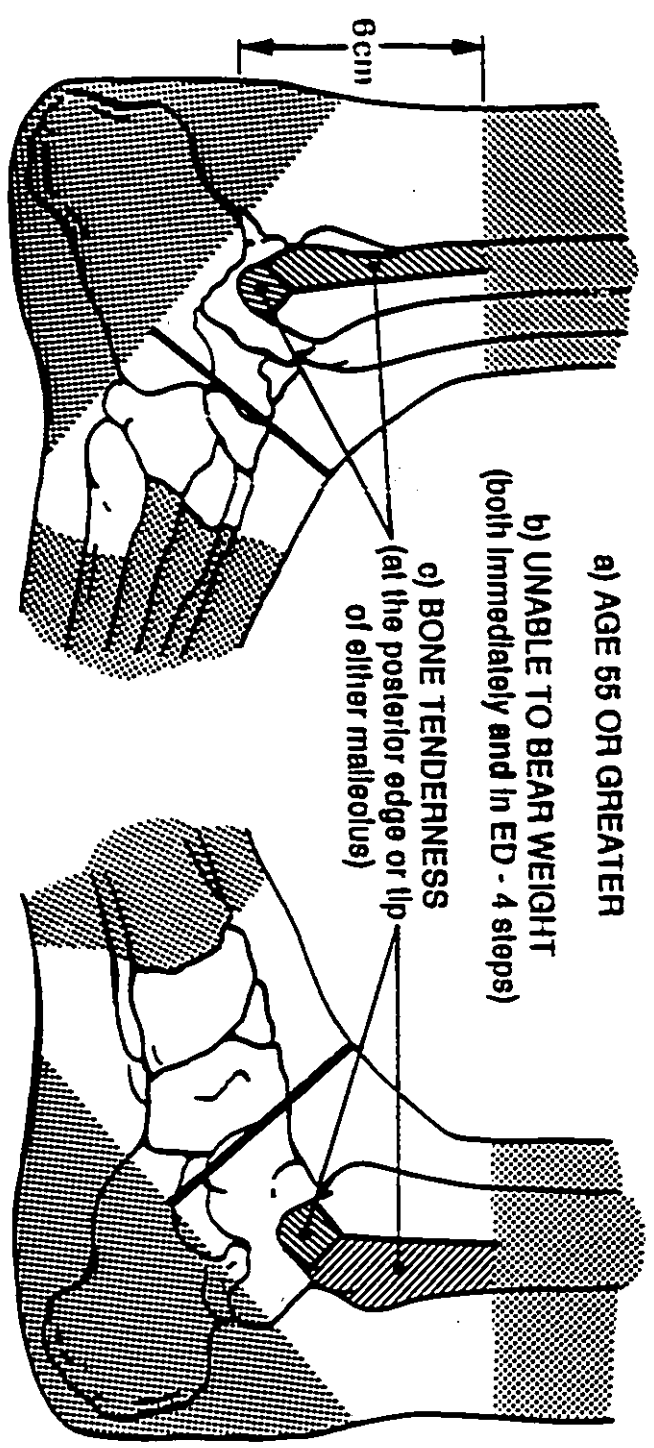


Figure 16. Clinical decision rule for ankle x-ray series in adults

7.5 Classification Performance

The classification performance of the ankle series decision rule as derived by recursive partitioning is shown in Table 20. We see that the goal of 100% sensitivity was achieved at a higher specificity (40.1%) than that of the logistic regression model but at a lower overall accuracy (46.2%). Much better accuracy could be achieved but only by reducing the sensitivity below our required level of 100%. An ROC curve was generated for the RP model by forcing the constituent variables into a logistic regression analysis and calculating sensitivity and specificity for the various fracture probabilities. An ROC curve was also developed for the LR model and was found to be similar to that of the RP rule except for achieving 100% sensitivity at a lower specificity level (Figure 17).

Table 20. Classification performance of clinical decision rule for ankle x-ray series

Predicted Fracture	Actual Fracture		
	Yes	No	
Yes	70	371	
No	0	248	
Total			689
Sensitivity			100%
Specificity			40.1%
Accuracy			46.2%
Negative predictive value			100%
Positive predictive value			15.9%

Similarly, the RP rule can be compared to the classification performance of the physicians' predicted probability, as shown in Tables 10, 11, and 19. Figure 18 compares the ROC curves of the RP Rule and the physicians' predicted probability: the physicians have better overall discrimination but fail to achieve 100% sensitivity (true positive rate of 1.0).

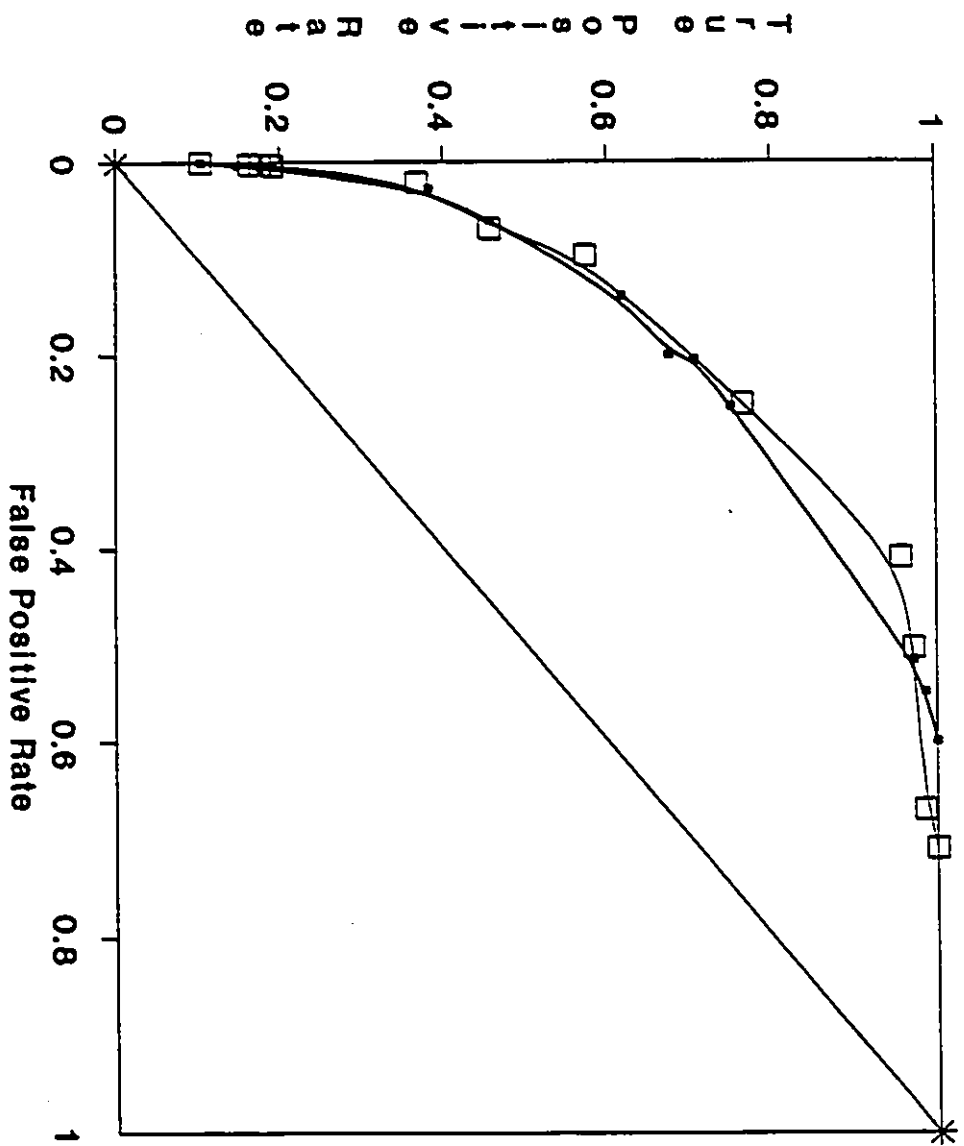


Figure 17. R O C curves comparing models derived by logistic regression and recursive partitioning

—●— RP Model
-□- LR Model

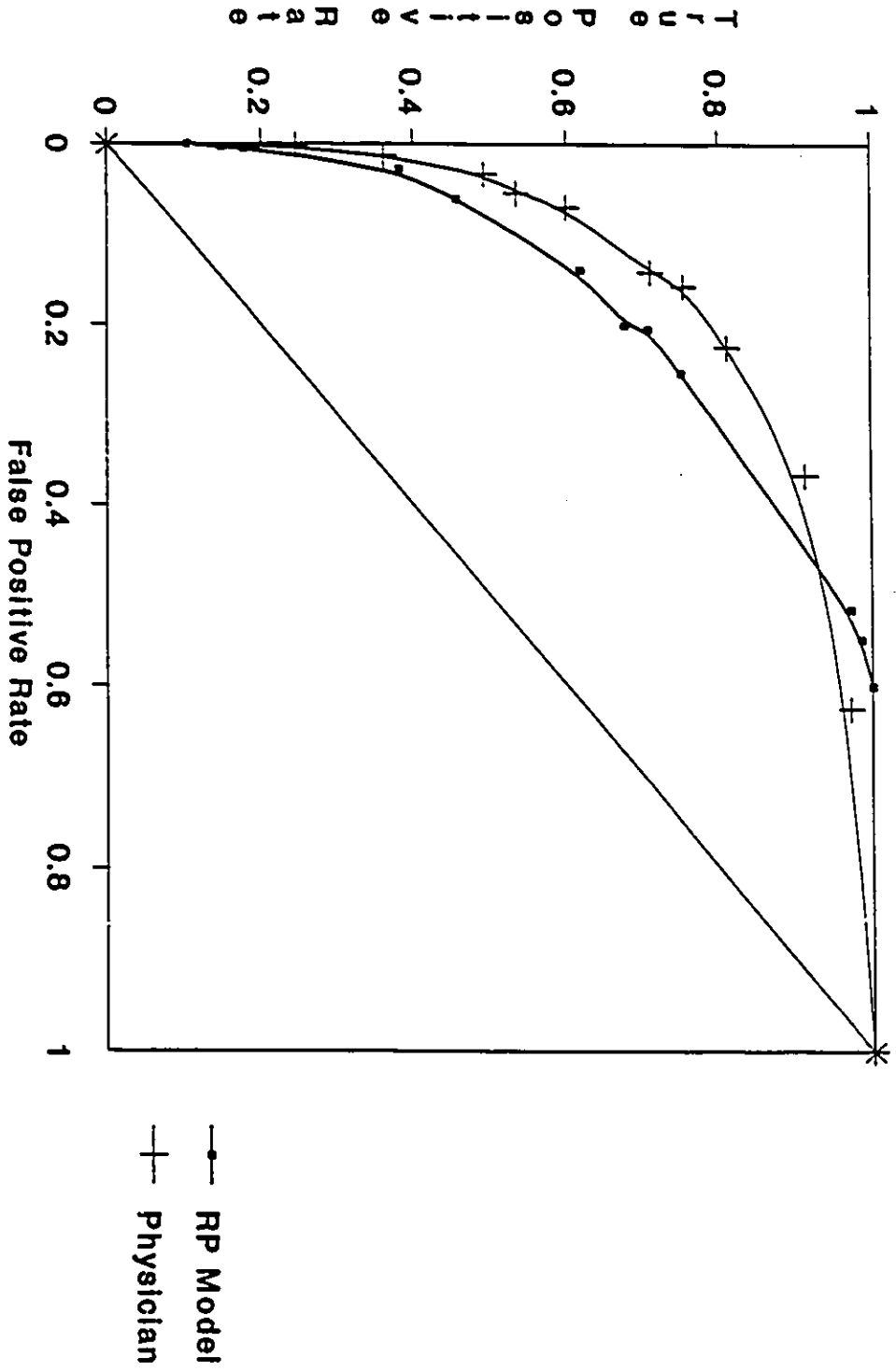


Figure 18. R O C curves comparing RP model and physicians' predicted probability for ankle series fracture

7.6 Derivation of Foot X-ray Series Rule

7.6.1 Derivation

A similar process determined the best predictor variables for fracture on a foot x-ray series (Table 21 and Figure 19). Neither age nor ability to bear weight were significantly associated with fractures, the majority of which were at the base of the fifth metatarsal. The only variable that had high interobserver agreement (kappa greater than 0.6) and was significantly associated with fracture was bone tenderness at the base of fifth metatarsal. However, using this variable alone would have led to a model that was only 93.8% sensitive. Recursive partitioning could only achieve a 100% sensitive model when bone tenderness at any one of the base of the fifth metatarsal, the cuboid, or the navicular were included. This was somewhat below our ideal standard for reliability as the kappa value for this combined variable was 0.53. Nevertheless, we felt that 100% sensitivity was paramount and that a small sacrifice in reliability was acceptable. No attempt was made to use logistic regression analysis for the foot series rule.

The final decision rule for foot x-ray series is presented in Figure 20. The classification performance for this rule (Table 22) shows a sensitivity of 100% achieved at a specificity of 24.7%.

Table 21. Univariate association of variables with significant midfoot fracture on foot x-ray series

Variable (%)	Significant Fracture (N = 32)	Other Cases (N = 198)
Mean age (\pm SD)	36.7 \pm 14.5	40.1 \pm 16.9
Unable to bear weight immediately	29	18
Ecchymosis	55	27*
Bone tenderness		
base 5th metatarsal	94	34**
cuboid	47	58
navicular	3	10
any one of above	100	75*
Unable to bear weight in ED	38	33

* P < 0.01, ** P < 0.0001 by continuity adjusted chi-square

Table 22. Classification performance of clinical decision rule for foot x-ray series

Predicted Fracture	Actual Fracture		
	Yes	No	
Yes	32	149	
No	0	49	
Total			230
Sensitivity			100%
Specificity			24.7%
Accuracy			35.2%
Negative predictive value			100%
Positive predictive value			17.7%

r

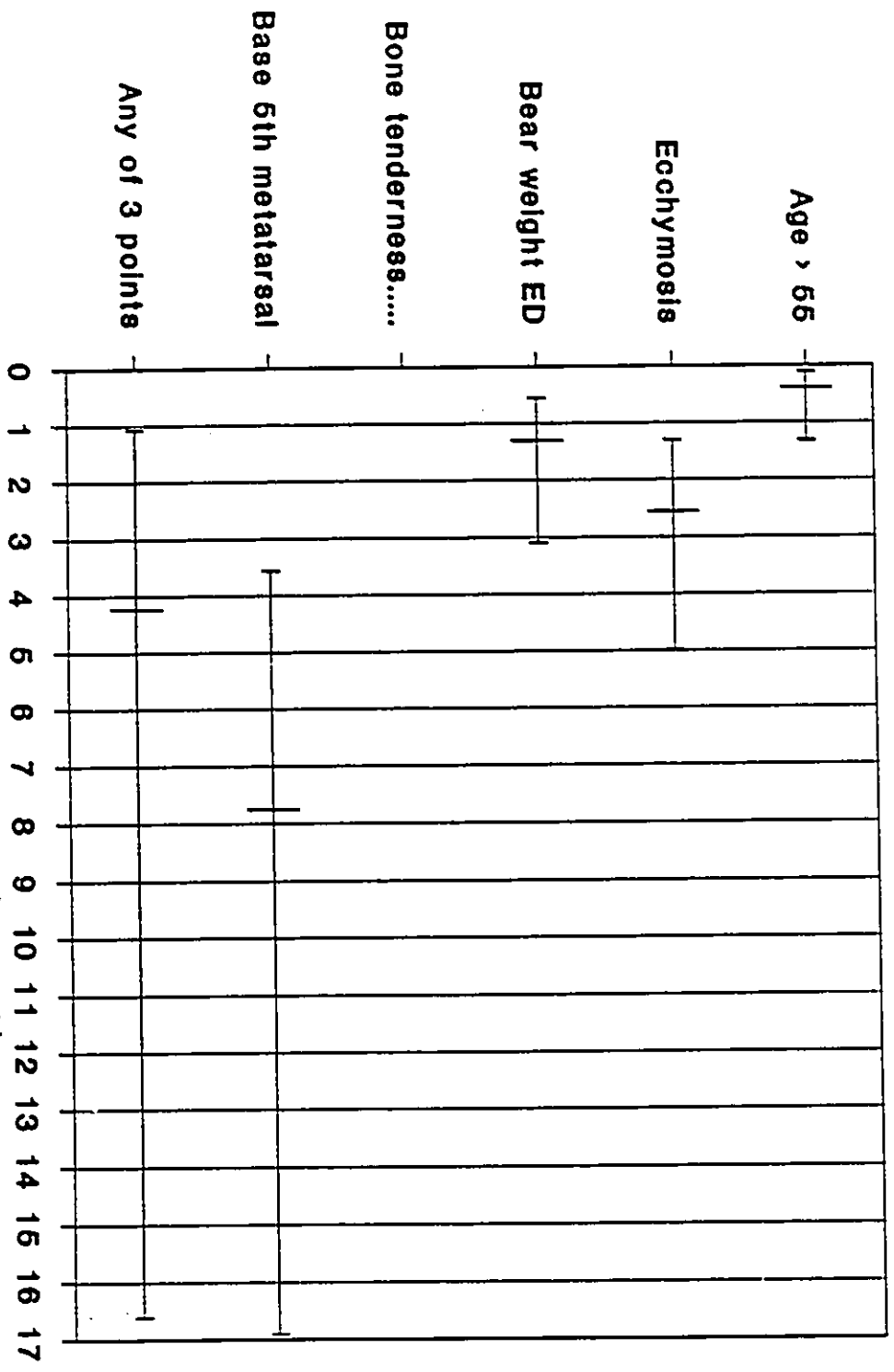
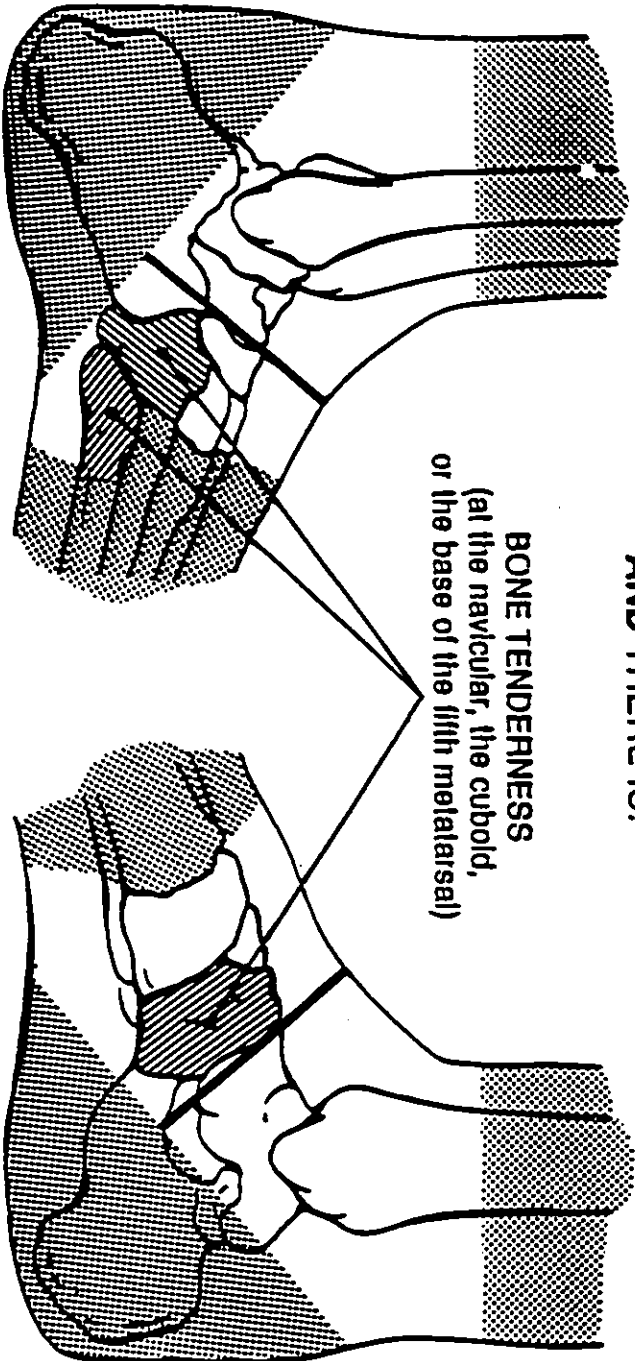


Figure 19. Relative risk with 95% confidence intervals for predictor variables of foot series fracture

2) A FOOT X-RAY SERIES IS ONLY NECESSARY IF:

**THERE IS PAIN IN THE MIDFOOT
AND THERE IS:**

BONE TENDERNESS
(at the navicular, the cuboid,
or the base of the fifth metatarsal)



LATERAL VIEW

MEDIAL VIEW

Figure 20. Clinical decision rule for foot x-ray series in adults

7.6.2 Classification Performance

Table 23 illustrates the classification performance of the two rules combined for the 919 x-ray series done in the study. None of the 102 fractures would have been missed, reaffirming our goal of 100% sensitive decision rules.

Table 23. Classification performance of ankle series and foot series decision rules combined

	Actual Fracture		
	Yes	No	
Predicted Fracture			
Yes	102	520	
No	0	297	
Total			919
Sensitivity			100%
Specificity			36.4%
Accuracy			43.4%
Negative predictive value			100%
Positive predictive value			16.4%

Table 24 compares the use of radiography in this study with what could potentially be achieved if the two decision rules had been invoked. The potential savings of x-ray series would be: ankle series 36.0%, foot series 21.3%, and both series combined 32.3%. 24.7% of the patients would require no radiography and only 7.5% would require both an ankle and a foot series. These potential reductions in use of radiography would represent some 30% less dollars spent on x-ray services for ankle injury patients.

The somewhat higher referral rates for radiography in phase I compared to phase 0 (ankle: 91.9% vs. 81.3%; foot: 30.7% vs. 28.9%) may have inflated the potential saving.

Table 24. Comparison of use of radiography in Phase I to that possible if decision rules had been implemented

Radiography	Actual Phase I (N = 750)	Applying Rules (N = 750)	Per- centage Saving
Patients referred for x-ray (%)			
Ankle series	689 (91.9)	441 (58.8)	36.0
Foot series	230 (30.7)	181 (24.1)	21.3
Either series	750 (100)	565 (75.3)	
Total x-ray series performed	919	622	32.3
Mean per patient	1.23	0.83	
X-rays positive for significant fracture (%)			
Ankle series	70 (10.2)	70 (15.9)	
Foot series	32 (13.9)	32 (17.7)	
Both series combined	102 (11.1)	102 (16.4)	
Pattern of series ordered (%)			
None	0 (0)	185 (24.7)	
Ankle only	520 (69.3)	385 (51.3)	
Foot only	61 (8.1)	124 (16.5)	
Ankle and foot	169 (22.5)	56 (7.5)	
Cost of x-ray series performed (\$)			
Ankle (@ \$33.30)	17,316.00	12,820.50	
Foot (@ \$33.30)	2,031.30	4,129.20	
Ankle and foot (@ \$55.40)	9,362.60	3,102.40	
Total	28,709.90	20,052.10	30.2

8. DISCUSSION

8.1 Phase 0: Baseline Use of Radiography

The baseline data collected in phase 0 depict a clear profile of the ankle injury patients that present to an adult emergency department. These patients are generally young, have usually suffered a twisting mechanism of injury, and are more frequently seen in the summer months. Only 14.5%, however, have suffered significant fractures. These more seriously injured patients are proportionately more common among those over age 55 and those seen in the winter months, the latter undoubtedly due to the hazards of ice and snow.

Use of radiography for ankle injury patients is ubiquitous in the emergency departments of the two study hospitals. 94.3% of patients are referred for at least one x-ray series even though the yield for significant fracture is only 13.2%. The total cost annually for the radiography of ankle injury patients at OCH and OGH can be estimated at \$110,000.00. For the province of Ontario, the cost of this one "low technology" procedure is in excess of \$4,000,000.00.

Physicians, no matter what their level of training or experience, continue to rely heavily on radiography despite their awareness that most of the x-rays will be negative. In our survey, we found that physicians estimated a zero or 0.10 probability of fracture in the majority of patients referred for an ankle series. Similarly, they claimed they would have been comfortable in ordering no x-rays in almost one-half of the patients.

Why do emergency physicians order so many apparently inappropriate and wasteful x-rays for ankle injury patients? Clearly, we believe that a major reason is the lack of dependable and widely accepted clinical guidelines and, thus, the rationale for this study.

Other non-clinical factors, however, likely play a significant role in the emergency physician's decision to order radiographs for most patients.^{18 19 20 65} The transient nature of the physician-patient relationship, the high patient volume, the lack of follow-up, the fear of medicolegal repercussions, and the demands of patients in emergency departments all contribute to the expediency of routine ordering of ankle x-rays. Hence,

any effort to alter physician x-ray ordering practices must consider these nonclinical determinants of decision-making. Consequently, we feel that physicians are unlikely to accept any clinical guidelines that do not offer a very high sensitivity to detecting fractures. Future research, such as the field trial proposed in section 8.3.2, may assess the nonclinical issues and the success with which they may be overcome.

8.2 Phase I: Development of the Rules

8.2.1 Interobserver Agreement

Data from this study demonstrate that there is a wide variation in interobserver agreement for the physical findings used to assess ankle injury patients. The best agreement was found for judging ability to bear weight four steps, an assessment that was also found to have good predictive value for the severity of injury. Agreement for most areas of bone tenderness was generally good and suggests that physicians should emphasize these findings in their assessments. The major exception was the proximal fibula, which may reflect the very few positives found. Combinations of two areas of bone tenderness tended to have higher kappa values than some of the individual points and indicates that improved reliability can be achieved by grouping points of tenderness.

The results clearly indicate that some findings are not dependable. Those that we found could not be relied upon in ankle injury patients are ecchymosis, range of motion, soft tissue tenderness, and the anterior drawer sign. Perhaps the lack of agreement reflects a lack of attention by physicians towards variables not deemed clinically useful. The drawer sign would appear to be extremely unreliable but this may, in part, be an artefact of the very low prevalence of this characteristic. In the 100 study patients, the drawer sign was only judged to be positive on six occasions yet never by the same two physicians. Kappa values are known to be dependent on the prevalence of the characteristic being assessed and this can complicate their interpretation.⁶⁶ As the prevalence of a characteristic approaches 0 or 1, the value of kappa approaches 0. Further assessment of the drawer sign is required before concluding that it is completely unreliable.

Reasonable interobserver agreement on swelling could only be achieved when the ordinal scale was dichotomized and then only for areas over the lateral and medial malleoli. This suggests that physician agreement is not good for physical findings rated by an ordinal scale and that a simple binary scale (e.g. "present" or "absent") may be more clinically useful. The gain in information provided by the use of an ordinal scale is offset by the loss in reliability.

This part of the study could have been statistically strengthened and have provided narrower confidence intervals by having more patients assessed by fewer physicians. Obtaining double assessments on more than 100 acutely injured patients in a reasonable period of time would be logistically formidable. The use of 21 experienced clinicians may lend more generalizability to the findings.

8.2.2 Clinical Decision Rules

The study successfully achieved the goal of developing clinical decision rules that are highly sensitive for predicting significant fractures in ankle injury patients. These rules may aid clinicians to be more selective and have the potential to reduce use of radiography by one-third in acute ankle injuries.

The final variables in the decision rules should be easy for physicians to remember and apply. The age guideline for ordering an ankle x-ray series is very clear and biologically plausible. Thirty-four percent of the fractures in the ankle series were seen in people aged 55 or greater, a group that constituted only 13% of the patients. The increased frequency of malleolar fractures above age 55 is likely to be due to osteoporosis. On the other hand, there was no age association for fractures of the midfoot.

Judging ability to bear weight for four steps in the emergency department was shown to be very reliable in our study but is still subject to interpretation and to the reluctance to ask patients to attempt to walk. Many patients were in wheelchairs but were able to bear weight for the requested four steps with little difficulty and minimal coaxing by the physicians. We also found that many patients with sprains had walked initially but had increased difficulty the next day, often associated with increased swelling. Hence, ability to bear weight was a more specific predictor of ankle fracture when one took into

account the patient's difficulty both initially and at the time of assessment. There was no association between ability to bear weight and midfoot fractures.

Localized bone tenderness, as expected, was found to be a good predictor of fracture, but less reliable than age or ability to bear weight. Soft tissue tenderness and swelling frequently make bone tenderness difficult to assess and this is likely the major reason that most ankle injury patients are referred for radiography. We believe that reliance on tenderness of the posterior edge (distal six centimetres) and inferior tip of the lateral malleolus will help to avoid confusion with the tenderness of the anterior talofibular ligament that is so often injured in ankle sprains. Bone tenderness of the medial malleolus is relatively uncommon but, if present, indicates that the patient is at high risk for a fracture.

Soft tissue swelling is highly associated with fractures, a finding frequently reported by radiologists.^{11 38} We found, however, that swelling was also influenced by time from injury and had less interobserver agreement than bone tenderness. Hence, we feel that our decision rules will be more reliable without the inclusion of swelling.

Physician judgement and common sense should always take precedence over clinical guidelines, which are not meant to be inflexible or dogmatic. For example, patients with gross deformity clearly need an x-ray without the need for invoking a decision rule. Caution must be used in interpreting physical findings of patients with altered mental status due to head trauma or drug intoxication.

We could have devised an ankle decision rule with much greater potential savings of x-rays. For example, dropping the inferior tip of the lateral malleolus would have improved the potential savings to 49.8% and improved the specificity to 55.7%. This, however, would have also missed three small fractures and reduced the sensitivity to 95.7%, which we feel would be unacceptable to physicians in the current context of North American medical practice. However, it may be that society will come to accept the small price of occasional missed fractures (which would have likely led to very little morbidity for the patients) in order to improve the efficiency and cost-effectiveness of the medical system.

8.2.3 Methodological Issues

We believe that the methodology of this study adhered to the principles and standards for clinical prediction rules as espoused by Wasson and Feinstein.^{28 29} The following discussion will review these principles and indicate the strengths and weaknesses of the study relative to previous research on ankle injuries.

The outcomes, clinically significant fractures of the ankle and midfoot, were clearly defined, based on objective criteria, and were assessed by an investigator blinded to the predictor variables. Likewise, these predictor variables were explicitly defined and collected in a prospective and standardized fashion by a committed group of experienced clinicians. These clinicians used specific data collection sheets and were blinded to the outcome at the time of the assessment.

The study assessed the reliability or consistency of the predictor variables by measuring the interobserver agreement and considering only those variables with high kappa values. No previous study, to the author's knowledge, has assessed the interobserver agreement of clinical findings in ankle injury patients.

The generalizability of the findings were increased by the use of a sample that included a wide spectrum of severity of injury. The clinical and demographic features of the patients were described and were typical of ankle injury patients commonly seen in emergency departments. This study has the largest reported, prospectively studied, sample of ankle injury patients.

The mathematical techniques, including the univariate analyses and the use of logistic regression and recursive partitioning, were identified and described. Logistic regression was unable to provide a satisfactorily sensitive model without a large decrease in specificity and increase in complexity. This appeared to be because of the tendency of logistic regression to achieve maximum accuracy in distinguishing between the two outcome groups, without regard to sensitivity and specificity separately. In our case, the nonfracture group was ten times as populous as the fracture group leading to a relative lack of accuracy for the latter patients. Therefore, despite the high overall accuracy of the model developed, we were disappointed by the poor sensitivity for fractures.

Recursive partitioning is a less commonly used multivariate method but seemed much better suited to our purpose of identifying all patients within one of the groups.

The investigator performed this procedure manually by repeated crosstabulations of the data, a rather intensive technique. Commercially available software would perform a similar analysis much more quickly, but likely would not afford the same insight into the data.^{63 64}

The study attempted to develop guidelines that are sensible for clinicians. The purpose and application of the rules have been clearly established and are relevant to current emergency department practice. The overt format should be easy to comprehend and reflects suitable component predictor variables. We believe that the simplicity of the rules, unlike the more complex model produced by logistic regression, will make them easy to remember and easy to use at the bedside.

The accuracy of the rules has been demonstrated, meeting the goal of 100% sensitivity for significant fractures. This is the first study to offer guidelines with such high sensitivity. The specificity is an acceptable 40% for the ankle series fractures and 24% for the foot series fractures. This performance classification, however, has only been assessed on the original derivation set of patients.

8.2.4 Limitations of the Study

Several potential areas of bias in the conduct of the study must be carefully considered. Selection of patients was based on the judgement of the assessor physicians who determined whether the patients had any pain or tenderness in zone I or zone II or both. The reliability of this determination has not been shown and ineligible patients may have been admitted and eligible patients overlooked. We feel, however, that the threat to the validity of the findings, from selection bias, should be minimal. We know that the demographic and clinical profile, including the prevalence of fractures, was very similar between the derivation set in phase I and the patients seen in the phase O baseline period.

While outcome assessments were made independently of knowledge of the predictor findings, one can argue that the potential for assessment bias exists because all 750 patients did not have both an ankle and a foot x-ray series. There is the theoretical possibility that zone I fractures were missed in the 61 patients who did not have an ankle series x-ray or that zone II fractures were missed in the 520 patients who did not have a foot x-ray series. We feel that this is very unlikely for several reasons. All patients

had at least one x-ray series and there is significant overlap in the anatomical areas included in the two series. Significant fractures close to the zone being x-rayed would likely have been apparent. Furthermore, the assessor physicians were liberal in their use of radiography, ordering an ankle or foot x-ray series if there was any subjective pain or detectable tenderness in the respective zones. Comparing Tables 7 and 9, we can see that the referral rates for ankle series (91.9% versus 81.3%) and for foot series (30.7% versus 28.9%) were higher during phase I than in phase O. Finally, ordering both x-ray series on all patients would have involved 63% more series than actually ordered. Such a large increase from the normal use of radiography would have been unacceptable to the physicians and to the ethics committees.

Clinicians may be concerned that the ankle series rule was derived from a patient set that had no significant fractures of the talus. Clearly, such fractures are relatively uncommon (none among 905 ankle injury patients in the pilot stage or main study of Phase I and three among 1290 patients in Phase O). Several significant talar fractures have been seen by the investigator since the completion of the study and would have been detected by the rule. This issue would be further addressed in any subsequent validation studies.

This study remains incomplete on two counts: prospective validation and demonstration of impact on patient care. No clinical decision rule can be considered valid until it has been prospectively assessed on a new set of patients. Previous ankle injury guidelines, other than the two studies on the general use of radiography in extremity injuries,^{8 42} have not been prospectively validated. The investigator proposes to conduct a prospective validation study of the decision rules on a new patient set, as described in section 8.3.1. Otherwise, consideration would have been given to the use of a cross-validation technique, such as the "jackknife" method. Should others choose to validate the rules in different settings, the generalizability of the findings will be enhanced.

Furthermore, one cannot be satisfied that a clinical decision is effective until an impact on patient care has been demonstrated. This has been attempted in only one other extremity injury study.⁸ Should these rules stand up to prospective validation, we

propose to then study their ability to actually change physician radiograph ordering behaviour, as described in section 8.3.2.

8.3 Future Research

In view of the limitations described above, one can argue that prospective validation and a field trial should follow the current study. The following sections briefly describe possible objectives and methods for such studies, phase II and phase III.

8.3.1 Phase II: Prospective Validation

The objective of this phase will be to prospectively assess the validity and reliability of the clinical decision rules on a new set of patients.

One thousand patients with acute ankle injuries will be assessed by the same emergency staff physicians at the Ottawa Civic and Ottawa General Hospitals. The physicians will apply the decision rules to each patient and record whether or not the decision rules would have required ankle or foot x-ray series. All patients will then undergo radiography on the same basis as in phase I, i.e. an ankle series if there is any pain or tenderness in zone I and a foot series if there is any pain or tenderness in zone II.

The outcome will be determined by a qualified radiologist who will evaluate each x-ray for the presence or absence of a clinically significant fracture. This evaluation will be made without knowledge of the result of the decision rules.

The validity of the clinical decision rule will be assessed by calculating the sensitivity and specificity for detecting which patients have a significant fracture. 100% sensitivity may be difficult to duplicate in a validation set of patients. Several other clinical variables, not included in the rules but close to contributing to the model (such as swelling), will be retained from Phase I to reassess their predictive value prospectively. If sensitivity proves to be inadequate in this phase, the rules may be derived anew, based on the total population of phases I and II.

The reliability of the decision rules will be assessed by measuring interobserver agreement of the physicians' application of the rule. 100 of the patients will be examined by a second physician who will also classify the patient according to the rule.

Data Analysis will be performed as follows:

a) Sensitivity, specificity, positive predictive value, and negative predictive value will be calculated for the clinical decision rules.

b) Kappa coefficient of interobserver agreement for the rule will be estimated by the methods described in Phase I.

c) The referral fraction for ankle and foot x-ray series and the predicted yield of the x-rays with and without the rules will be calculated. This will permit an estimate of the potential to reduce the cost and the number of x-rays ordered, if the rule had been applied.

8.3.2 Phase III: Field Trial

The objective of this phase will be to test the clinical and cost effectiveness of implementing the decision rules into practice.

The proposed unit of study will be the hospital, using before-after and concurrent control groups. Randomization of patients or physicians to study and control groups, using or not using the decision rule, is not likely feasible within the same institution. Contamination by the previous experience of the physician or by fellow physicians would heavily bias application of the rule. Therefore, all patients within the intervention and control hospitals will be treated in the same fashion.

The rule will be introduced to all physicians (interns, residents, staff) working in the Ottawa Civic and Ottawa General Hospital emergency departments. We will evaluate the effectiveness of implementing this rule on all ankle injury patients presenting over a five month period, which will parallel the months studied in Phase 0. Patients classified as being unlikely to have a fracture will not be sent for x-ray (assuming the co-operation of patient and physician). All other patients will have standard ankle x-rays.

Outcomes will be assessed by classifying the x-rays or by following those patients not x-rayed, at five day intervals. This latter group will be followed until clinical improvement confirms the absence of a fracture. These patients will be x-rayed if their

clinical status does not improve. Note will also be made of patients presenting to another facility and having an x-ray.

The referral fraction for ankle and foot x-ray series will be evaluated for each hospital and compared to the baseline data of Phase 0. Similar data will be collected before and during the field trial, at a control hospital, the Queensway-Carleton Hospital. This community hospital will serve as a concurrent control to demonstrate that the x-ray referral rates do not fall other than in the intervention groups.

This comparison to Phase 0 will allow an estimate of the actual reduction in numbers of x-rays ordered. Cost effectiveness will be determined by comparing the cost of all x-rays required (both at the initial visit and on follow-up) to the baseline costs. We will also compare, with the baseline data, the actual time spent in the emergency department for all ankle injury patients. Finally, we will survey the physicians and the patients to determine the acceptability of the rule. The physicians will be interviewed to assess the impact of non-clinical factors in their decision-making process.

8.4 Conclusion

Phase 0 of this study has confirmed our belief that physicians continue to order radiographs for the vast majority of patients presenting to emergency departments with acute ankle injuries. This practice continues despite the low yield of these radiographs for significant fracture. This practice also continues despite the belief of the emergency physicians that the majority of these radiographs will be negative.

Phase I has successfully achieved the goal of developing 100% sensitive decision rules for the use of radiography in acute ankle injury patients. The data indicate that patients with pain near the malleoli only require an ankle x-ray series if they have one or more of these criteria: age 55 or greater, unable to bear weight both immediately and in the emergency department, or bone tenderness at the posterior edge or inferior tip of either malleolus. Likewise, ankle injury patients with pain in the midfoot only require a foot x-ray series if they have bone tenderness at the base of the fifth metatarsal, the cuboid, or the navicular.

These clinical decision rules have the potential to reduce the use of radiography and the associated costs by more than 30% in adult ankle injury patients. The rules, however, must now be prospectively validated, ideally in several different settings. The final test of these rules will be to demonstrate their impact on patient care, in terms of actually reducing the use of radiography in ankle injury patients.

REFERENCES

1. Stother IG. Incidence of minor fractures in twisting injuries of the ankle. *Injury* 1974;5:213-4.
2. Brooks SC, Potter BT, Rainey JB. Inversion injuries of the ankle: clinical assessment and radiographic review. *Br Med J* 1981;282:607-8.
3. Vargish T, Clarke WR, Young RA, Jensen A. The ankle injury - indications for the selective use of x-rays. *Injury* 1983;14:507-12.
4. Montague AP, McQuillan RF. Clinical assessment of apparently sprained ankle and detection of fracture. *Injury* 1985;8:545-6.
5. Dunlop MG, Beattie TF, White GK, Raab GM, Doull RI. Guidelines for selective radiological assessment of inversion ankle injuries. *Br Med J* 1986;293:603-5.
6. Sujitkumar P, Hadfield JM, Yates DW. Sprain or fracture? An analysis of 2000 ankle injuries. *Arch Emerg Med* 1986;3:101-6.
7. Diehr P, Highley R, Dehkordi F, et al. Prediction of fracture in patients with acute musculoskeletal ankle trauma. *Med Decis Making* 1988;8:40-7.
8. Brand DA, Frazier WH, Kohlhepp WC, et al. A protocol for selecting patients with injured extremities who need x-rays. *N Engl J Med* 1982;306:333-9.
9. Gratton MC, Salomone JA, Watson WA. Clinically significant radiograph misinterpretations at an emergency medicine residency program. *Ann Emerg Med* 1990;19:497-502.
10. Abrams HL. The "overutilization" of x-rays. *N Engl J Med* 1979;300:1213-6.
11. Cockshott WP, Jenkin JK, Pui M. Limiting the use of routine radiography for acute ankle injuries. *Can Med Assoc J* 1983;129:129-31.
12. Gleadhill DNS, Thomson JY, Simms P. Can more efficient use be made of x-ray examinations in the accident and emergency department? *Br Med J* 1987;294:943-7.
13. Ontario Ministry of Health, personal communication, March 1991.
14. Bell RS, Loop JW. The utility and futility of radiographic skull examination for trauma. *N Engl J Med* 1971;284:236-9.
15. Masters SJ, McClean PM, Argarese JS, et al. Skull x-ray examinations after head trauma. *N Engl J Med* 1987;316:84-91.

16. Lloyd S. Acute ankle injuries: clinical/radiologic assessment in diagnosis. *Can Fam Physician* 1988;34:2261-5.
17. Feinstein AR. The "chagrin factor" and qualitative decision analysis. *Arch Intern Med* 1985;145:1257-9.
18. Lloyd S. Selective radiographic assessment of acute ankle injuries in the emergency department: barriers to implementation [Editorial]. *Can Med Assoc J* 1986;135:973-4.
19. Long A. Radiographic decision-making by the emergency physician. *Emerg Med Clin N Am* 1985;3:437-46.
20. Svenson J. Need for radiographs in the acutely injured ankle [Letter]. *Lancet* 1988;1:244-5.
21. Goldman L, Weinberg M, Weisberg, et al. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med* 1982; 307:588-96.
22. Pozen MW, D'Agostino RB, Selker HP, Sytkowski PA, Hood WB. A predictive instrument to improve coronary-care-unit admissions practice in acute ischemic heart disease: a prospective multicenter clinical trial. *N Engl J Med* 1984; 310:1273-8.
23. Goldman L, Cook EF, Brand DA, et al. A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N Engl J Med* 1988; 318:797-803.
24. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. *Ann Emerg Med* 1989;18:13-20.
25. Heckerling PS, Tape TG, Wigston RS, et al. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med* 1990;113:664-70.
26. Bates DW, Cook EF, Goldman L, Lee TH. Predicting bacteremia in hospitalized patients: a prospectively validated model. *Ann Intern Med* 1990;113:495-500.
27. Fischl MA, Pitchenik A, Gardner LB. An index predicting relapse and need for hospitalization in patients with acute bronchial asthma. *N Engl J Med* 1981;305:783-9.
28. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules: application and methodological standards. *N Engl J Med* 1985;313:793-9.
29. Feinstein AR. *Clinimetrics*. New Haven: Yale University Press, 1987.

30. Hennekens CH, Buring JE. *Epidemiology in medicine*. Toronto: Little, Brown, 1987.
31. Mulrow CD, Dolmatch BL, DeLong ER, et al. Observer variability in the pulmonary examination. *J Gen Intern Med* 1986;1:364-7.
32. Feinstein AR. *Clinical epidemiology. The architecture of clinical research*. Philadelphia: WB Saunders, 1985.
33. Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology. The essentials*. 2nd ed. Baltimore: Williams & Wilkins, 1988:22-9.
34. Charlson ME, Ales KL, Simon R, Mackenzie R. Why predictive indexes perform less well in validation studies. *Arch Intern Med* 1987;147:2155-61.
35. Lee TH. Evaluating decision aids: the next painful step [editorial]. *J Gen Intern Med* 1990;5:528-9.
36. Garfield JS. Is radiological examination of the twisted ankle necessary? *Lancet* 1960;2:1167-9.
37. Lettin AWF. Diagnosis and treatment of sprained ankle. *Br Med J* 1963;1:1056-60.
38. deLacey G, Bradbrooke S. Rationalising requests for x-ray examination of acute ankle injuries. *Br Med J* 1979;280:1597-8.
39. Charny MC, Ennis WP, Roberts CJ, Evans KT. Can the use of radiography of arms and legs in accident and emergency units be made more efficient? *Br Med J* 1987;294:291-3.
40. Beaulieu M-D, Corriveau A, Nadeau P-O. Evaluation et traitement de l'entorse externe de la cheville dans un milieu de soins de premiere ligne: la radiographie systematique est-elle essentielle? *Can Med Assoc J* 1986;135:1003-6.
41. West A. Assessing the injured ankle without x-rays. *Br J Clin Pract* 1988;43:360-2.
42. McConnochie KM, Roghmann KJ, Pasternack J, Monroe DJ, Monaco LP. Prediction rules for selective radiographic assessment of extremity injuries in children and adolescents. *Pediatrics* 1990;86:45-57.
43. Rockwood, CA, Green, DP, editors. *Fractures in adults*. 2nd ed. Philadelphia: JB Lippincott, 1984.
44. Simon RR, Koenigskecht, SJ. *Emergency orthopedics: the extremities*. 2nd ed. Norwalk: Appleton & Lange, 1988.

45. SPSS Inc. SPSS Data Entry II. Chicago: SPSS Inc, 1988.
46. Schedule of benefits: physician services. Ontario Ministry of Health, 1988.
47. Hulley SB, Cummings SR, editors. Designing clinical research: an epidemiologic approach. Baltimore: Williams and Wilkins, 1988.
48. Armitage P, Berry G. Statistical methods in medical research. 2nd ed. Oxford: Blackwell Scientific. 1987.
49. Flack VF, Afifi AA, Lachenbruch PA. Sample size determinations for the two rater kappa statistic. *Psychometrika* 1988;53:321-5.
50. SPSS Inc. SPSS Manual. Chicago: SPSS Inc., 1990
51. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
52. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons, 1981:212-37.
53. Kramer MS, Feinstein AR. Clinical biostatistics: LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1982;29:111-23.
54. Sackett DL, Haynes RB, Tugwell P. Clinical epidemiology. A basic science for clinical medicine. Toronto: Little, Brown and Co., 1985:17-45.
55. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons, 1989.
56. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, 1986.
57. SPSS Inc. SPSS/PC+ Update for V3.0 and 3.1. Chicago: SPSS Inc., 1989.
58. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston: Butterworths, 1988.
59. Friedman JH. A recursive partitioning decision rule for nonparametric classification. *IEEE Trans Comput* 1977;16:404-8.
60. Ciampi A, Thiffault J, Nakache J-P, Asselain B. Stratification by stepwise regression, correspondence analysis and recursive partition: a comparison of three methods of analysis for survival data with covariates. *Comput Stat Data Anal* 1986;4:185-204.
61. Ciampi A, Chang C-H, Hogg S, McKinney S. Recursive partition: a versatile method for exploratory data analysis in biostatistics. *Biostatistics* 1987:23-50.

62. Ciampi A, Hogg SA, McKinney S, Thiffault J. RECPAM: a computer program for recursive partition and amalgamation for censored survival data and other situations frequently occurring in biostatistics. I. methods and program features. *Comput Meth Prog Biomed* 1988;26:239-56.
63. Magdison J. SPSS/PC+ CHAID. Chicago: SPSS Inc., 1989.
64. FirstMark Technologies Ltd. Knowledgeseeker User's Guide Version 2.0. Ottawa: FirstMark, 1990.
65. Matthews MG. Guidelines for selective radiological assessment of inversion ankle injuries[Letter]. *Br Med J* 1986;293:959.
66. Thompson WD, Walter SD. A reappraisal of the kappa coefficient. *J Clin Epidemiol* 1988;41:949-58.

LIST OF APPENDICES

Appendix A. Pilot stage data collection sheets	ii
Appendix B. Final data collection sheets	iv
Appendix C. Sample calculation of kappa coefficient (after Sackett et al. ⁵⁴)	vi
Appendix D. Interobserver agreement data for 100 patients	viii

Appendix A. Pilot stage data collection sheets

UNIVERSITY OF OTTAWA ANKLE INJURIES STUDY

FOR FULLTIME EMERGENCY STAFF PHYSICIANS ONLY

INCLUSION
CRITERIA

All patients who have suffered acute blunt ankle injuries, including twisting, falls from a height, direct blows, and vehicle accidents.

See back of form for anatomical structures to be included.

EXCLUSION
CRITERIA

- 1) Patients under the age of 18 years
- 2) Pregnant women
- 3) Injuries of the skin (superficial lacerations or abrasions, puncture wounds, burns)
- 4) Patients referred from outside the hospital and who have already had ankle x-rays
- 5) Patients whose ankle injury occurred more than 10 days previously
- 6) Patients who return for the same ankle injury

PATIENT
INFORMATION

- 1) Name: _____ 2) Char/Unique No: _____
- 3) Able to bear weight immediately after injury? YES / NO
- 4) Cracking or popping sound at time of injury? YES / NO
- 5) Mechanism of injury: TWISTING / FALL FROM A HEIGHT / DIRECT BLOW
/ VEHICLE ACCIDENT / OTHER
- 6) Direction of injury: INVERSION / EVERSION / PLANTARFLEXION / UNKNOWN
- 7) Time from injury to assessment: _____ HOURS _____ MINUTES
- 8) Application of ice in previous 6 hours? YES / NO
- 9) Use of analgesics in previous 6 hours? YES / NO
- 10) Ankle fracture in previous 12 months? YES / NO
- 11) History of impaired sensation in lower limbs (e.g. multiple sclerosis, diabetes)? YES / NO

PHYSICAL
EXAMINATION

- 12) Altered level of consciousness (e.g. head injury, intoxication)? YES / NO
- 13) Bone deformity? YES / NO
- 14) Visible swelling medially? YES / NO
- 15) " " over lateral malleolus? YES / NO
- 16) " " over anterior talo-fibular ligament? YES / NO
- 17) " " anteriorly? YES / NO
- 18) Ecchymosis? YES / NO
- 19) Range of motion (plantar-dorsiflexion): NORMAL / LIMITED
- 20-29) Bone tenderness - mark diagram on back of form
- 30-34) Soft tissue tenderness - mark diagram on back of form
- 35) Wound overlying bony tenderness? YES / NO
- 36) Anterior drawer sign positive (compared to other ankle)? YES / NO
- 37) Able to bear weight now (can take at least 4 steps)? YES / NO

PHYSICIAN
INFORMATION

- 38) Name: _____
- 39) Interobserver reliability tested? YES / NO
- 40) If "YES", name of physician: _____
- 41) Predict probability of significant fracture:
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

COMMENTS:

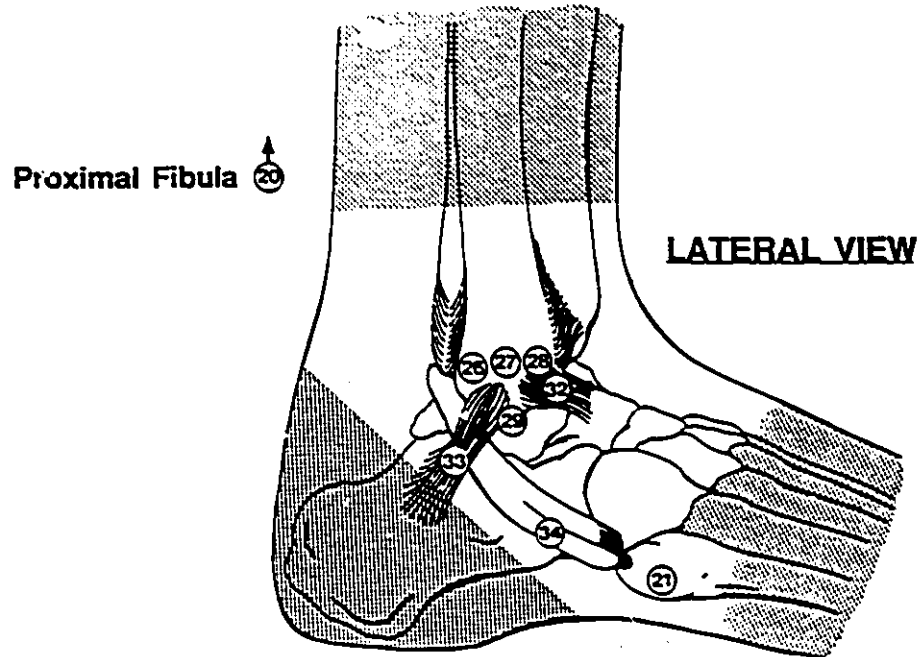
****THANK-YOU FOR YOUR HELP****

RESEARCH
ASSISTANT

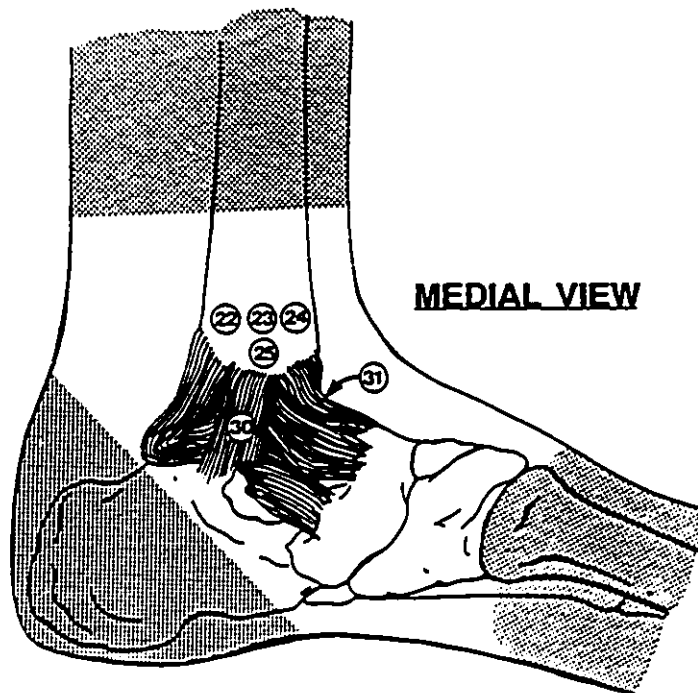
- 42) Age: ____ 43) Sex: ____ 44) Date: ____ / ____ / ____ 45) Hospital: OCH / OGH
- 46) X-ray result for fracture: NONE / INSIGNIFICANT / SIGNIFICANT

SITES OF LOCALIZED TENDERNESS

(Please place an X in the appropriate circles.)



N.B. Shaded areas not included in study



Appendix B. Final data collection sheets

UNIVERSITY OF OTTAWA ANKLE INJURIES STUDY

FOR FULLTIME EMERGENCY STAFF PHYSICIANS ONLY

INCLUSION CRITERIA All patients who have suffered acute blunt ankle injuries, including twisting, falls from a height, direct blows, and vehicle accidents. See back of form for anatomical structures to be included.

EXCLUSION CRITERIA

- 1) Patients under the age of 18 years
- 2) Pregnant women
- 3) Isolated injuries of the skin (superficial lacerations or abrasions, puncture wounds, burns)
- 4) Patients referred from outside the hospital and who have already had ankle x-rays
- 5) Patients whose ankle injury occurred more than 10 days previously
- 6) Patients who return for the same ankle injury

PATIENT INFO

- * 1) Name: _____ 2) Chart/Unique No: _____ 3) Date: ___/___/___
- 4) Able to bear weight immediately after injury?
YES (CONTINUED CURRENT ACTIVITY) / YES (WITH DIFFICULTY) / NO
- 5) Cracking or popping sound at time of injury? YES / NO
- 6) Mechanism of injury:
TWISTING / FALL FROM A HEIGHT / DIRECT BLOW / VEHICLE ACCIDENT / OTHER
- 7) Direction of injury: INVERSION / EVERSION / PLANTARFLEXION / UNKNOWN / N/A
- 8) Time from injury to assessment: _____ HOURS
- 9) Application of ice in previous 6 hours? YES / NO
- 10) Ankle fracture in previous 12 months? YES / NO

PHYSICAL EXAM

- * 11) Ecchymosis YES / NO
- * 12) Visible swelling
 - a) medially? NONE / MINIMAL / MODERATE / MARKED
 - b) over lateral malleolus? NONE / MINIMAL / MODERATE / MARKED
 - c) over anterior talo-fibular ligament? NONE / MINIMAL / MODERATE / MARKED
 - d) anteriorly? NONE / MINIMAL / MODERATE / MARKED
- * 13) Restriction of active plantar-dorsiflexion? NONE / MINIMAL / MODERATE / MARKED
- * 14) Bone tenderness - MARK DIAGRAM on back of form
- * 15) Soft tissue tenderness - MARK DIAGRAM on back of form
- * 16) Anterior drawer sign positive (compared to other ankle)? YES / NO
- * 17) Ask the patient: "Are you able to walk now, unaided, or did you bear weight on your way into the E.D.?" YES / NO If "NO", skip question #18.
- * 18) Ask the patient to walk and rate the degree of difficulty:
NONE / MINIMAL / MODERATE / MARKED / UNABLE / REFUSED / N/A

PHYSICIAN INFO

- 19) Name: _____ 20) Interobserver reliability tested? YES / NO
- * 21) If "YES", name of physician: _____ (DO QUESTIONS MARKED WITH *)
- * 22) Predict probability of significant fracture in:
 - a) Zone I (True Ankle) 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
 - b) Zone II (Midfoot) 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
- * 23) How comfortable would you be in not ordering any x-rays for this patient?
(1)Very Comfortable (2)Comfortable (3)Neutral (4)Uncomfortable (5)Very Uncomfortable

****PATIENTS WITH ZONE I TENDERNESS NEED AN ANKLE X-RAY AND THOSE WITH ZONE II TENDERNESS NEED A FOOT X-RAY. DO BOTH AT YOUR DISCRETION****
****THANK-YOU FOR YOUR HELP****

RESEARCH ASSISTANT

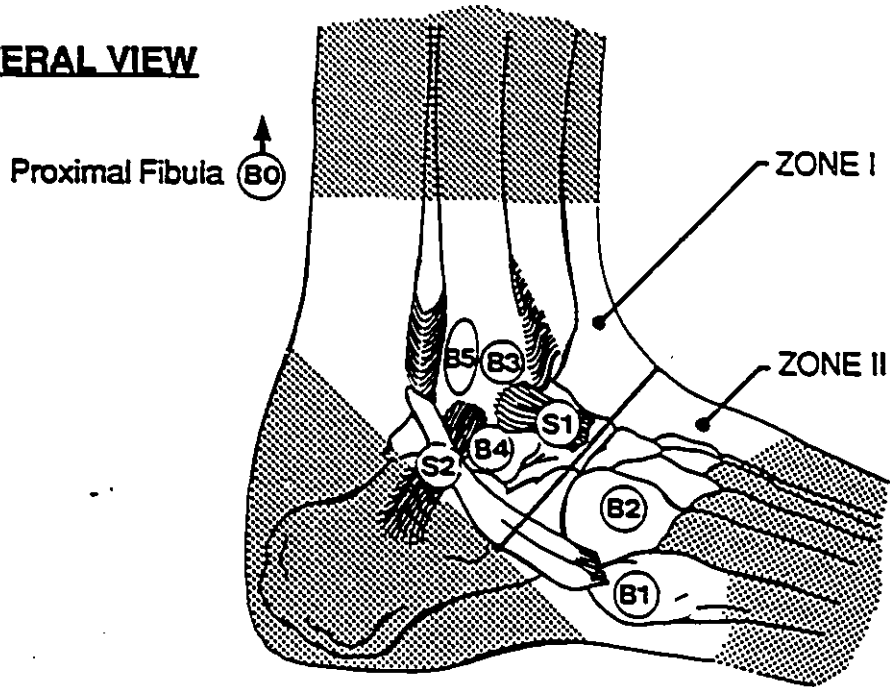
- 24) Age: ___ 25) Sex: ___ 26) OCH/OGH 27) Ankle x-ray: ND/N/NS/S 28) Foot x-ray: ND/N/NS/S
- 29) Fracture: a) Lat Mall NS/S b) Med Mall NS/S c) Post Mall NS/S d) Talus NS/S e) Calc NS/S
f) Nav NS/S g) Cun NS/S h) Cub NS/S i) 5th Met NS/S

v

SITES OF LOCALIZED TENDERNESS

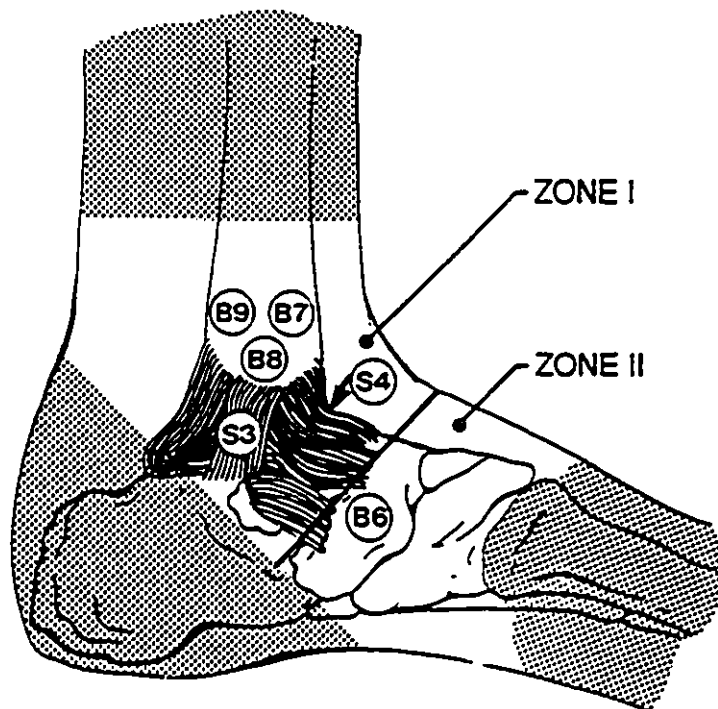
(Please place an X in the appropriate circles.)

LATERAL VIEW



N.B. Shaded areas not included in study

MEDIAL VIEW



Appendix C. Sample calculation of kappa coefficient (after Sackett et al.⁵⁴)

The accompanying figure illustrates the calculation of kappa, applied to the interobserver agreement on the presence or absence of moderate to marked swelling over the medial malleolus. Whereas the observed agreement between physicians appears to be 89%, this is inflated by the amount of agreement that could occur by chance, given the two physicians' overall observations. In the example, the physicians might be expected to agree, based purely on chance, that 86% of 83 (71.4) patients would have no or minimal swelling and 14% of 17 (2.4) patients would have moderate or marked swelling. Therefore, the overall agreement expected on the basis of chance is calculated to be 74% ($71.4 + 2.4 / 100$). This means that the actual agreement beyond chance is $89 - 74$, or 15%. The kappa value (0.58) represents the ratio of actual agreement beyond chance (15%) to the potential maximum agreement beyond chance (26%).

		Second Physician		
		none or minimal swelling	moderate or marked swelling	
First Physician	none or minimal swelling	79	4	83
	moderate or marked swelling	7	10	17
		86	14	100 Patients

$$\text{OBSERVED AGREEMENT} = \frac{79 + 10}{100} = 89\%$$



$$\text{AGREEMENT EXPECTED ON THE BASIS OF CHANCE} = \frac{(86\% \times 83) + (14\% \times 17)}{100} = 74\%$$



$$\text{ACTUAL AGREEMENT BEYOND CHANCE} = 89\% - 74\% = 15\%$$



$$\text{POTENTIAL AGREEMENT BEYOND CHANCE} = 100\% - 74\% = 26\%$$



$$\text{KAPPA} = \frac{\text{ACTUAL AGREEMENT BEYOND CHANCE}}{\text{POTENTIAL AGREEMENT BEYOND CHANCE}} = \frac{15\%}{26\%} = 0.58$$

Appendix D. Interobserver agreement data for 100 patients

Table i. Number of cases rated as present or absent for general physical findings by two physicians*

Variable	1st Present 2nd Present	1st Present 2nd Absent	1st Absent 2nd Present	1st Absent** 2nd Absent	Kappa
Ecchymosis	14	9	14	59	0.39
Soft tissue tenderness					
Anterior talofibular ligament (S1)	62	13	10	15	0.41
Calcaneofibular ligament (S2)	6	9	9	76	0.29
Deltoid ligament (S3)	14	10	16	59	0.34
Anterior aspect of ankle (S4)	18	14	10	58	0.42
Anterior drawer sign	0	2	4	85	-.03
Ability to bear weight in ED	34	4	4	56	0.83

* Not all variables were assessed in 100 patients

** 1st = present or absent according to first physician; 2nd = present or absent according to second physician

Table ii. Number of cases rated as none-minimal or moderate-marked for physical findings by two physicians

Variable	1st Marked 2nd Marked	1st Marked 2nd None	1st None 2nd Marked	1st None* 2nd None	Kappa
Restriction of range of motion	39	20	13	28	0.34
Swelling					
Medial malleolus	10	7	4	79	0.58
Lateral malleolus	45	7	7	41	0.72
Anterior talofibular ligament	27	22	16	35	0.24
Anterior aspect of ankle	3	7	13	77	0.12

* 1st = rating of first physician; 2nd = rating of second physician;
None = rated as none or minimal for abnormality;
Marked = rated as moderate or marked for abnormality

Table iii. Number of cases rated as present or absent for bone tenderness by two physicians

Variable	1st Present 2nd Present	1st Present 2nd Absent	1st Absent 2nd Present	1st Absent* 2nd Absent	Kappa
Proximal fibula (B0)	0	2	1	97	-.01
Midfoot					
Base 5th metatarsal (B1)	8	2	2	88	0.78
Cuboid (B2)	8	5	10	77	0.43
Navicular (B6)	1	3	1	95	0.32
Either B1 or B2 or B6	14	7	9	70	0.53
Lateral Malleolus					
Anterior edge (B3)	52	10	7	31	0.64
Inferior tip (B4)	33	13	13	41	0.48
Posterior edge (B5)	34	10	2	54	0.75
Either B4 or B5	50	6	6	38	0.76
Medial Malleolus					
Anterior edge (B7)	7	4	7	82	0.50
Inferior tip (B8)	15	3	8	74	0.66
Posterior edge (B9)	7	5	3	85	0.59
Either B8 or B9	17	5	17	71	0.66
Any one of B4 or B5 or B8 or B9	58	5	6	31	0.76

* 1st = present or absent according to first physician; 2nd = present or absent according to second physician