WHAT IS THE BEST COMORBIDITY INDEX FOR RETROSPECTIVE SURVIVAL STUDIES IN HEAD AND NECK ONCOLOGY?

by

Stephen F Hall MD FRCSC

A thesis submitted in conformity with the requirements for the degree of Master of Science
The Graduate Department of Health Administration
Clinical Epidemiology and Health Care Research Program, Faculty of Medicine
University of Toronto

© Copyright by Stephen Frederick Hall, 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

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

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

0-612-58820-3
Comorbidities are diseases or conditions that co-exist with a disease of interest. The importance of comorbidities is that they alter treatments, change resource utilization and confound the results of treatment survival analysis. The objective of this study was to determine the best comorbidity index to use in survival analysis of patients with squamous cell carcinoma of the Head and Neck. Five validated indexes, with very different methodologies, including the Charlson Index, the Cumulative Illness Rating Scale, the Kaplan-Feinstein Classification, the Index of Co-existent Disease, and the Chronic Disease Scale were critically reviewed and then tested. Data from 379 unselected consecutive patients from the Kingston Regional Cancer Center with complete 3 year follow-up were used. Kaplan-Meier analysis and Cox Proportional Hazards Regression were used to stratify patients into three levels of increasing severity of comorbidity for each index. The Kaplan-Feinstein Classification was the most successful in stratifying patients in this population.
Table of Contents

Abstract ........................................ ii
Table of Contents ................................ iii
Acknowledgments ................................. v
List of Tables ................................... vi
List of Figures ................................... vii
List of Appendices ............................... viii
Glossary of Abbreviations ...................... ix

Chapter 1: Introduction
1.1 Rationale .................................. 1
1.2 The Study Objectives ....................... 4
1.3 Overview .................................. 5

Chapter 2: Literature Review
2.1 Search strategy and results .............. 7
2.2 Critical appraisal of the five indexes 9
   2.2.1 The Cumulative Illness Rating Scale (CIRS) 
   2.2.2 The Kaplan-Feinstein Classification (KFC) 
   2.2.3 The Charlson index .................... 22
   2.2.4 The Index of Co-Existent Disease (ICED) 
   2.2.5 The Chronic Disease Scale (CDS) .... 34
2.3 Comparisons of the five indexes in the Literature 37
2.4 Measuring Comorbidity in Head and Neck Oncology 42
2.5 Summary .................................. 45

Chapter 3: Methods
3.1 Data collection ............................ 50
   3.1.1 patient population ....................... 54
   3.1.2 clinical database ........................ 55
   3.1.3 study population ......................... 56
   3.1.4 variables ............................... 57
   3.1.5 data accuracy and completeness .......... 59
3.2 Development of abstraction and entry module 
3.3 An example of index score calculation 60
3.4 Abstraction ................................ 62
3.5 Plan of analysis ........................... 63
3.6 Ethics approval ........................... 64

Chapter 4: Results
4.1 Demographics and Variable distributions 64
4.2 Kaplan-Meier Analysis ...................... 65
4.3 Multivariable Analysis ..................... 66
4.4 Radio Operating Characteristics .......... 67
4.5 Percent Variance Explained .............. 68
4.6 Effect of follow-up time ................. 69
4.7 Effect of disease specific survival ...... 70
4.8 Effect of treatments ....................... 71
4.9 Effect of intent of treatment ............. 72
4.10 Effect of age ................................ 73
4.11 Correlation of Index scores .............. 74
4.12 Further Analysis .......................... 75
4.13 Comments of the Abstractor .............. 76
Chapter 5: Discussion

5.1 What is the best comorbidity scale for retrospective survival studies in Head and Neck Oncology?

5.2 How does a summative scale perform in comparison to a scale using the score from the most severe illness?

5.3 How does a pharmacy based scale perform compared to a clinical record based scale?

5.4 Clinical and Research Implications of this study

5.5 Strengths and Weaknesses

5.6 Future directions

5.7 Conclusion

Bibliography

Appendices
Acknowledgments

The preparation and completion of this thesis would not have been possible without the insight and patience of my supervisor Dr. Paula Rochon, the wisdom and humor of Dr. David Streiner, and the knowledge and experience of Dr. Lawrence Paszat. Thank you for your guidance.

I would like to express my appreciation to the following people who were responsible for my success in the Clinical Epidemiology Program: Dr. Marsha Cohen who ‘took a chance with me’ in 1996. Dr. Claire Bombardier, my advisor in the Certificate Program. The three successive Deans of Medicine at Queen’s University who funded this opportunity for me. Dr. Patti Groome, my inspiration.

I am grateful to all the staff of the Radiation Oncology Research Unit of Queen’s University, especially Ms Susan Rohlands who designed the PC version of the data abstraction/entry module.

I would like to thank Mrs. Louise Telford who performed the data abstraction, Dr. Wico Mulder who reviewed the patient database and Dr. Duncan Sinclair who proofread the manuscript.

I acknowledge the Claire Nelson Bequest of the Kingston General Hospital in Kingston Ontario that funded the development of the abstraction module and abstraction phases of the project.

I cannot thank my wife Lynn enough for her love, patience and understanding over the past 4.5 years.
List of Tables

2.1 Relevant studies that have used the Cumulative Illness Rating Scale to measure comorbidity
2.2 Relevant studies that have used the Kaplan-Feinstein Classification to measure comorbidity
2.3 Relevant studies that have used the Charlson Index to measure comorbidity
2.4 Relevant studies that have used the Index of Coexistent Disease to measure comorbidity
2.5 Relevant studies that have used the Chronic Disease Scale
2.6 Studies that compare another Comorbidity Index with the Charlson Index
2.7 Studies that measure comorbidity in Head and Neck Oncology
2.8 Summary of the critical appraisal of the 5 indexes
2.9 Summary of Reliability Studies

4.1 Clinical information on 379 patients with squamous cell carcinoma of the Head and Neck
4.2 the frequency distribution of scores and univariate survival for 3 levels of each index
4.3 the p values of log-rank tests for the best 3 survival curves for each index
4.4 the results of the Cox Proportional Hazards regressions for 383 patients with Head and Neck Cancer
4.5 the Percent Variance Explained from the Cox models for the CIRS, the KFC, the Charlson Index and the ICED
4.6 the correlation of index scores
4.7 summary of the analysis using 2 prognostic levels
List of Figures

2.1 Search Strategy for Literature Review
2.2 Studies that compare another comorbidity index to the Charlson Index

3.1 The Study Design

4.1 the distribution of the Charlson Index and
the Chronic Disease Scale scores
4.2 the distribution of the Kaplan-Feinstein Classification and
the Index of Co-existent Disease scores
4.3 the distribution of the Cumulative Illness Rating Scale scores
4.4 the probability of survival for all patients
4.5 the probability of survival by T stage
4.6a the probability of survival by the CIRS for the training set (n=258)
4.6b the probability of survival by the KFC for the training set (n=258)
4.6c the probability of survival by the Charlson Index for the training set (n=258)
4.6d the probability of survival by the ICED for the training set (n=258)
4.7a the probability of survival by the CIRS for the test set (n=121)
4.7b the probability of survival by the KFC for the test set (n=121)
4.7c the probability of survival by the Charlson Index for the test set (n=121)
4.7d the probability of survival by the ICED for the test set (n=121)
4.8a the probability of survival by the CIRS for all patients (n=379)
4.8b the probability of survival by the KFC for all patients (n=379)
4.8c the probability of survival by the Charlson Index for all patients (n=379)
4.8d the probability of survival by the ICED for all patients (n=379)
4.9 the log (-log) plot by the KFC
4.10 the relative risk point estimates and the 95% confidence intervals for each index (n=363)
4.11 the relative risk point estimates and the parameter estimates from the Cox model by level for each index
4.12 the probability of survival by the Charlson Index (disease specific survival)
4.13 the relative risk point estimates and the 95% confidence intervals
for each index (patients treated for cure)
4.14 the ROC curves for the KFC and the CIRS
4.15 the ROC curves for the ICED and the Charlson Index
List of Appendices

1. The Cumulative Illness Rating Scale
2. The Kaplan-Feinstein Classification
3. The Charlson Index
4. The Index of Co-existent Diseases
5. The Chronic Disease Scale
6. Medlog data entry sheets
7. Paper version of the comorbidity abstraction module
8. Electronic pages of the comorbidity abstraction module
Glossary of Abbreviations

AAA - abdominal aortic aneurysm
ADL - Activities of Daily Living Scale
APACHE II - Acute Physiology and Chronic Health Evaluation (severity of illness index for patients in Intensive Care Units)
ASA - American Society of Anesthesiology physical status scale
CCI - Charlson Combined Index (with age)
CDS - Chronic Disease Scale
CIRS - Cumulative Illness Rating Scale
CIRS-G - Miller version or Geriatric version of CIRS
ECOG-PS - Eastern Cooperative Oncology Group Performance Scale
FS - Functional Severity (part of the ICED scale)
HMO - Health Management Organization
ICC - Intraclass Correlation Coefficient - reliability test for multiple observations by multiple independent observers that accounts for the variability of responses
ICD9 - International Classification of Disease
ICED - Index of Co-existent Illness
IDS - Individual Disease Severity (part of the ICED scale)
KFC - Kaplan-Feinstein Classification
KM - Kaplan-Meier (bivariate analysis for time to event data)
KRCC - Kingston Regional Cancer Center
mKFC - modified KFC (Piccirillo)
MI - myocardial infarction
N stage - regional lymph node metastasis status portion of TNM classification
OR - odds ratio
PC - personal computer
PL - prognostic level
PVE - percent variance explained
ROC - receiver operating characteristics (curve)
SEER - Surveillance, Epidemiology and End Results Program of the National Cancer Institute
THR - total hip replacement
T stage - size of primary tumor in TNM classification
TNM - anatomic staging system for the classification of malignant tumors (American Joint Committee for Cancer Staging and the Union International Contre le Cancer)
Chapter 1 Introduction

The objectives of this chapter are:

i. to provide the rationale for the study;
ii. to outline the objectives of the project;
iii. to provide an overview of the thesis;

1.1 Rationale

JP, a 58 year old businessman and life-long smoker, was admitted to the Coronary Care Unit with his second myocardial infarction. At the time of admission he also complained of hoarseness, sore throat, occasional shortness of breath on exertion and a hard mass on the left side of his neck. When his throat was examined, he was found to have a very extensive carcinoma of the larynx. The recommended treatment was major surgery (laryngectomy) followed by radiotherapy. In preparation for his cancer treatment, he had Coronary Artery Bypass surgery, but suffered a small stroke intra-operatively and a deep vein thrombosis during his recovery. Surgery for his throat cancer was not possible, and therefore he was treated with radiotherapy. He had a good initial response, but the cancer quickly returned. He subsequently had a tracheotomy and five months later was admitted to hospital where he died.

Comorbidities are diseases or disorders that co-exist with a disease of interest. The importance of comorbidities are that they can alter overall survival, can influence treatments offered to patients, are related to the incidence of complications of treatments, can interfere with diagnosis, can alter resource utilization, and can confound the results of treatment outcome analysis. Mr JP exemplifies some of the problems related to comorbid illness. First, because of his heart disease, he did not receive the recommended treatment. Second, if his medical record became part of a retrospective study of radiotherapy treatment effectiveness, he is recorded as a treatment failure. Mr JB’s pre-therapeutic condition determined his treatment and may have determined his outcome. Unfortunately, throughout the medical literature and especially in oncology research, we have not developed, tested or effectively used comorbidity measurements to adjust the assessment of treatments for the impact of pre-therapeutic health status. Feinstein summed up the problem as follows:

“When treatments are compared ... no one can tell whether the post-therapeutic differences are due to the treatment or to clinically important pre-therapeutic prognostic distinctions. ... In every branch of scientific biology, when two different variables are important, we classify and analyse both. Yet the prognostic and therapeutic statistics of cancer persist in classifying only the morphology of the tumor while ignoring the rest of the patient.”

An index or scale to measure comorbidity should reduce a patient’s known total medical burden to a single number on a severity scale that can be used to stratify patients. There are four validated indexes or scales to measure comorbidity: the Cumulative Illness Rating Scale, the Kaplan-Feinstein Classification, the Charlson Index, and the Index of Co-existent Illness. The Cumulative Illness Rating Scale (CIRS), reported in 1968, was the first index to be described. Linn et al. created a scale to measure “total medical burden” or physical impairment. The scale consists of a combination of 14...
disease systems with a severity of illness scale across each system and was tested on a series of 172 male discharges. The CIRS summative score correlated with subsequent mortality. The next scale to be described was the Kaplan-Feinstein Classification (KFC) published in 1974. This scale was tested on a series of 201 newly diagnosed male diabetics and consists of the combination of 12 disease systems and a severity of illness scale. They found that the KFC scores correlated with mortality. The best known index, the Charlson Index, was published in 1987. Charlson et al. devised a scale of 16 specific medical conditions that most predicted mortality, based on 607 admissions to a general medical service, and validated it in a study of patients with breast cancer. This scale has been widely used in studies that control for differences in comorbidity and has been adapted to a format for administrative data. A report using The Index of Co-existent Disease (ICED) was published in 1993 by Cleary and Greenfield. This scale is similar to the KFC, but includes more disease systems and a functional disability scale. Cleary et al. found that the ICED could be used to help predict the incidence of post-operative complications.

Head and Neck Cancer represents approximately five percent of all cancers (excluding skin and thyroid). In Ontario, there are approximately 1,800 new cases per year of which squamous cell carcinoma is the most common type (85% of cases). Typically in published series of patients, the mean age is approximately 65, of which 75% are males. Patients present with symptoms of hoarseness, sore throat, sore mouth, dysphagia or neck masses. The most common sites for the primary cancers are larynx (35%), oral cavity including tongue (25%), oropharynx including tonsil (20%) and hypopharynx (10%). Tumor size is staged using the TNM Classification of Malignant Disease (see glossary), with the incidence of early stage disease (T1 and T2) approximately 60% and the incidence of neck metastases less than 30%. The incidence of distant metastases at presentation is low. The traditional curative treatments for Head and Neck cancer have been radiotherapy, surgery, or planned sequential surgery with radiotherapy. Until recently, chemotherapy had no proven role in the treatment of these patients, but in the past 18 months concomitant chemotherapy with radiotherapy has been proven by randomized trials to improve survival. In general, the overall treatments are effective. The mean five year disease-specific survival for all patients is approximately 60%, but prognosis is dependent on T stage, N stage and site. For example, patients with risk factors of T4, N3, or site hypopharynx tend to do poorly regardless of treatment. The treatments are often associated with significant post-therapeutic short- and long-term complications such as dry mouth due to radiotherapy or the permanent loss of voice with surgery. The treatments for Head and Neck cancer can be very debilitating, have an incidence of mortality for those who complete treatments, and are contraindicated in some patients due to general health. Therefore a knowledge of the mix of patients in any data presentation is essential to understand the effectiveness of any reported treatment.

Patients with Head and Neck cancer constitute a different study population compared to other common cancer sites because the same habits that caused the cancers also cause many co-existent illnesses. In North America, squamous cell carcinoma of the Head and Neck is caused primarily by smoking cigarettes and by excessive alcohol consumption. Blot et al. reported a series of 1,114
patients with cancer of the oropharynx and found that 79% of patients with cancer of the oropharynx had smoked for over 20 years; only 104 patients claimed they had never smoked. It is well know that smoking causes multiple health problems. In the British Doctors Study \(^8\), the relative risk of death in male smokers compared to non-smokers for lung cancer, Chronic Obstructive Pulmonary Disease, Abdominal Aortic Aneurysm, and Atherosclerotic Heart Disease was 15, 12.7, 4.1 and 1.6 respectively, and for 'partial cause of death' for stroke and cancer of the esophagus the increased risk was 2.2 and 7.6. Alcohol is the other major etiological factor in North America for patients with squamous cell carcinoma. Deleyannis \(^9\) found that 45% of patients with Head and Neck cancer were or had been alcoholics, and Andre \(^10\) reported a case control series of 299 patients with Head and Neck cancer showing that the risk of developing cancer correlated with increasing amounts of alcohol consumed. Alcoholic patients have an increased risk of death from suicide, cirrhosis, hemorrhagic shock, pneumonia, lung cancer and digestive cancer \(^11\). It is not surprising that Piccirillo \(^12\) found that patients with Head and Neck cancer had the second highest level of comorbid illness compared to patients with gynecological cancer, breast cancer, lung cancer and rectal cancer. It is also not surprising that their co-existent illnesses have a measurable impact on survival. In previous work, Hall et al. \(^13\) quantified the risk by comparing the mortality data from a series of 655 new patients with Head and Neck cancer from the Kingston Regional Cancer Center to the expected survival data of age/sex matched residents of Ontario using data from the Canadian Institute of Actuaries. We estimated, at five years, that of those patients who died, 18% had died of co-existent disease associated with Head and Neck cancer, not of the cancer itself.

It is not just the patient habits and their comorbid illnesses that make Head and Neck cancer unique in Oncology. Unlike other common cancer sites, such as breast and lung, randomized clinical trials of traditional and common treatments of radiotherapy and surgery have never been done. Therefore, our treatment decisions are guided by large studies, commonly from single institutions and often reporting one type of treatment. These reports by design suffer from institution selection bias, treatment selection bias and patient selection biases. Institution selection bias occurs when an institution only offers select treatments, such as a center that is known for surgery or one that offers radiotherapy almost exclusively, although both might treat similar patients. In Head and Neck cancer, it is well known that different countries have very different overall treatment policies, as has been demonstrated by O'Sullivan et al. \(^14\) using an international survey of 160 Head and Neck oncologists from five countries; by Groome et al. \(^15\) using administrative data from Ontario and the SEER database from the USA; and by Hall et al. \(^16\) who compared treatments in Southeastern Ontario to Southeastern Norway. Patient selection bias occurs when patients refuse or choose their own treatments. In Head and Neck cancer this is an important consideration as patients will often refuse to have certain treatments such as disfiguring surgery, even when they understand that another option may not be as effective in curing the cancer. In any large series, some patients will have directed their own treatment and their comorbidity status could confound the results of treatment analysis. Treatment selection bias occurs when an effective treatment is not given to a patient for reasons such as age. This has been
demonstrated by Greenfield et al. who showed that older patients with breast cancer often do not receive the same investigations and treatments as younger patients. Treatment selection bias also occurs due to comorbid illness, such as the example of Mr JP but the influence of this bias in Head and Neck Oncology research has been overlooked. Elias et al. is one of few authors to acknowledge the potential influence of treatment selection bias. He reported a retrospective series of 101 patients with cancer of the hypopharynx that compared radiotherapy, surgery and planned sequential treatment using the outcome of three year overall survival. They found a very large and statistically significant difference between treatment types and stated "there is little doubt that the combination of surgery with radiotherapy gives the best results ... of course we cannot rule out the selection of patients on the basis of their general health influencing therapy choices". This was a statement of how selection bias contaminates results, specifically for studies that compare surgical patients who were well enough to have major surgery, to the radiotherapy patients that includes all those unfit for surgery. A similar statement is found at the end of a report of a national survey conducted jointly by the American College of Surgeons Commission on Cancer and the American Cancer Society on the treatments of cancer of the hypopharynx in the United States from 1987 to 1991. In this study, 576 hospitals provided information on 3,046 patients, making it the largest group of patients with cancer of the hypopharynx ever reported. The objective was to "identify general standards of care and offer insights into...the comparison of outcomes from differing treatments". They concluded that patients treated with surgery, or surgery with radiotherapy, had twice the survival of radiotherapy alone, but in the last sentence of the last paragraph of the conclusion they state "Because these studies were not randomized clinical trials selection bias likely accounts for some of the differences." It is likely that most readers of this large multicenter study sponsored by the American Cancer Society would conclude that surgery was better than radiotherapy for cancer of the hypopharynx. In summarizing the problem of ignoring the treatment selection bias of comorbidity, Piccirillo stated "this omission continues to produce major imprecision in the classification of patients and the subsequent interpretation of both 5 year survival rates and therapeutic effectiveness."

1.2 Study Objectives

Clinical Epidemiology has been defined as "the study of variation in the outcome of illness and of the reasons for that variation". To understand the variation in the outcomes in Head and Neck cancer, it essential to address treatment selection bias, and in the absence of randomized clinical trials, one option is multivariable analysis that controls for biases such as comorbidity. To control for comorbidity requires an measurement index that is valid, feasible and generalizable. The four indexes listed above are all extremely different. Some use a summative approach, others the most severe single illness. Two are continuous and two are ordinal. Two are restrictive (using select information) and two are general (using all information). Despite these differences, there are only 13 publications in which the performance of some of the indexes have been compared. In the Head and Neck oncology
literature there are only five studies that have published data using the indexes. Considering the problem of treatment selection bias and the high incidence of comorbid illness, the objective of this project was to determine which is the best index for the Head and Neck cancer patient population. The best index would have to valid both on its own and compared to the other indexes based on a thorough literature review, and would have to predict survival in the patient population. Added to the study was the Chronic Disease Scale, described by Von Korff in 1992 based on pharmacy data as a proxy for health status. The Chronic Disease Scale was chosen for its potential as a more objective assessment of medical burden than the abstraction of information from the medical record. It has not been used previously in clinical studies.

To design a study to identify the best index of comorbidity to predict survival for patients with Head and Neck cancer would require four components. First, to provide a source of the comorbidity information, a high quality, accessible, and consistent patient record had to be identified. Second, a process for rapid and accurate data abstraction and data entry had to be established. Third, all baseline data, treatment information, and follow-up status had to be available. Finally, criteria for the best index had to be established. This study, based on those four components, was designed with the following objectives:

- The primary objective is to determine the best comorbidity scale for retrospective survival studies in Head and Neck Oncology.
- The secondary objectives are:
  1. To compare the performance of a summative scale to a scale using the score from a single most severe illness;
  2. To compare the performance of a pharmacy based scale to a clinical record based scale.

To accomplish these objectives, the outpatient chart of the Kingston Regional Cancer Center (KRCC) was chosen, a PC-based data abstraction/entry module was developed, and prospective data from all patients from the Head and Neck clinic at the KRCC were utilized. Finally, it was decided that performance would be assessed as the ability to stratify patients on survival into three levels of comorbidity severity (low, medium and high) in both bivariate analysis (Kaplan-Meier method), and multivariate analysis (the Cox Proportional Hazards Model).

1.3 Overview

The text of this thesis begins with a review of the relevant literature in Chapter 2. This review was restricted to clinical publications and does not include a review of measurements in, or the problems associated with, administrative data. This study involves patients with cancer, and therefore more general scales such as the ASA (American Society of Anesthesiology) physical status scale that is commonly used for hospital-based outcome studies, or more specific scales such as the APACHEII (see glossary) that is used to stratify patient severity of illness in Intensive Care Units, are not reviewed.
Likewise the literature on Functional assessment scales such as the ECOG-PS (see glossary), Kamofsky, and the ADL is not reviewed. These scales are often used in clinical trials but are designed to be prospective and were not incorporated into the study design. The literature review begins with a critical appraisal of each the five indexes and is followed by a review of the 13 studies that compare the performance of the indexes. The final section in the literature review is a summary of the studies that measure comorbidity in Head and Neck Oncology. Chapter 3 describes the methods used to obtain and analyse the data, including the development of the electronic data entry system. The results are presented in Chapter 4, following the outline of the analysis method described in Chapter 3. Chapter 5, the discussion, combines the data analysis with the critical appraisal to determine which of the present five indexes is the best for the patient population. The secondary questions are addressed, followed by sections on the strengths, limitations and clinical relevance. Finally, some directions for subsequent research leading from this project will be described.
Chapter 2  

Literature Review

The objective of this chapter is to present a comprehensive review of the relevant literature, using a critical appraisal format, to determine the relative validity of five comorbidity indexes in the setting of survival analysis and Head and Neck Oncology. The following headings will be used:

i. Search Strategy

ii. Critical Appraisal of the five Indexes
   - The Cumulative Illness Rating Scale (CIRS)
   - The Kaplan-Feinstein Classification (KFC)
   - The Charlson Index
   - The Index of Coexistent Disease (ICED)
   - The Chronic Disease Scale (CDS)

iii. Comparisons of the five Indexes in the Literature

iv. Measuring Comorbidity in Head and Neck Oncology

v. Summary

2.1 Search Strategy

A comprehensive MEDLINE search was performed for publications on measurements of comorbidity between January 1987 and November 2000. The search was performed using three levels of specificity and utilized OVID (Electronic Information Retrieval System, OVID Technologies). The final results are summarized in Figure 2.1.

The initial search was comorbidity index specific. MeSH words for the named indexes did not exist and therefore the search consisted of the following keywords: 'Charlson Index', 'Charlson Comorbidity Index', 'Cumulative Illness Rating Scale', 'Kaplan Feinstein Index', 'Kaplan-Feinstein Index', 'Index of Co-existent Disease', 'Index of Co-existent Diseases', 'Index of Coexistent Disease', 'Index of Coexistent Diseases' and 'Chronic Disease Scale'. This search identified 93 articles. The abstract for each article was reviewed. Sixty-six articles were identified for the Charlson Index, four for the Kaplan Feinstein Classification (KFC), 21 for the Cumulative Illness Rating Scale (CIRS) and two for the Index of Coexistent Illness (ICED). The sum of the references by index was greater than the actual number of references (103 compared to 93) indicating that 10 articles referenced more than one index. All the publications on the CIRS, ICED and KFC were reviewed, except those pertaining to subjects such as administrative data methodology, cardiac surgery outcomes, homes for the elderly, or chronic psychiatric disorders. A subgroup of the 66 articles on the Charlson Index were reviewed that represented the spectrum of domains and settings.

A nonspecific search for comorbidity measurements was performed and consisted of the keywords 'comorbidity index', 'co-morbidity index', 'comorbidity scale', 'comorbidity score' and 'co-morbidity score'. The keywords 'comorbid illness scale', 'comorbid illness index', 'co morbidity index', 'co morbidity scale', 'co-morbidity scale' and 'co morbidity score' were tested but added no references. This search identified 113 references, 60 of which were not identified on the specific index search. Each of these 60 new abstracts was reviewed. The Charlson Index was cited in 12 references under names such as Charlson's Index and Index of Charlson. Other or unnamed
Figure 2.1
Search Strategy for Literature Review

180 abstracts identified

index specific 103
general index 60
comorbidity measures 17

112 abstracts reviewed

CIRS
21

KFC
7

Charlson
81

ICED
3

CDS
0

plus bibliographies

minus non relevant

9*
14*
21*
12*
4

* = includes review article by Exterman (70) not listed in tables
indexes were mentioned 44 times including "the weighted index", "the total illness burden scale", the Sartariano scale, the Duke Severity of Illness Checklist and the precursor to the ICED called The Comorbidity Index. The KFC was identified in two more references and one more relevant publication comparing indexes was found. The remaining abstracts did not identify the measurement used despite the keywords. In this nonspecific search, the most common study populations were dialysis patients, cardiovascular illness/surgery and prostate surgery/cancer, but the domains included a range of topics including HIV, breast cancer, ICU admissions, hip fracture, stroke, tracheostomy in children, anticoagulants, sub-dural hematoma, pancreatitis and diabetes. Some of the creative domains included the recurring visits to the ER, patients with swallowing disorders, and erectile function in castrated men. The constructs for the studies also varied and included treatment outcome, treatment variation, patient selection bias, complications, functional outcomes, and decision making. From the 60 references, 15 further publications were selected for review.

The third search was a general search for other comorbidity measurements and consisted of MeSH words 'comorbidity', 'co-morbidity', 'severity of illness index', 'health status index', 'prognosis', 'treatment outcome' and 'risk factors'. This search identified 17 more references. Each abstract was reviewed and two more references to the Charlson Index were found (the original article and the Deyo modification of the Charlson Index for administrative data). No other publications were felt to be relevant.

In summary, a careful review of 172 abstracts identified by a MEDLINE search plus articles identified in bibliographies was performed. When reviewed, the number of relevant papers was reduced to 21 papers on the Charlson Index, nine on the CIRS, 14 on the KFC, 12 on the ICED and four on the CDS. Of these 61, 13 are articles that compare two or more indexes, leaving 40 individual publications in this literature review. It is important to note that using the actual published names of these five indexes as keywords in a MEDLINE search did not identify the majority of the publications that used them.

### 2.2 The Critical Appraisal of the Five Comorbidity Indexes

In this section, each of the five indexes will be evaluated following a modified version of "A Methodological Framework for the Critical Appraisal of Classification Systems". Buchbinder et al. designed a generic approach for the critical appraisal of classifications and used it to specifically assess the classifications of multi-system clinical disorders, a setting not unlike the subject of this thesis. For this appraisal, content validity refers to the completeness of the scale to measure what it claims to measure, face validity is based on the question "is this a reasonable measure of the concept being measured", reliability refers to reproducibility of results, and construct validity is broadly defined for the clinical setting as how well the index stratifies in similar or different constructs. Construct validity in this thesis therefore includes criterion validity (how well the measurement performs in different settings). Indexes of comorbidity are designed to stratify patients to predict outcomes so the majority of
references are examples of predictive validity (how well does the index predict the outcome under study). Only one publication compared the total medical burden to autopsy findings (the gold standard) and therefore is the only identified example of 'pure' concurrent validity (section 2.2.1). Similarly there are a few examples of convergent validity (how well does the index perform compared to another measurement in a similar construct) comparing a comorbidity index to a known functional scale such as the Activities of Daily Living Scale. These are noted in the appropriate sections. When publications used specific examples of mini-theories, extreme groups and discriminative validity (ie does the index change with time?), these are noted. Relevant information on articles that compare the performance of two or more indexes that might normally be included under construct validity are found in Section 2.3. The critical appraisals for each index are summarized in a table within each subsection that also lists all the references in chronological order. The Summary section (2.5) compares the findings in critical appraisals to determine the index with the best validity based on the literature. This final assessment is summarized on Table 2.8.

The following format will be used to critically evaluate each of the five comorbidity indexes:

1. Original description and purpose:

2. Content Validity:
   a. What method was used for item generation and reduction?
   b. Is it a complete description of the concept or domain being measured?
   c. Are the categories mutually exclusive?
   d. What modifications have been made to the index?

3. Face Validity:
   a. Does it seem to be a reasonable measure?
   b. Are the elements clear and specific?
   c. If there are weights, are they sensible?
   d. Is the scoring sensible?
   e. Are the criteria dated?
   f. Is the score continuous, ordinal or nominal?

4. Reliability:
   a. What tests of inter and intra observer reliability have been reported?

5. Construct Validity:
   a. How does it perform in other settings or with other outcomes?

6. Feasibility:
   a. Is it simple to understand and use?
   b. Does it require extra training or specialized personal?
   c. Has time of administration been reported?

7. Summary, Strengths and Weaknesses:
2.2.1 The Cumulative Illness Rating Scale

The MEDLINE search identified 21 references for the Cumulative Illness Rating Scale (CIRS). Additional references were identified via bibliographies. The most relevant nine articles were reviewed. Table 2.1 lists the eight studies that are included in the discussion below. A review article by Extermann 26 is included in Section 2.5. A summary version of the CIRS-G is found in Appendix 1.

Original description and purpose:

The Cumulative Illness Rating Scale was reported by Linn et al. in 1968. 26 The CIRS was the first attempt to estimate or quantify the concept of overall physical impairment. In the original study, the scale was designed to estimate the "total medical burden" or the "capacity for survival in the elderly" in order to test the hypothesis that biological aging is different than chronological aging.

The CIRS is a multi-item summative predictive index consisting of 13 domains (13 independent organ systems), each with a 0-4 severity scale. The individual system severity scales are single item category scales that are summed to give a total numerical index to represent the overall medical burden. The final score is a continuous variable of increasing comorbidity. A patient can score from 0-56, and high scores (over 10) are associated with poor prognosis. The index is reported as a continuous variable and is usually a normal distribution skewed slightly right. The severity scoring is loosely defined and intentionally depends on the judgement of the physician examining the chart. For the severity scale, zero refers to no impairment, two represents an impairment that interferes with everyday life and requires treatment, and four is for life threatening illness with grave prognosis. If a patient has more than one disease or condition in one domain a judgement is required to determine the overall severity within that domain compared to all other patients. In its original format, the sum represented the score for total impairment, but in subsequent studies the mean of the involved items and the number of items with scores of three and four have been used. The CIRS was designed to be used by physicians to extract information from medical records.

The original CIRS study by Linn et al was based on chart extraction data from 172 male discharges from a Veteran's Hospital in the Southeastern USA in 1964. The objective of the study was to determine if older patients were more 'resistant' to illness. They compared in-hospital mortality and divided the study group into age groups of 55-65, 66-75 and over 75. They used the total sum CIRS score. They found that mortality correlated with the CIRS score and not with age, and therefore concluded that biological age was different than chronological age.

Content Validity:

The term content validity refers to the completeness and adequacy of an index to describe the concept or domain it is attempting to measure. For this discussion the original description plus the modification used in this thesis will be reviewed. The original presentation of the CIRS in 1968 describes the process of item generation and item reduction with the statement "modifications and revisions resulting from a series of preliminary computer analyses". It is essentially a grouping of
<table>
<thead>
<tr>
<th>Author</th>
<th>Date</th>
<th>Patient Population &amp; Setting</th>
<th>Number of Patients</th>
<th>Outcome</th>
<th>Statistical Groups</th>
<th>Distribution of Scores</th>
<th>Pearson Rho</th>
<th>Correlation with Pathology</th>
<th>Mortality and Morbidity</th>
<th>Post Discharge Mortality, Length of Stay, Pressure Sores</th>
<th>Normal Dist, Mean 6.7</th>
<th>Normal Dist, Median 4</th>
<th>Variation in Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miller</td>
<td>1992</td>
<td>Psychogeriatric outpatients</td>
<td>151</td>
<td>morbidity</td>
<td>mean 5.1</td>
<td>Various</td>
<td>&gt;8</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
</tr>
<tr>
<td>Conwell</td>
<td>1992</td>
<td>Completed suicides</td>
<td>72</td>
<td>morbidity</td>
<td>439</td>
<td>Various</td>
<td>0.75</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
</tr>
<tr>
<td>Parmalee</td>
<td>1995</td>
<td>Elderly multicare center</td>
<td>196</td>
<td>morbidity</td>
<td>362</td>
<td>Various</td>
<td>0.49</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
</tr>
<tr>
<td>Rochon(2)</td>
<td>1996</td>
<td>Spinal cord injured, hospitalized &amp;</td>
<td>1996</td>
<td>morbidity</td>
<td>203</td>
<td>Various</td>
<td>0.46</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
</tr>
<tr>
<td>Exterman*</td>
<td>1998</td>
<td>Geriatric cancer</td>
<td>1998</td>
<td>morbidity</td>
<td>500</td>
<td>Various</td>
<td>0.46</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
</tr>
<tr>
<td>Gunwitz</td>
<td>1997</td>
<td>Long term care residents with chronic atrial fibrillation</td>
<td>500</td>
<td>morbidity</td>
<td>500</td>
<td>Various</td>
<td>0.46</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
<td>Various</td>
</tr>
</tbody>
</table>

* article comparing the CIRS to other indexes of comorbidity
diseases or illnesses by independent organ systems created using the “judgmental approach”. All systems are included, but overlap is possible. A ‘miscellaneous’ domain is included for completeness. The categories for severity are general, as the index is meant to be based on judgement, not an exhaustive list of specific criteria.

The most important modification to the CIRS was by Miller, who created the CIRS-G version 27,28. Miller used a panel of specialists, to produce a 26 page manual in an attempt to “fine tune the appropriateness and reasonableness of each organ-system category” as well as the severity scales. He also added new items of hematopoetic disease and psychiatric disorders. A copy the manual for the CIRS-G was provided by the Department of Geriatric Psychiatry at the University of Pittsburgh and was used for this project. A summary version of the CIRS-G is found in Appendix 1.

Parmalee et al. 29 added an item for hypertension, modified some of the items and further defined some of the scoring for some items. Parmalee found, using the Cox Proportional Hazards model, that some items, such as deafness, blindness and major musculoskeletal problems actually had a negative impact on mortality, but she did not suggest modifying the scale. Hypertension was one of the items that they added using their judgmental approach, only to find in the end that it did not influence mortality.

Face Validity:

The term face validity is based on the question “is this a reasonable measure of the construct?”, and therefore refers to the details of items and the scoring. What the CIRS lacks in item generation and/or item reduction is gained in sensibility. Sensibility, or “enlightened common sense”, is defined as “a mixture of ordinary common sense plus a reasonable knowledge of pathophysiology and clinical reality” 30. The CIRS is simple and comprehensible - a summative score, using a severity scale, from each independent organ systems, with increasing numbers meaning a greater medical burden. It would seem to be a reasonable measure given the overall complexity of total medical burden. When the Miller adaptation is used, there are detailed descriptions for each overall severity level and specific descriptions of the criteria within each domains. In the Miller version, under the item of peripheral vascular disease, claudication is assigned severity two and bypass surgery is assigned severity four.

For the item liver disease, categories can be assigned by history, laboratory values or alcohol consumption based on history.

There are however three limitations due to the categories of severity - limitations that in fact apply to all of the indexes that use severity scales. First, the scales and scoring for severity within items have never on their own been validated. For example, Miller classified the burden of cancer into four levels. Level 2 refers to cancer in the past without recurrence for five years and Level 3 means treatment in the past five years. Although these levels reflect the same philosophy as the overall scale, in reality they may not be valid due to different types and behaviours of cancers. The second problem is that the CIRS weights all of the items and scales the same. Does a poorly controlled insulin dependent diabetic (level 4) have the same medical burden as an institutionalized psychiatric patient (level 4)? Does a patient with both mild renal failure (level 2) plus stable diverticular disease (level 2)
have the same prognosis as a level 4 patient on home oxygen or intractable congestive heart failure? Finally, some severity category labels within different domains are assigned by different processes. Examples include historical evidence, such as the ability to read fine print, radiological evidence, laboratory values and assumed diagnoses such as diverticulitis.

There is an interesting misprint in the two original articles by Linn et al. The second article, which actually describes the scale, and is the most common reference to the CIRS, refers to the previously reported study as the validation study of 472 patients. This was incorrect as the study actually had only 172 patients!

The CIRS is a continuous variable and therefore must be divided by cut-points for survival analysis. It was designed as a measure of overall health status or total medical burden, but by removing the disease of interest for the sum has become an index of comorbidity.

Reliability:
Reliability is defined as "the extent to which repeated measurements of a stable phenomenon - by different people, at different times and places - get similar results", and refers to the reproducibility of the results within and between observer test events. The five publications below tested the CIRS for reliability. The studies will be described in subsequent sections. In the original Linn study of 172 veteran hospital discharges, no information is given on who actually extracted the data or what patients were used for the inter-rater reliability. Three residents and three physicians performed independent ratings on 20 patients. They used Kendall's W and claimed high "consistency" between raters (all values were over 0.83), but tests for agreement and chance were not done. It is not clear what measure of the CIRS was even used as they refer to "ranked ratings" not the sum. The Miller study used the ICC test to examine the inter-rater reliability for both the summative score and the number of categories with high scores. In 10 outpatient charts assessed by two trained research nurses, the ICC was 0.78 for the summative score and 0.81 for the number of categories with high scores. For 10 inpatient charts, tested by three physician assistants and two physicians, the ICC for the summative score was 0.88 and for the number of high categories was 0.83. The tables and data are included in the paper. In Conwell's autopsy study, two physicians rated the CIRS by chart review. It is stated that there was no statistically significant difference between the raters (0.3) but no details of the statistic test, the reviewers, the charts or the score used are provided. Rochon used a second experienced chart reviewer to extract CIRS scores from 30 random charts of the 362 patients with spinal cord injuries. The inter-rater reliability of 0.8 was calculated using the Spearman coefficient for the summative score. Extermann's study of elderly patients with cancer also used the ICC for two raters. They found inter-rater reliability for the summative score was 0.76 and the test-retest reliability was 0.95.

Construct Validity:
How well an index responds in an other construct, another setting or with another outcome is the broad definition of construct validity used in this discussion. Nine publications are of relevance in this section.
Miller et al. \textsuperscript{27} tested the CIRS in a study of psycho-geriatric outpatients in the Northeastern USA in what was primarily a study of construct validity. The objective of the study was to demonstrate the feasibility and reliability of the CIRS to provide a quantitative rating of medical burden for their patient population. They collected data from the outpatient charts of 40 geriatric general medical clinic patients, 45 depressed patients, 21 spousally bereaved patients and 35 healthy elderly controls. They derived five different scores from the CIRS, including the sum, the mean of the number of identified items, and the number of items with scores over two. They found that the CIRS was higher in the geriatric medical clinic population than in all other groups, and concluded that the CIRS could be used to both rate and track medical problems in elderly depressed patients. One hypothesis was that the extreme groups (healthy elderly controls vs. multidisciplinary geriatric medical) would have different scores, with the scores of the depressed patients falling in between the two. The mean score for the medical patients was 10.6, for the controls it was 4.5 and for the depressed patients it was 5.2. Another mini-hypothesis was that the scores of younger patients at the medical clinic would be lower, which they were. One year later they collected CIRS data on 70 of the original patients and found the CIRS scores had increased in the controls and in the depressed groups as expected (discriminant validity).

Conwell et al. \textsuperscript{32} compared the CIRS scores (obtained by a combination of chart review and extensive interviews) to the findings at autopsy for 72 completed suicides to validate the CIRS as an objective measure of illness burden. This study is an example of concurrent validity. The clinical CIRS score was established from charts, family interviews, and physician interviews, and the autopsy results (the gold standard) were graded in a similar way to the CIRS with a 0-4 severity scale for each of the 12 systems (psychiatric item could not be included) to create a 'pathological' CIRS. They used the summative score and found that the clinical CIRS was a predictor of the pathology CIRS (Rsq =0.74, p<.0001). The clinical CIRS tended to overestimate when low levels of disease were present and underestimate when disease burden was high, but the differences were small. They concluded that the CIRS was a valid measure of objective physical illness burden.

Rochon et al. \textsuperscript{33,34} studied 362 discharges from a spinal cord rehabilitation hospital. The CIRS was extracted from the chart information at the time of admission, was reported as the summative score, and was compared to mortality up to 18 months after discharge. They found a normal distribution of scores with a slight right skew and mean 6.7 (SD 2.9). Using linear regression models, they found that the CIRS correlated with both mortality and length of stay. Rochon et al. \textsuperscript{34} used the CIRS to identify risk factors for the development of pressure sores. A CIRS score over eight was an independent predictor in the logistic regression model (OR=3.7) along with other factors such as nutrition, level of consciousness and immobility.

Parmalee et al. \textsuperscript{29} used a population of 439 elderly Jewish patients from a large multicare center in Northeastern USA to investigate the CIRS as a predictor of morbidity and mortality. The score was extracted from the inpatient medical record and the scoring was slightly modified as noted above. They used the mean score of the 14 items and the number of items with scores over two instead of the
sum. The raw data or summative scores were not published. The scores were divided into tertiles for multivariable analysis and correlated with 1-2 year post assessment mortality and morbidity (hospitalization). They concluded that the CIRS was a valid indicator of health status. The correlation between the two scoring systems was 0.89.

Extermann et al. reported the CIRS scores for a series of 203 older cancer patients with a spectrum of primary cancers and mean age of 75 years. The CIRS-G version was used. The CIRS sum scores had a normal distribution (median 4). Other strata reported were mean scores, the number of items with scores over three and the means of the items across the study cohort.

The CIRS has been used to control for comorbidity in a variety of other settings such as laminectomy for spinal stenosis, and assessment of the appropriateness of drug treatments in elderly patients. In the study of stroke prevention using a group of patients with mean age 84, the continuous sum scores CIRS scores were divided into tertiles of <11, 11-14, >14. In the study of the outcome of laminectomy, the mean age was 69.3, all had surgery and the tertiles were defined by 0-2, 3-5 and >5.

The CIRS has been compared to other indexes measuring functional status. Miller correlated the CIRS with an index of capability, the Older American Activities of Living Scale (OARS-ADL). The correlation was 0.58 using a group of general medical outpatient clinic patients. Parmalee found no correlation between the CIRS and the Physical Self-Maintenance scores in elderly patients. Extermann compared the CIRS to the ECOD-FS (a functional assessment score), the ADL and the IADL in elderly patients with cancer. No correlation was found. These studies confirm that the CIRS, that was not designed as a measure of functional status, does not assess functional status.

**Feasibility:**

The classification is easy to understand, abstract, perform and calculate. It requires judgement for many items, so an experienced and trained abstractor is essential.

**Summary:**

The original scale by Linn was created in 1968 to be a simple overall measure of the complex phenomenon of overall medical burden. Its assumption was that physicians, "by virtue of their training and orientation ... everyday make reasonably accurate, uniform and global judgements" of patient’s health or system impairment. It was designed to convert those general judgements to a very general and simple scale. Because it was created to use judgement, and not explicit items, the scale suffers from problems with content validity and within-item face validity. It was not originally designed to be a measure of comorbidity, as comorbidity studies remove the index disease/disorder from the scale and assume it is still valid. With the Miller revision some of the content problems were resolved, construct validity was demonstrated, and reliability was tested appropriately. It is because of the overall sensibility and simplicity of the CIRS that it has been used as an index of comorbidity. As seen in Table 2.3, the scores and best cut-points vary with the patient population. A small number of studies have shown that the CIRS scores predict survival. It is the only index to have a concurrent validity study compared to the gold standard of autopsy findings. The CIRS has been successfully utilized in many
studies involving geriatric patients, but its validity in an Oncology setting has not been described. The CIRS might be a useful scale for measuring comorbidity in the Head and Neck cancer population because the patients tend to have multiple illnesses due to smoking and alcohol abuse. A study with high quality, consistent data on medical burden should be designed to test it.

2.2.2 The Kaplan Feinstein Index

The MEDLINE search identified seven articles citing the Kaplan-Feinstein Classification (KFC) and another seven studies were found via bibliographies. All fourteen studies were reviewed. Nine of these studies are listed on Table 2.2. The five studies that utilized part of or a modified version of the original scale are reviewed in the text, but are not found in the summary Table.

Original Description and Purpose:

The Kaplan-Feinstein Classification of comorbidity was published in 1974. The KFC is a single item ordinal scale derived from the analysis of disease in 12 systems. The domains include hypertension, cardiac, central nervous system, respiratory, renal, hepatic, gastrointestinal, peripheral vascular disease, malignancy, locomotor, alcoholism and a miscellaneous group (Appendix 2). Each item has a categorical severity scale of Grades 0 to 3 with criteria that are explicit for that domain. These categories are used to describe the coexistent co-morbidities ("co-morbidities that might be expected to impair a patient's long term survival"). The presence of comorbid disease at a defined severity within any item and the item with the greatest severity defines the final score. The score can be reported numerically as Grade 0 to 3 or by the text words of none, moderate or severe. None refers to the absence of comorbidity in any system, moderate includes the presence of any Grade 1 or any Grade 2, in any system and severe means the presence of Grade 3 illness. A patient with two or more Grade 2 comorbidities is moved up to Grade 3. The Grade 3 or Severe group has also been labelled 'Prognostic Comorbidity'.

The KFC was developed using a series of 188 men with newly diagnosed diabetes mellitus. The objective of the study was "to demonstrate that co-morbidity is a crucial confounding variable which has been omitted from most statistical analyses". They found that increases in the index were associated with a higher incidence of mortality at five years after diagnosis. In the creation of the KFC, diabetes was the disease of interest, so was not included in the scale, and therefore in subsequent studies using the KFC, diabetes has been added with appropriate severity levels. In the original description the distribution of scores was not reported.

Content validity:

The KFC was developed "after years of experience with the concept of prognostic comorbidity". A prognostically comorbid illness is "one that predisposes, either by itself or in combination with the main disease, to the future development of adverse target events". One reference described Dr. Feinstein keeping a list of such illnesses over many years. The items and severity scales were,
<table>
<thead>
<tr>
<th>author</th>
<th>date</th>
<th>patient population &amp; setting</th>
<th>number of patients</th>
<th>outcome</th>
<th>distribution of scores (Gr 0,1,2,3)</th>
<th>statistical groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clemens</td>
<td>1986</td>
<td>prostate cancer</td>
<td>280</td>
<td>mortality</td>
<td></td>
<td>Gr 0&amp;1, 2, 3</td>
</tr>
<tr>
<td>Charlson*</td>
<td>1987</td>
<td>breast cancer</td>
<td>685</td>
<td>10 yr mortality</td>
<td>550,36,56,43</td>
<td>Gr 0,1&amp;2,3</td>
</tr>
<tr>
<td>Pompei</td>
<td>1988</td>
<td>hospitalized medical ward</td>
<td>559</td>
<td>post discharge mortality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waite*</td>
<td>1994</td>
<td>discharge after emergency admission</td>
<td>79</td>
<td>readmission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krousel-Wood*</td>
<td>1994</td>
<td>prostatectomy</td>
<td>302</td>
<td>5 yr mortality</td>
<td>21%,37%, 23%, 18%</td>
<td>0&amp;1, 2&amp;3</td>
</tr>
<tr>
<td>Albertsen*</td>
<td>1996</td>
<td>prostate cancer</td>
<td>451</td>
<td>mortality</td>
<td></td>
<td>Gr0, 1&amp;2, 3</td>
</tr>
<tr>
<td>Newschaffer*</td>
<td>1997</td>
<td>breast cancer</td>
<td>404</td>
<td>mortality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singh* (2)</td>
<td>1998 &amp; 1999</td>
<td>young Head &amp; Neck cancer</td>
<td>70</td>
<td>mortality</td>
<td></td>
<td>0&amp;1, 2&amp;3</td>
</tr>
<tr>
<td>Piccirillo</td>
<td>2000</td>
<td>Head &amp; Neck cancer</td>
<td>203</td>
<td>2 yr mortality</td>
<td>55%,24%, 16%, 18%</td>
<td>Gr0,1&amp;2,3</td>
</tr>
</tbody>
</table>

* article comparing the KFC to other indexes of comorbidity
however, created by the “judgmental approach” of one individual and there has been no formal or reproducible item generation/item reduction process. The KFC does not include newer diseases, diagnoses or major treatment advances since 1974. Like the CIRS, the severity categories have not been tested and may not be comparable across different items.

A modified version of the KFC has recently been used in reports by Piccirillo. Items were added such as AIDS and dementia, to update or complete the original KFC. In 1994 he began a process of prospective coding comorbidity in all newly diagnosed cancer patients at his medical center. He reported on 3378 patients with lung, breast, gynecological, colorectal, prostate and Head and Neck cancer. The distribution of the modified KFC levels (mKFC) for each cancer site are reported, and showed, for example, that 21% of patients with Head and Neck cancer had moderate or severe comorbidity. Patients with Head and Neck cancer had the highest mean comorbidity compared to all other patients except those with lung cancer. Unfortunately, the details of the criteria for the modified version of the KFC are not reported or available.

Face Validity:

Like the CIRS, what is lost without formal item generation and item reduction in the KFC is gained in sensibility. A severity index is applied across 12 domains with explicit criteria and the single highest severity score is the final score (plus an adjustment for multiple illnesses). Like the CIRS, the severity categories have not been tested and there is an untested assumption of equal weights.

The combination of two or more Grade 2s to become Grade 3 is sensible, as it addresses the problem of multiple comorbidity when an index is not summative. However, it makes the index ordinal when most indexes are thought to be continuous, so assumptions of linearity and the assumptions for Cox proportional hazards models may not be valid.

The scale was created based on male patients and the now dated knowledge of medicine in 1974.

Reliability:

There are only two references to reliability for the KFC. These studies are described in more detail in subsequent sections. In the study of visits to the Emergency Department for admission after discharge from the hospital, Waite et al. found 50.6% agreement for KFC across five abstractors and the ICC was 0.82. Newschaffer assessed inter-rater reliability using a second reviewer and the weighted kappa was 0.82.

Construct Validity:

The following 11 publications used the KFC in outcome analysis and are relevant to this thesis. Ten were survival studies and one examined the use of the KFC for predicting readmission to the hospital.

Clemens et al used the KFC as the comorbidity scale in the development of a Clinical Anatomic Staging System for prostate cancer. The records of 280 patients with complete follow-up were examined. The study involved correlating and combining signs, symptoms, comorbidity and
clinical TNM stage into one scale (also known as a clinical severity index). The actual numbers per group are not reported. The five year survival declined with increasing comorbidity and statistically separate survival curves were created when the Grade 0 and Grade 1 were combined.

Charlson et al. in the validation portion of the original paper on the Charlson Index, also used KFC to measure comorbidity on the cohort of 685 patients with breast cancer followed for 10 years. The study population had a low incidence of comorbidity as 80% were Grade 0 and only 5% Grade 1. The four KFC Grades split into three distinct survival curves (Grades 1 and 2 overlapped) using both Kaplan-Meier methods and Cox proportional hazard regression models.

Pompei et al. coined the term “susceptibility bias” to introduce the problems of comorbidity and accounting for comorbid illnesses in clinical research. They examined the records of the same general medical patients used for the creation of the Charlson Index. One year follow-up was available for 599 patients. They used the KFC to estimate comorbidity status and added AIDS as a Grade 3 disorder. The univariate analysis showed declining one year survival with increasing KFC scores.

Krousel-Wood et al. tested the KFC as a predictor of survival in a series of 302 patients after prostate surgery. The distribution of the scores for Grades 0-3 was 21%, 37%, 23% and 18%. They found that five year mortality increased with the score (Grades 0 and 1 were similar). The area under ROC curve for predicting survival was 0.75. When Grades 0 and 1 were compared to Grades 2 and 3, using Cox proportional hazards model and controlling for both age and procedure, the relative risk of five year mortality was 4.23 (p<.0005).

“A new clinical severity staging system for cancer of the larynx” was published by Piccirillo et al. in 1995. In fact, this was not a new system at all - it was an exact replica of both the methodology and results of study published in 1977 by Feinstein. Clemens et al. used the same methodology for a study of prostate cancer as noted above. The objectives were to validate the previously published system that combined comorbidity, TNM Stage, signs and symptoms into the clinical severity index. They used the data from 193 patients with cancer of the larynx from a single institution that had five year follow-up. All patients who were treated were included. Time zero was the date of first treatment. Only Prognostic Comorbidity (KFC Grade 3) was recorded, and the distribution data for the other Grades were not included. Fourteen percent of patients had Prognostic Comorbidity (Grade 3). They found that 74% of patients without Prognostic Comorbidity survived five years compared to 15% with Prognostic Comorbidity. They also demonstrated that the clinical severity index outperformed the TNM staging system for predicting survival.

The same methodology was subsequently used by Pugliano et al. to validate the clinical severity index in 296 patients with cancers of the oropharynx and 277 patients with cancer of the oral cavity. Unlike the previous studies, they defined patients who were “too sick” to tolerate treatment as Grade 3. Eleven percent of the oropharynx cancer patients and 7% of the oral cavity cancer patients were Grade 3. They found for the oropharynx patients that the five year survival of those with Grade 3 was 18% compared to those with less than Grade 3 whose survival was 40%. Similarly, for oral cavity
cancer patients, the five year survival for Grade 3 was 10% and when less than Grade 3 was 49%.

Albertsen et al. evaluated the use of the KFC to stratifying patients with prostate cancer on survival. Data on 451 men with localized prostate cancer were studied. The distributions are not included. Using Kaplan-Meier curves, the KFC split patients into three strata for both ‘all cause death’ and ‘non-cancer caused death’ (the curves for Grades 1 and 2 overlapped).

Newschaffer et al. examined the usefulness of the KFC in the survival analysis of 404 patients over 65 with breast cancer. They used the Cox proportional hazards regression with age, stage, and surgical treatment with the outcome of all cause death. Unfortunately, the distribution of scores was not reported and follow-up was less than five years for all patients. They used the KFC scores as a continuous variable and found that it did not produce a statistically significant increase in the relative risk of death. They felt that the failure of the KFC was due to its need for specific or detailed information. This study is noted again in Section 2.3.

Singh et al. used the KFC to control for comorbidity in a survival study of patients under 45 years of age with squamous cell carcinoma of the Head and Neck. Seventy patients were identified, 25 of which had HIV, and they used information obtained up to six months after diagnosis. Patients with unresectable disease or deemed “too sick” for treatment, were excluded from the study. When two groups (Grades 0 plus 1 and 2 plus 3) were created they found differences in survival (Kaplan-Meier and Cox regression), differences in the incidence of major complications and differences in cancer relapse. They did not find any difference in the actual treatments offered to patients when stratified by KFC score. This is an important study for this thesis because it is one of the five published studies in Head and Neck cancer that controlled for comorbidity and that used survival as the outcome of interest. Unfortunately, the problems of selection bias (patients with advanced disease were excluded), inclusion of comorbidity data obtained after time zero, the unusual patient mix with a very high incidence of AIDS, and very short follow-up (median 13 months), reduce its usefulness, generalizability and validity.

Piccirillo used the KFC, with modified items as noted in the Content Validity Section, to study survival in patients with Head and Neck cancer. The distribution of Grades 0-3 over 341 patients was 55%, 24%, 16% and 18%. Two year survival data were reported for 203 of the patients. The Kaplan-Meier curves for Grades 1 and 2 overlapped. Follow-up was too short, and the sample size was too small, for the results to be significant on multivariable analysis. This is an important study for the critical appraisal of the KFC in the setting of Head and Neck cancer, but the lack of definitions of modifications, incomplete follow-up, and short follow-up, reduce its usefulness.

Unlike the other indexes in this project, the KFC has rarely been utilized beyond oncology studies. Waite et al. reported a study on the risk of hospital readmission to the General Internal Medicine service that compared the performance of the KFC to other indexes in predicting ‘at risk’ patients. They reviewed 79 patient records and 40 controls from a Veterans’ hospital. There were 39 patients who had at least one readmission within six months. The KFC, like all the indexes tested, was of no predictive value.
Feasibility:

The KFC is simple to understand, abstract, perform and calculate. The abstractor needs to trained, but training is less than in some scales such as the CIRS, as the items are more explicit. The score is easy to calculate. Time for completion according to Waite was 8.9 minutes.

Summary

The KFC is a single item ordinal scale with 12 items and a severity scale within each. The final score is the highest score of all the items. The KFC has been successful in its aim “to demonstrate that comorbidity is a confounder” as it does stratify patients on mortality. It has been used almost exclusively in oncology and by disciples of Dr Feinstein. Its weaknesses include inadequate content validity and face validity, and it is an ordinal scale. Its performance has shown that, although the KFC does stratify patients on survival, in no instance were the four original Grades statistically useful (Clemens, Charlson, Piccirillo and Krousal-Wood, Albertson all combined at least 2 Grades). Furthermore, in every publication, the combinations of Grades are different (see Table 2.2). The two survival studies in the breast cancer population by Charlson and Newschaffer had conflicting prognostic results with Charlson finding 4 statistically distinct strata and Newschaffer finding no relationship. In the Head and Neck cancer population, the Singh study is biased and the follow-up of the patients in the Piccirillo study is too short to be useful. The KFC may be a useful index in Head and Neck cancer, but a study with complete follow-up and without selection bias would be required to demonstrate its utility.

2.2.3 The Charlson Index

As noted in section 2.1, the MEDLINE search identified 81 references for the Charlson Index across a wide range of settings. In fact, of the 156 articles that were identified using a named index, the Charlson was featured in over 50%. Thirty articles were selected based on relevance to the overall issues of comorbidity, oncology, a wide range of study constructs, or comparisons with other indexes. The numerous papers on administrative data, costs of treatment, and drug tests were excluded. From the 30, a selection of the 21 most relevant (Table 2.3), chosen as they represent the spectrum of uses through the relevant literature, are reviewed briefly below. The publications specifically on Head and Neck Oncology are reviewed only in section 2.5.

Original Description and Purpose

The Charlson Index was published in 1987 as a “taxonomy for comorbid conditions that might alter the risk of short term mortality” 4. It is a multi-item summative scale and consists of a list of 16 specific, predefined medical conditions. Each of the 16 conditions is weighted and each is explicitly defined with a series of directives. The final score (the sum) is a continuous variable of increasing comorbid illness, and a patient can have a result of 0-31. The scores are typically skewed far right; in
Table 2.3
Relevant studies that have used the Charlson Index to measure comorbidity (1987-2000)

<table>
<thead>
<tr>
<th>author</th>
<th>date</th>
<th>patient population &amp; setting</th>
<th>number of patients</th>
<th>outcome</th>
<th>groups studied</th>
<th>distribution of scores</th>
<th>statistical groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlson* (test set)</td>
<td>1987</td>
<td>medical admissions</td>
<td>607</td>
<td>1 yr mortality</td>
<td>0,1-2,3-4,&gt;4</td>
<td>181,125,71,82</td>
<td>0,1-2,3-4,&gt;4</td>
</tr>
<tr>
<td>Charlson (validation set)</td>
<td></td>
<td>breast cancer</td>
<td>685</td>
<td>10 yr mortality</td>
<td>&quot;</td>
<td>588,54,25,18</td>
<td>0,1-2,3-4,&gt;4</td>
</tr>
<tr>
<td>Rabeneck</td>
<td>1993</td>
<td>acute pancreatitis</td>
<td>162</td>
<td>mortality</td>
<td></td>
<td></td>
<td>0,1-2,&gt;2</td>
</tr>
<tr>
<td>Waite*</td>
<td>1994</td>
<td>readmission after medical discharges</td>
<td>74</td>
<td>readmission</td>
<td></td>
<td></td>
<td>mean 2.1</td>
</tr>
<tr>
<td>Krousel-Wood*</td>
<td>1996</td>
<td>prostatectomy</td>
<td>302</td>
<td>mortality</td>
<td>0,1,2,&gt;2</td>
<td>54%,23%, 12%,11%</td>
<td>0-1, &gt;1</td>
</tr>
<tr>
<td>Rochon* (2)</td>
<td>1996 &amp; 1994</td>
<td>spinal cord injured, hospitalized</td>
<td>362</td>
<td>mortality and length of stay, pressure sores</td>
<td>74% +0</td>
<td></td>
<td>&gt;2</td>
</tr>
<tr>
<td>Albertsen*</td>
<td>1996</td>
<td>prostate cancer</td>
<td>451</td>
<td>mortality</td>
<td>0,1,2,&gt;2</td>
<td></td>
<td>0,1,2,&gt;2</td>
</tr>
<tr>
<td>Poses</td>
<td>1996</td>
<td>patients in intensive care unit</td>
<td>201</td>
<td>mortality</td>
<td></td>
<td></td>
<td>continuous</td>
</tr>
<tr>
<td>Lawrence</td>
<td>1996</td>
<td>abdominal surgery</td>
<td>82</td>
<td>post-operative chest complications</td>
<td></td>
<td>mean 2.7</td>
<td>continuous</td>
</tr>
</tbody>
</table>

* article comparing the Charlson Index to other indexes of comorbidity
<table>
<thead>
<tr>
<th>Author</th>
<th>Date</th>
<th>Patient Population &amp; Setting</th>
<th>Number of Patients</th>
<th>Outcome</th>
<th>Groups Studied</th>
<th>Distribution of Scores</th>
<th>Statistical Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newschaffer</td>
<td>1997</td>
<td>breast cancer</td>
<td>404</td>
<td>mortality</td>
<td>0,1-2,3-4,&gt;4</td>
<td>mean 2, 28% &gt;4</td>
<td>continuous</td>
</tr>
<tr>
<td>Singh* (2)</td>
<td>1997 &amp; 1998</td>
<td>young Head &amp; Neck cancer patients</td>
<td>70</td>
<td>mortality</td>
<td>0,1-2,3-4,&gt;4</td>
<td>mean 2.7</td>
<td>0-2, &gt;2</td>
</tr>
<tr>
<td>Polancyk</td>
<td>1998</td>
<td>congestive heart failure</td>
<td>2320</td>
<td>mortality</td>
<td></td>
<td>mean 1.24</td>
<td>continuous</td>
</tr>
<tr>
<td>Kiefe</td>
<td>1998</td>
<td>women in a general practice</td>
<td>1764</td>
<td>cancer screening</td>
<td></td>
<td>0,1,2,&gt;2</td>
<td>0,1,2,&gt;2</td>
</tr>
<tr>
<td>Newschaffer</td>
<td>1998</td>
<td>breast cancer</td>
<td>1948 control &amp; 3310 cancer</td>
<td>survival</td>
<td>0,1,2,&gt;2</td>
<td>breast (83%, 10%, 5%, 2%) control (73%, 16%, 7%, 4%)</td>
<td>0,1,2,&gt;2</td>
</tr>
<tr>
<td>Singh</td>
<td>1999</td>
<td>H&amp;N cancer patients having major surgery</td>
<td>200</td>
<td>complications and length of stay</td>
<td>0,1&amp;2,&gt;2</td>
<td>131,45,24</td>
<td>0,1&amp;2,&gt;2</td>
</tr>
<tr>
<td>Gabriel x2*</td>
<td>1999</td>
<td>arthritis patients</td>
<td>891</td>
<td></td>
<td></td>
<td>64% =0</td>
<td>continuous</td>
</tr>
<tr>
<td>Extermann*</td>
<td>1999</td>
<td>geriatric cancer</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sabin</td>
<td>1999</td>
<td>larynx cancer</td>
<td>152</td>
<td>survival</td>
<td>0,1-2,3-4,&gt;4</td>
<td>83% =0, 17% =4</td>
<td>1&amp;2,3&amp;4</td>
</tr>
</tbody>
</table>
most series about 80% of patients have a score of zero. A score over five is considered high and is usually associated with a poor prognosis. The Charlson Index is the most widely used index of comorbidity and might be considered by some to be the gold standard at present. It has been found to correlate with mortality in many studies.

Content validity:

Charlson created the index using the 'training set' of a series of 607 patients admitted to a general medical service in a large New England teaching hospital during one month in 1984. The illnesses present at the time of admission were recorded at the time of discharge and mortality was analysed with one year follow-up. The most prognostic illnesses were identified using Cox regression, and then weights were assigned using the relative risks as a guide. The final items are mutually exclusive, and, as they are clearly defined, are specific (Appendix 3). In the same publication the Charlson Index was validated using a 'test set' of 685 patients with cancer of the breast from one institution treated from 1962 to 1969. Ten year follow-up data were obtained for 681 patients. The Charlson Index scores were extracted retrospectively from medical records. They found that 86% of patients had a score of zero and that the survival curves split into four statistically separate strata of comorbidity (scores 0, 1, 2, >2).

Two important modifications of the Charlson Index have been published. Deyo et al. adapted the Charlson Index to the ICD9-CM codes for use in administrative data including hospital separations. This methodology has been adopted widely and is referred to as the Deyo Index or the Deyo modification. They tested the new index using the records from 27,111 Medicare patients who had lumbar spine surgery in 1985. The distribution of the scores showed the typical right skew, with 83% of patients having zero comorbidity. They used a cut-points of 0, 1, 2 and >2, and showed that complications, six week mortality and length of stay all increased with increasing scores.

The second modification was published by Charlson et al. and consisted of adding a score per decade of life to the original index to create a combined age plus comorbidity index. This had, in fact, been reported within the original publication, where it was stated that any study with follow-up greater than five years should combine age with comorbidity status. Charlson validated the modified index using a cohort of 207 patients from a general surgery service, between 1982 and 1985, who had either diabetes or hypertension. This was a curious choice of diseases as under the Charlson Index hypertension is not, but diabetes is, a predictor of mortality. These patients had been involved in a study of post-operative complications and had been thoroughly investigated pre-operatively. Using the Cox Regression, it was found that each decade of life over 40 was equivalent to adding 1 unit to the original index. The Kaplan-Meier curves for five year survival divided into three groups using the age plus original index with cut-points of 0-4, 5-7 and >7. The Cox regression analysis showed that the point estimate of the relative risk of death increased progressively with increasing scores and that the age plus index score was statistically significant (p<.0001). Unfortunately, the 95% confidence limits all overlapped.
Face Validity:

Given the methodology that produced the list of prognostic illnesses and the summative approach, the Charlson Index would seem to be a reasonable instrument. However, there are four problems with face validity:

1. Some of the items are not homogeneous on survival. For example, patients “solid tumors” of different stages and different sites have different prognoses.

2. The weights were assigned using the relative risk estimates of the Cox output (1 for 1.2-1.5, 2 for 1.5 to 2.5, 3 for 2.5 to 3.5 and 6 for everything else), but the assignments are not justified and have not been tested.

3. The list and weights were created on medical patient admissions to one hospital in one month in 1984, and it not surprising that the generalizability of the index has been questioned by many authors.

4. The scores are invariably skewed right, often with over 80% of patients having a score of zero. This floor effect means that only a small, highly select group of patients are available for statistical analysis and often reducing the ability of the some groups to achieve statistical significance.

The Charlson Index is a continuous variable and therefore must be divided by cut-points for survival analysis.

Reliability:

In the 30 studies reviewed for this thesis, only three studies reported reliability testing. These three are reviewed in more detail in the next subsection. The study by Waite et al. used hospital readmission data, showed the Charlson Index (compared to the KFC and the ICED) to have the highest absolute agreement rate over five observers (58.2%) and the highest ICC (0.93). Ninety-eight percent of the time, the absolute agreement was within 1 Charlson unit for the 5 observers which was the highest score among all the indexes tested. Extermann reported test-retest reliability of 0.86 and ICC of 0.74 for two abstractors. Newschaffer assessed inter-rater reliability using 40 charts using two abstractors. The weighted kappa was 0.945.

Construct Validity:

Eleven publications are reviewed in this section. The four studies using Head and Neck cancer patients are reviewed in Section 2.4. Although the majority are survival studies, there are examples of the Charlson Index being used to predict readmission to the hospital, complications of surgery and medical office visits. The studies are in chronological order by publication date.

Rabeneck et al. used the Charlson Index as part of a predictive model for the survival of patients with acute pancreatitis. They reviewed the charts of 162 patients from a single institution. The score was calculated at the time of diagnosis and had the typical right skew. They found that mortality increased with increases in scores, and that cut-points of 0, 1-2 and >2 were associated with different outcomes.

Krousel-Wood compared five year survival in prostatectomy patients. The Charlson scores were divided into four groups with distributions of 54.3%, 22.5, 12.3% and 10.9% of patients, for levels
with scores of 0, 1, 2, and >2 respectively. Mortality increased with increasing scores. The area under ROC curve was 0.67. Cox modeling (controlling for age and surgical procedure type) showed that groupings of levels 0 plus 1 vs. 2 plus 3, and 0 plus 1 vs. >3 had statistically significantly different relative risks of death.

Rochon et al.\textsuperscript{33} used the Charlson Index to study outcomes of length of stay and mortality in a series of hospitalized spinal cord injured patients. The scores had a skewed distribution to right, with 71.4\% having a score of zero. Mortality increased with increasing scores, but the length of stay did not. Rochon et al.\textsuperscript{34} used the same study population to examine the risk factors for the development of pressure sores. They used a cut-point of >2 and found an OR of 2.2 using a linear regression model.

Waite et al.\textsuperscript{35} tested the Charlson Index for its ability to predict readmission to hospital. The mean value score was 2.1 across all abstracters. This was a negative study as the Charlson Index (along with the KFC and the ICED) did not predict readmission.

Poses et al.\textsuperscript{53} assessed the use of the Charlson Index in the Intensive Care Unit (ICU) to predict in-hospital mortality. They reviewed the charts of 201 patients (excluding cardiac surgery patients) admitted to an ICU, and used the area under ROC curve to compare the Charlson to the portion of the APACHE II (see glossary) that measured "similar disorders". This study is an example of convergent validity. Mortality increased with increasing Charlson Index scores. The area under the ROC curve, used a measure of accuracy of the index in predicting mortality) was 0.67 for the Charlson and 0.57 for the APACHEII. There was no statistical difference between these 2 scores. The complete APACHE II score had area under ROC curve of 0.87 as might be expected for an index designed specifically for ICU patients.

Lawrence et al.\textsuperscript{54} examined the records of patients who had abdominal surgery to estimate the risk of, and to create a model to predict pulmonary complications of abdominal surgery. Of 2,291 unselected patients, 82 had pulmonary complications. These 82 were compared to 82 controls selected from 412 random chart reviews. The mean Charlson score in those with complications was 2.7 and in those without was 1.6. The Charlson Index had the most significant p value of all potential predictors in a univariate analysis. Using a logistic regression, the model (including abnormal clinical examination, abnormal chest x-ray, a cardiac risk index and the Charlson Index) explained 50\% of the variance. The OR for the Charlson Index alone was 1.6.

Albertsen et al.\textsuperscript{40} evaluated the Charlson Index in a survival analysis of men with early prostate cancer. No raw data were reported. Using cutpoints of 0, 1, 2, and >2, they found the Charlson Index predicted survival and that no patients with scores >2 survived over the follow-up period(mean 15 years). The curves for scores 1 and 2 appear to overlap but no log-ranks tests were reported.

Newschaffer et al.\textsuperscript{41} examined the performance of the Charlson Index on mortality in a cohort of older breast cancer patients. The study was designed primarily to compare performance and reliability of the Charlson Index, the KFC and the Satariano indexes. [The Satariano Index is a modification of the Charlson Index not discussed in this thesis.] Using the Cox proportional hazards regression and the Charlson Index as a continuous variable, each unit increase of the Charlson Index increased the relative
risk of death by 1.48, when the model controlled for age, surgical site and stage.

Polanczyk et al. 55 used the Charlson Index to identify predictors of mortality in patients with congestive heart failure. They examined the discharge records of 2,320 patients, created both test and validation groups using the Deyo version of the Charlson Index for administrative data. This publication is typical of the literature on the Charlson Index as they used administrative data, modified the original scale for their specific patient set, and compared their new index to the original. They found that stepwise increases in scores were associated with increases in mortality. They also found their own customized heart failure scale designed from data on their own patients outperformed the Charlson Index.

Kiefe et al. 56 used the Charlson Index to determine if chronic disease was a barrier to cancer screening in women. In the setting of an academic general internal medicine/family medicine clinic, the charts of 1,764 patients were examined for mammogram (within two years), PAP smear within three years and breast examination within one year. The mean Charlson score was 1.24. For all of the three screening tests, the proportion of women presenting for each test fell within each consecutive increase in Charlson score. Using a logistic regression model, each unit increase of the Charlson Index was associated with decline in the likelihood of tests (17% for mammogram), after controlling for race, insurance and age.

Gabriel et al. 57,58 published to reports on comorbidity in the arthritis population. These two publications are discussed in more detail in Sections 2.2.4 and 2.5. The studies showed that over 50% of patients had at least one of the Charlson Index items other than the disease of interest (rheumatoid arthritis), that the number of patients with scores >0 increased with follow-up time (discriminative validity), and that the original score predicted subsequent increases in Charlson scores. The convergent validity study by Extermann was noted in the previous section on the CIRS. This study calculated and compared scores for the Charlson Index the ECOG-FS, the ADL and IADL using the charts from 203 elderly cancer patients. There was no correlation between the Charlson Index and the other scores (Spearman Rank 0.14 comparing Charlson to ECOG-FS). The Charlson Index does not reflect functional or performance status.

Feasibility:

The classification is simple to understand, perform and calculate. Training of the abstractor is required, but less than some indexes, as the items are explicit. The score is easy to calculate. The time for abstraction according to Waite (5.9 min) was the shortest of all the indexes tested (KFC, ICED and Charlson).

Summary:

The Charlson Index was designed to be an index of comorbidity to predict mortality and has excellent content validity and reliability. It has been used extensively throughout the literature and has been shown to predict survival in many settings, including oncology and specifically in Head and Neck Oncology. One of its deficiencies is the face validity of a very skewed score distribution. It identifies the sickest patients, but has been described as too restrictive because it is not sensitive to lesser degrees of
comorbidity. The Charlson Index is a continuous variable, and therefore is analysed using cut-points for time-to-event analysis, but those divisions and the number of divisions vary from publication to publication. The Charlson Index has been criticized for being out of date and not generalizable, but continues to be the most common index used, probably because it is easy to use and is reliable. It has been reported in four publications using the Head and Neck cancer population, but a study using a series of patients without selection bias would be needed to establish its value in survival studies.

2.2.4 The Index of Coexistent Disease

The MEDLINE search identified only three publications using the Index of Coexistent Disease (ICED), but after reviewing the bibliographies of these and other references, 9 further studies were identified. All 12 publications were reviewed, and excluding the review article by Extermann, 9 further studies were identified. All 12 publications were reviewed, and excluding the review article by Extermann, are found in Table 2.4. Original Description and Purpose

The ICED was developed by Greenfield et al. 58 as an improvement to his previously reported Comorbidity Index 59. The objective was to create a scale that reflected the “increasing risk of mortality or resource utilization”. Coexistent disease is defined as “illness due to diseases other than the primary disease that may affect the outcome of interest over the period of observation”, and severity is defined as a “concept that reflects the increasing probability of experiencing outcomes such as complications or poor functioning that increase the probability of mortality and increased resource allocation”. The ICED is the combination of two scales (the Individual Disease Severity Scale (IDS) and the Functional Severity Scale (FS)). The IDS portion consists of 14 items with a five level severity scale (0-4) consisting of (i) no coexistent disease, (ii) asymptomatic to mild disease, (iii) mild to moderate disease with symptoms and requiring treatment, (iv) moderate to severe disease or an uncontrolled condition on treatment, and finally, (v) severe uncontrolled disease. Within each domain the items are defined explicitly. The Functional Scale consists of 12 items (circulation, respiration, neurological, mental status, urinary, fecal, feeding, ambulation, transfer, vision, hearing, speech) over a severity range (0, 1, 2) reflecting no impairment, mild to moderate impairment and severe/serious impairment. Both the scales use the highest single result over all items. The two scores are then “condensed” on a table to create the final four level ICED score of 1-4 reflecting zero, mild, moderate and severe coexistent illness. The actual ICED scoring system is not published, but a copy was kindly sent to me by Dr Greenfield. A summary of the ICED is found in Appendix 4.

The ICED was designed to stratify risk for outcomes such as complications, length of stay, and costs, as well as mortality. The ICED was first reported in 1991 by Cleary et al. 5 in a study comparing length of stay and complications among six large USA teaching hospitals. The objective was to investigate the hypothesis that differences in length of stay and complications might be explained by differences in case-mix as measured by the ICED. The medical records of 2,484 patients, between 1985 and 1987, admitted with one of acute myocardial infarction, coronary artery bypass surgery, total
Table 2.4
Relevant studies that have used the Index of Co-existent Disease to measure comorbidity

<table>
<thead>
<tr>
<th>author</th>
<th>date</th>
<th>patient population &amp; setting</th>
<th>number of patients</th>
<th>outcome</th>
<th>distribution of scores (1-4)</th>
<th>statistical groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cleary</td>
<td>1991</td>
<td>6 hospitals and 6 conditions</td>
<td>2484</td>
<td>complications and length of stay</td>
<td>23, 46, 125, 61</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>Nicolucci</td>
<td>1992</td>
<td>renal dialysis</td>
<td>255</td>
<td>survival</td>
<td>103, 131, 104, 17</td>
<td>1-4</td>
</tr>
<tr>
<td>Greenfield</td>
<td>1993</td>
<td>total hip replacement</td>
<td>356</td>
<td>complications</td>
<td></td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>Krousel-Wood*</td>
<td>1994</td>
<td>prostatectomy</td>
<td>302</td>
<td>survival</td>
<td>38, 130, 40, 45</td>
<td>1, 2, 3, 4</td>
</tr>
<tr>
<td>Waite*</td>
<td>1994</td>
<td>medical discharges</td>
<td>79</td>
<td>readmission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen</td>
<td>1996</td>
<td>angina</td>
<td>55</td>
<td>patient preference for symptom free state</td>
<td>13, 36, 36, 16</td>
<td>1-4</td>
</tr>
<tr>
<td>Albertsen*</td>
<td>1996</td>
<td>prostate cancer</td>
<td>451</td>
<td>survival</td>
<td></td>
<td>1-4</td>
</tr>
<tr>
<td>Imamura</td>
<td>1997</td>
<td>total hip replacement</td>
<td>49</td>
<td>reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imamura</td>
<td>1998</td>
<td>total hip replacement</td>
<td>863</td>
<td>complications</td>
<td>29%, 37%, 29%, 5%</td>
<td>1-4</td>
</tr>
<tr>
<td>Gabriel* x2</td>
<td>1999</td>
<td>arthritis</td>
<td>891</td>
<td>mortality and increased comorbidity</td>
<td>38% (1&amp;2), 45%, 17%</td>
<td>1, 2, 3, 4</td>
</tr>
</tbody>
</table>

* article comparing the ICED to other indexes of comorbidity
hip replacement, cholecystectomy, or transurethral prostatectomy were reviewed. Follow-up information was obtained through mailed questionnaires. Little detail is given on the scores, except that the mean ICED per disorder ranged from 1.6 to 2.7. The study demonstrated large differences in length of stay among centers within the admission diagnosis groups, but these differences could not be explained by any of the variables including the ICED. It is stated that the ICED score was correlated with the Physical Status Classification of the American Society of Anesthesiologists (ASA), but no data is given. Greenfield in a second publication reported the outcomes from some of the total hip replacement patients. The inpatient medical record of each of 356 patients was reviewed, and one year follow-up was obtained through questionnaires. The number of patients with each ICED score was 103, 131, 104 and 17 for scores 1-4. They found that the ICED score was an independent predictor of major in-hospital non-fatal complications using both univariate analysis and logistic regression models.

Content validity: 

There is no reference to the process of the development of the ICED, the items, or the severity levels, except that it was derived from the previously reported "Comorbidity Index". In the two publications that used the index for outcome analysis "the Comorbidity Index" is described with little detail and no supporting documentation. This index is not available in the literature or from Dr Greenfield. The ICED was created using the "judgmental approach" by an individual. There has been no formal item generation or item reduction. Like the KFC, it covers a wide spectrum of medical illness, but because it explicitly defines each severity level of each item, it risks being incomplete.

Face Validity: 

The ICED does make sense, as it includes a broad range of common illnesses that influence mortality and understandable levels of severity. Furthermore, it has very extensive and multiple sections on heart disease, which is a major component of overall mortality. Its most important feature is that, unlike all other indexes, it includes a functional status component. The two sections of diseases and functional status each have different severity scales and like other indexes (CIRS and KFC) the categories for the severity scales have not been tested. The assumption of equal weights for different items and the assumption of a scale using the highest scores have not been tested. The ICED has explicit definitions for it's categories, so may become out of date.

The ICED has a major flaw in face validity. The final score is derived from a condensation of the scores from the 2 sections but there is no validation of this 'condensation' of scores. An explanation of the determination of a final score ought to be a major component of its development, but it is not reported. The ICED, because of its condensation of scores, is an ordinal scale and assumptions of linearity may not valid when it is used in multiple regression analysis.

Reliability: 

Reliability has been addressed in four of the studies. The details of these four publications are described in more detail in other sections. The ICED manual states that the inter-rater reliability was 90% over 30 patients retested but no information on the methodology of the test is provided. Imamura
et al. trained two orthopedic surgeons and one family physician on the ICED. They reviewed two sets of charts on patients who had total hip replacement. Kappa was only 0.57 and the ICC was 0.75 for the initial study involving two abstractors. In the second part of the study using three abstractors, the ICC was 0.57 and they agreed on only 53% of cases. The author discussed the possible explanations of the poor results, especially noting the poorest agreement between the two orthopedic surgeons, and concluded that the ICED needed to be improved. Waite et al. used four abstractors in their study of hospital readmissions. The overall agreement in the ICED was only 52% and the ICC was 0.79. Krousel-Wood et al. tested inter-rater reliability using 11 charts and found kappa to be 0.80.

**Construct Validity:**

Nine other studies, using a variety of constructs have used the ICED to stratify patients.

Nicolucci et al. tested the ICED in a series of patients with chronic end stage renal disease on dialysis. The medical records of 255 patients were reviewed. The number of patients in each ICED level was 23, 46, 125, and 61. When the two least severe groups were combined, increasing ICED scores were associated with declining median survival, separation of Kaplan-Meier survival curves and increasing relative risk of death using the Cox model.

Krousel-Wood et al. reported a study on mortality after prostatectomy. The objective of the study was to examine the influence of comorbidity on both patient selection for different surgical procedures and on five year survival. Three hundred and two patient records were examined. The numbers in each score level were 38, 130, 40 and 45. They found that mortality increased with increasing ICED score, that the ICED did demonstrate selection bias (patients with higher ICED scores did have less invasive surgery) and that the area under ROC curve for survival was 0.68. Using the Cox model, when the ICED was reduced to 2 groups (0 plus 1 vs. 2 plus 3) and was combined with age to compare the procedures, no difference in the mortality between surgical treatments was found.

A unique use of a comorbidity index and the most interesting paper I reviewed on comorbidity measurements is by Chen et al. They investigated preferences for a symptom free state in patients with angina and used the ICED to control for comorbidity. This study examined individual patient interpretations and reportings of their own health status. Chen et al. postulated that different patients with chronic stable angina would report different levels of ‘anticipated gain’ (comparison between their present health state and imagined health state if their angina was relieved). The number of patients with each ICED score 1-4 was 13, 36, 36 and 16. They found, using both univariate analysis and a linear regression model, that the ‘anticipated gain’ did decline with increasing ICED; ie patients with more medical burden had lower expectations. They also found that comorbidity explained more of the variation than the severity of the angina.

Albertsen et al. tested the ability of the ICED to stratify patients on survival using a series of 451 men with localized prostate cancer. Using univariate analysis, they found that the ICED predicted age-adjusted survival, and using bivariate analysis, that the ICED stratified patients into four levels for all cause death. The distributions of the raw scores and log rank tests were not published.

In the reliability study by Imamura, as noted in the previous section, the charts of a small
series of patients after total hip replacement (THR) were reviewed. The distribution of numbers of patients in the four ICED levels were 17, 4, 18 and 4. In Imamura’s second study, he first described and compared the ICED scores in patients after THR in Japan, the UK and the USA, and then correlated the scores with post-operative complications. The scores from different countries were very different. The patients from Japan had significantly less comorbidity and the patients from the UK had higher scores. The distribution of the scores from the USA patients was 29.0%, 36.9%, 29.3 % and 4.8%. The ICED scores from patients in the UK and the USA predicted serious complications using both univariate analysis and logistic regression.

Gabriel et al. 57, 58 published two papers on comorbidity in patients with arthritis. The first article was a study that followed 891 patients and 891 age/sex matched controls over 10 years to document the development of comorbid illness in the study population. The study also compared comorbid illness in rheumatoid arthritis and osteoarthritis. They found that increasing comorbidity scores at time zero, as measured by the ICED, predicted increasing comorbid illness (a higher ICED score) over time, and that patients with different arthritis types developed different comorbid illnesses. The details of the scores were not included. The second paper, using the same patient data-sets found, using the Cox model and controlling for age, sex and disease, that the ICED (as a continuous variable) was a significant predictor of mortality. Unfortunately, these two articles are flawed as the disease of interest (rheumatoid arthritis) was included in the scoring. Furthermore, they assumed linearity of an ordinal variable for the multivariable analysis.

Dr Greenfield enclosed a copy of a poster from the American Nephrology Society (1997) with the manual for the ICED, but the paper has yet to be published. The poster presented data from the HEMO study (a multicenter hemodialysis trial) on 1000 patients in a randomized control trial assessing variations of dialysis techniques on survival. The ICED had been modified to include 5 new items (nonvascular central nervous system disease, HIV, eye disease, blood diseases and anticoagulant drugs). The distribution of scores of levels 1-4 was 1%, 35%, 31%, 33%. The mean score was 1.99. No outcomes are reported.

Waite’s study 36 of readmissions to hospital as described in sections on the KFC and the Charlson Index found that the ICED could not be used to predict re-admissions.

Feasibility:

The classification itself is simple in theory to understand, but is very time consuming to use because it has 27 items. Training of abstractors is essential due to the complexity of having two different sections with different severity scales. To determine the final score requires the ‘condensation chart’. The time for abstraction, according to Waite 36 is 9.5 minutes, which was longest in their study that compared the ICED, the KFC, and the Charlson Index.

Summary:

The ICED is a complex scale, but is important since it is the only comorbidity scale that incorporates a functional component in addition to the more traditional concept of disease-burden.
comorbidity. Its weaknesses lie with content validity, feasibility and especially with face validity (the 'condensation' of scores problem). It is an ordinal scale and assumes that the single greatest illness carries more impact than any combination of illnesses. Some published studies have shown that it does correlate with mortality. The combinations of levels required to obtain separation of survival curves were not the same in the two studies in prostate disease (Krousel-Wood and Albertsen). Albertsen reported that it was a better scale than the KFC and the Charlson Index for predicting survival, but surprisingly did not report the statistical significance of the differences in the curves. Both of the studies using ICED in total hip replacement showed that the ICED stratified patients for post-operative complications. The Gabriel study of arthritis patients has interesting findings, but has a serious methodological error. Imamura questioned the reliability of the ICED and suggested that it needed to be revised. In the latest publication using the ICED, it has been revised (again). In summary, the ICED is a complex scale with some merit and some success, but has problems of validity.

2.2.5 The Chronic Disease Scale

Original Description and Purpose:

The Chronic Disease Scale (CDS) was published by VonKorff et al. in 1992 and unlike scales that abstract diseases or illnesses is based on pharmacy data. A multidisciplinary consensus group of physicians, pharmacists and health services researchers examined the formulary of the Group Health Cooperative of Puget Sound (GHC), and using predetermined guidelines, identified prescription medications used for treating chronic illnesses. Twenty-six drugs or groups of drugs were selected and then weighted 1-5 to indicate the severity of the chronic illness they were used for. (Appendix 5) The sum of the weighted medication classes is a proxy measure of the severity of chronic illness. A patient's score is the sum of all the weights of all the drugs they have taken in the past year, and can range from 0-36. They excluded anti-inflammatory drugs, analgesics, antidepressants, non-prescription drugs and sedative/hypnotics because they were used to treat symptoms not the disease itself. In the GHC, prescription medications are covered by the insurance and over 90% of prescriptions are filled at GHC pharmacies. The pharmacy records from two regions of the GHC consisting of 122,000 patients, were examined for one year, along with hospitalization data and the Washington State death registry for the next year. They found that the total score, after age/sex adjustment, correlated with hospitalizations and mortality. Comparing patients with CDS = 0, patients with CDS = 2 had an OR of 1.96 for mortality, and patients with CDS = 6 had an OR of 5.62.

There are only four publications on the CDS as found on Table 2.5.

Content Validity:

The item generation, item reduction and weighting of the CDS was accomplished by a multidisciplinary consensus panel.

The CDS was revised by Clark et al. The drug list was expanded from 17 to 28 disease
Table 2.5
Relevant studies using the Chronic Disease Scale

<table>
<thead>
<tr>
<th>author</th>
<th>date</th>
<th>patient population &amp; setting</th>
<th>number of patients</th>
<th>outcome</th>
<th>statistical groups</th>
<th>distribution of scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vankorff</td>
<td>1992</td>
<td>HMO</td>
<td>122,000</td>
<td>1 yr mortality and hospitalization</td>
<td>0,1-3,&gt;3</td>
<td>40%, 31%, 29% for males 65-74</td>
</tr>
<tr>
<td>Clark</td>
<td>1995</td>
<td>HMO</td>
<td>125,000</td>
<td>6 month hospitalization, mortality, costs and office visits</td>
<td>0,1-3,&gt;3</td>
<td>43%, 37%, 19% for males 65-74</td>
</tr>
<tr>
<td>Johnson</td>
<td>1994</td>
<td>HMO</td>
<td>6324</td>
<td>hospitalizations and office visits</td>
<td>0,1-3,&gt;3</td>
<td>45%, 31%, 24%</td>
</tr>
<tr>
<td>Tabin</td>
<td>1995</td>
<td>HMO</td>
<td>1255</td>
<td>cost of cancer treatment</td>
<td>0,1-3,&gt;3</td>
<td>45%, 31%, 24%</td>
</tr>
</tbody>
</table>
groups (included psychiatric illnesses) and the medication list increased from 26 to 81. They tested the new classification using 6 months of outpatient pharmacy data and found the enhanced CDS correlated with mortality.

Face Validity:

The CDS is a novel idea, was well designed and should, given the completeness of the data available, stratify patients. If a complete drug list were available, it should be more objective and accurate than a clinical records based system. The weights make sense, although they were not tested or justified.

One potential problem with the drug list and the weights (see Appendix 5) that is not addressed in the publication involves the drug family of ACE inhibitors that appears in both the Heart Disease and Hypertension items. These drugs are used for treatment of both disorders. If a patient is taking an ACE inhibitor for hypertension their score is 2 and if they are taking the drug for heart disease they have a minimum score on that item of 3. It is possible that a patient is taking the drug for both, and it is not clear how to deal with that now common practice.

The CDS score is a continuous variable, and although the scores are low, cutpoints may be needed for survival analysis. In the original description, they used three cutpoints (0, 1-3, >3).

Reliability:

Reliability tests were not performed as all data were obtained from HMO mainframe databases.

Construct Validity:

A study using data from the Kaiser Permanente Northwest Region of Oregon and the CDS was reported by Johnson et al. The data were obtained from a survey (n=6,324) linked to the pharmacy data from the HMO database. Over 90% of prescriptions are filled at the HMO pharmacy and 80% of members use the drug plan. They divided the continuous scores into 0, 1-3 and >3, and stated that they found similar distributions of scores compared to the original description by Von Korff. Using logistic regression modelling, and adjusting for age and sex, they found that increasing scores of CDS predicted health care visits and following-year hospitalizations. The OR of hospitalization in the following year for patients with CDS over 6 was 5.95.

Taplin et al. used the CDS to control for comorbidity in a study examining costs of cancer treatment in the same HMO population as the original Von Korff study. Cancer data were identified from the regional tumor registry. Costs of hospitalization, office visits, ambulatory treatments and home care were available in the HMO records. The study enrolled patients with breast, colon and prostate cancer during the calendar year 1991. Age and sex matched controls were used to establish the baseline costs of patients without cancer during that year. They found, in general, that the overall costs of initial treatment and continuing care increased with stage and comorbidity and not with age. The net costs were not influenced by the CDS and in fact a paradoxical result of less net costs with increasing CDS, was found. One explanation offered was treatment selection bias and the differences in treatments within an HMO.
Feasibility:

The classification is simple to understand, abstract and perform. No further training is required if the abstractor has some knowledge of prescription drugs. The score is easy to calculate. Time for abstraction is not available but would be brief, if a drug list was available on each patient and would be negligible if the drug list was in a database.

Summary:

The objective of the CDS was to create a proxy for health status that was objective and measurable based on prescription drug use. The scale is a list of drugs or groups of drugs, and the score - the sum of all items present in any patient over one year - serves as a proxy for severity of chronic illness. The CDS is an index that provides information without chart review or the problems associated with chart review (time, costs, completeness, subjective judgements of abstractors) if a complete drug list is available and therefore has the potential to provide an accurate estimate of comorbidity. It is untested in the clinical research.

2.3 Comparisons of the Indexes in the Literature

There are 13 papers in the literature that compare results across indexes. As shown in Figure 2.2 and Table 2.6, all these studies compare one or more indexes to the Charlson Index. In this section, the articles that compare the Charlson and one other index (the KFC, the ICED or the CIRS) will be reviewed first, followed by the comparisons among three indexes. There are no comparisons to the CDS. Following the comparisons is a section on published correlations of the scores. The setting and methodologies of all 13 publications have been reviewed in the previous sections on each individual index. This section highlights the comparison portions of each publication. The article by Extermann is a review of the methodology and validity of select indexes and is not included in either Table 2.6 or this section, but is included in the summary Section 2.8.

Comparisons Between the Charlson Index and the Kaplan-Feinstein Classification:

In the original publication by Charlson et al., the validation study compared the Charlson to the KFC. This study examined the survival of 685 women with breast cancer with 10 year follow-up. The indexes were compared using tables, Kaplan-Meier survival curves, Cox regression and variance explained. Both indexes were successful in stratifying for comorbidity using all four statistical methods. There was little difference in the results, especially at the extremes of comorbidity scores. In the Kaplan-Meier analysis, the KFC Grades 2 and 3 curves overlapped, and only scores 0 - 3 are reported for Charlson Index. Charlson concluded that the clinical judgement methodology of the KFC had similar results to the "empiric" methods of Charlson Index.

Singh et al. compared the KFC and the Charlson Index using a series of patients under 45 years of age with Head and Neck cancer. They used Kaplan-Meier and Cox models to analyse
Figure 2.2
Studies that compare another comorbidity index with the Charlson Index

KFC
Newschaffer
Charlson
Singh

ICED
Albertsen
Waite
Krousel-Wood

CIRS
Gabriel

Rochon
Extermann
<table>
<thead>
<tr>
<th>author</th>
<th>date</th>
<th>comparison index(es)</th>
<th>patient population &amp; setting</th>
<th>number of patient</th>
<th>outcome</th>
<th>tests used for comparisons **</th>
<th>conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlson</td>
<td>1987</td>
<td>KFC</td>
<td>breast cancer</td>
<td>685</td>
<td>survival</td>
<td>univariate, KM and CPHM</td>
<td>no difference</td>
</tr>
<tr>
<td>Singh(2)</td>
<td>1998 &amp; 1999</td>
<td>KFC</td>
<td>young H&amp;N cancer</td>
<td>70</td>
<td>survival</td>
<td>KM and CPHM</td>
<td>Charlson more feasible, no difference in results</td>
</tr>
<tr>
<td>Newschaffer</td>
<td>1997</td>
<td>KFC</td>
<td>breast cancer</td>
<td>404</td>
<td>survival</td>
<td>CPHM, reliability</td>
<td>Charlson better</td>
</tr>
<tr>
<td>Gabriel (2)</td>
<td>1999</td>
<td>ICED</td>
<td>arthritis</td>
<td>891</td>
<td>survival</td>
<td>CPHM</td>
<td>Charlson slightly better</td>
</tr>
<tr>
<td>Rochon (2)</td>
<td>1994 &amp; 1996</td>
<td>CIRS</td>
<td>institutionalized spinal cord injured</td>
<td>362</td>
<td>survival, length of stay</td>
<td>linear regression</td>
<td>CIRS better for length of stay and Charlson better for survival</td>
</tr>
<tr>
<td>Extermann</td>
<td></td>
<td>CIRS</td>
<td>elderly cancer</td>
<td>203</td>
<td>descriptive</td>
<td>univariate distributions</td>
<td>Charlson less sensitive (restrictive)</td>
</tr>
<tr>
<td>Waite</td>
<td>1994</td>
<td>KFC &amp; ICED</td>
<td>medical discharges</td>
<td>79</td>
<td>readmissions</td>
<td>univariate, reliability</td>
<td>Charlson the most reliable</td>
</tr>
<tr>
<td>Krousall-Wood</td>
<td>1994</td>
<td>KFC &amp; ICED</td>
<td>prostatectomy</td>
<td>302</td>
<td>survival</td>
<td>univariate, CPHM, ROC</td>
<td>KFC best</td>
</tr>
<tr>
<td>Albertson</td>
<td>1996</td>
<td>KFC &amp; ICED</td>
<td>prostate cancer</td>
<td>451</td>
<td>survival</td>
<td>univariate, KM, CPHM</td>
<td>ICED slightly better</td>
</tr>
</tbody>
</table>

** KM= Kaplan-Meier, CPHM = Cox proportional hazards model,
survival, and frequency tables to examine the rates of complications of treatments. They found no
difference in the ability to predict survival, but as noted in sections 2.2.3 and 2.2.5, there are
methodological problems with these studies. They concluded that the KFC may be slightly better in
predicting major complications of treatment, but noted that it was not as easy to use.

Newschaffer et al. 40 tested the ability of the KFC and the Charlson Index to predict survival in a
series of 404 patients with breast cancer. They reported the results of the Cox models using the
indexes as continuous variables, controlling for age, stage and surgical treatment. The Charlson
outperformed the KFC as the relative risk of death of a one unit increase in score was higher and the
95% confidence limits did not cross zero. The Charlson index also had superior inter-rater reliability.
They concluded that the Charlson was better than the KFC as it was easier to use, more accurate and
more reliable. They thought the problem with the KFC was related to the availability and accuracy of
the increased amount of data required.

Comparisons Between the Charlson Index and the Index of Co-existent Disease:

Gabriel et al.57 58 compared the Charlson and the ICED in two publications using the same series
of patients. Both studies compared 891 patients with either Rheumatoid Arthritis or Osteoarthritis from a
hospital based population database to 890 age/sex matched controls. They used Cox models with the
two indexes as continuous variables. The outcome for the first publication was 10 year mortality. They
reported that both the Charlson Index and the ICED increased the Cox chi-square estimate when added
to models of age plus disease type. The increase was slightly greater for the Charlson, but it was
concluded that both instruments were effective. In the second publication, the outcome was
development of more comorbidity (as measured by the indexes) over 10 years. They found that both
indexes predicted rises in comorbidity. As noted in section 2.2.4, these studies have a methodological
flaw and their conclusions have to be considered with caution.

Comparisons Between the Charlson and the CIRS:

Rochon et al.33 34 studied outcomes for 362 patients with spinal cord injuries in a rehabilitation
hospital to compare the CIRS with the Charlson Index. They examined the impact of comorbidity on
mortality, length of stay and the development of pressure sores. They used linear regression models,
and found that the Charlson predicted survival best. On the other hand, the CIRS with age, best
predicted length of stay. Both indexes (Charlson score >2, CIRS score >8) were predictors of pressure
sores using univariate analysis.

Extermann 70 26 compared the score distributions of the CIRS-G and the Charlson using a series
of 203 elderly cancer patients in a geriatric assessment program. The objective of the study was to
examine correlations among other indexes of comorbidity and functional status (ECOG-PS, ADL and
IADL). No raw data are presented and no correlations were found. The authors concluded that
performance status and comorbidity were independent factors. Extermann identified the polarized
results (the floor effect) of the Charlson Index as its biggest problem and used the term “restrictive scale”. He concluded that, although both were reliable, the Charlson might be better for large scale projects, whereas the CIRS (“a comprehensive scale”) might be better in select populations such as “academic oncology centers” or randomized trials.

Comparisons between the Charlson, the KFC and the ICED:

Waite et al. \(^{30}\) compared the Charlson Index, KFC and ICED for their ability to predict readmission within six months after hospital discharge from a general medical service. They used 39 cases and 40 controls. The comparisons used frequency distributions and reliability tests. Not one of the three indexes was successful in predicting readmission. The study did show that the Charlson outperformed the others on reliability testing.

Krousel-Wood et al. \(^{43}\) compared the KFC, Charlson and ICED using a series of patients who had surgery for benign prostate disease. Their primary interest was to compare survival by procedure selection (TURP vs open prostatectomy) and the outcome was five year survival. The records of 302 patients were examined. They compared the indexes using frequency distributions of scores, ROC curves and Cox models. The Charlson Index scores were divided into levels of 0, 1, 2, and >2. They found that increasing scores in all indexes were associated with increasing mortality. The raw scores for the area under ROC curve for survival were 0.69, 0.68, and 0.75 for the Charlson, ICED and KFC respectively. However, when the 95% limits were included, there was no difference in the predictive values of the indexes using the ROC method. Each index was tested in Cox models, controlling for age and procedure (variables such as race, smoking, payer, and surgeon were not significant). Multiple models were attempted using combinations of index results. The Charlson scores when divided into groups of both 0 and 1 vs. 2 and 3 or 0 and 1 vs. >3 had a statistically significant correlation with mortality. Similar results were found for the ICED and the KFC when grouped into 0 and 1 vs 2 and 3. The KFC had the highest relative risk of death (4.23), compared to Charlson (2.23) and ICED (2.26), when controlling for age and procedure. Treatment selection bias was examined using the distribution of scores and relative risk of death within each surgical procedure group, controlling for age. The ICED best stratified patients by procedure (univariate) and reduced the effect of the treatment selection more (reduced the relative risk by procedure) than the other two indexes.

Albertsen et al. \(^{48}\) also studied the three indexes in a prostate disease cohort. The 451 patients had early stage cancer, had no initial surgical or radiotherapy treatment, and had mean follow-up of 15.5 years. The indexes were compared using the Kaplan-Meier method and Cox models for all-cause death, disease-specific survival and other-cause death. All indexes produced 4 separate survival curves (KFC Grades 0-3, ICED levels 1-4 and Charlson using cut-points of 0, 1, 2, >2). All nine Kaplan-Meier curves are published, but no log rank tests are reported. The ICED appears to have the best separation of curves. The indexes did not stratify patients on disease-specific survival.
Correlation of Index scores

Only three authors published the correlations between the indexes they compared. Rochon found the Pearson correlation between the Charlson Index and the CIRS to be 0.51. Gabriel found a Spearman coefficient of 0.58 between the ICED and the Charlson Index. Extermann found the Spearman coefficient of 0.39 between the Charlson Index and the CIRS-G. These modest correlations are explained by the differences in methodology and scoring of these indexes.

Summary

There is surprisingly little in the literature that compares the performance of the four established indexes. The common statistical methods to compare indexes are univariate distributions of scores, ROC curves, Kaplan-Meier survival curves, and the Cox proportional hazards regression. Throughout all of the 13 studies, little detail is presented to guide the selection of cut-points or groups, and no study used a test/validation group format to justify such decisions. Different studies come to different conclusions about the indexes compared to one another (three claimed the Charlson the best, two claimed the KFC the best, one claimed the ICED the best and three stated no difference between the four indexes), and there is no common theme except that all four indexes do stratify patients when mortality is the outcome. There is a need for a well constructed survival study, without selection bias, that compares all four indexes using the range of statistical methodologies. If this was performed using one cancer site (ie head and neck, breast, lung etc) using data from one center, the methodology would set a standard for other comparisons.

2.4 Measuring Comorbidity in Head and Neck Oncology

Comorbidity measures and indexes have been used in nine outcome studies in Head and Neck Oncology (Table 2.6). Of these, four studies are identical in methodology, and will be combined for this review.

Feinstein, Piccirillo and Pugliano have published a series of four studies using data from three different sites (larynx, oral cavity and oropharynx) on the concept of a "Clinical-Severity Index" in Head and Neck cancer. In each of these studies, the methodology was identical. A cohort of patients from a single institution was identified using a tumor registry, the charts were reviewed, and those charts with complete five year follow-up were selected. The KFC was used. Patients were stratified by the presence or absence of Grade 3 ("prognostic comorbidity"). The Clinical Severity Index was calculated combining data from a list of clinical signs, clinical symptoms, TNM stage, and the KFC score. In each case, a final composite index demonstrated predictive ability on survival. In each study, the proportion of patients with Grade 3 comorbidity was low (larynx=27/193, oral cavity=21/276 and oropharynx=33/281), and in each case mortality was greater in those patients than the those without Grade 3. The four studies are not completely comparable, as the details of the scoring systems are slightly different.
<table>
<thead>
<tr>
<th>author</th>
<th>date</th>
<th>patient population &amp; setting</th>
<th>number of patients</th>
<th>outcome</th>
<th>index used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feinstein</td>
<td>1977</td>
<td>larynx</td>
<td>192</td>
<td>survival</td>
<td>KFC</td>
</tr>
<tr>
<td>Piccirillo</td>
<td>1994</td>
<td>larynx</td>
<td>193</td>
<td>survival</td>
<td>KFC</td>
</tr>
<tr>
<td>Pugliano</td>
<td>1997</td>
<td>oropharynx</td>
<td>281</td>
<td>survival</td>
<td>KFC</td>
</tr>
<tr>
<td>Singh</td>
<td>1997 &amp; 1998</td>
<td>young patients, all sites</td>
<td>70</td>
<td>survival</td>
<td>KFC and Charlson</td>
</tr>
<tr>
<td>Pugliano</td>
<td>1999</td>
<td>oral cavity</td>
<td>276</td>
<td>survival</td>
<td>KFC</td>
</tr>
<tr>
<td>Singh</td>
<td>1999</td>
<td>major surgery, all sites</td>
<td>200</td>
<td>post-operative complications</td>
<td>Charlson</td>
</tr>
<tr>
<td>Sabin</td>
<td>1999</td>
<td>larynx</td>
<td>152</td>
<td>survival</td>
<td>Charlson</td>
</tr>
<tr>
<td>Piccirillo</td>
<td>2000</td>
<td>all sites</td>
<td>203</td>
<td>survival</td>
<td>modified KFC</td>
</tr>
</tbody>
</table>
For example, the two larynx studies do not combine two Grade 2s into a Grade 3, and both of the papers by Pugliano state that patients too “ill” for complete treatment were assigned Grade 3. Furthermore, patients not having curative treatment do not appear beyond the original study sample in some of the articles and are not mentioned at all in the others. These 4 studies are biased in their patient selection, but they reported their incidence of Grade 3 comorbidity and they ‘set the stage’ for future investigations into the effect of comorbidity in the Head and Neck Oncology.

Singh et al. ⁴⁹-⁵⁰ published two studies on the impact of comorbidity on the survival, and compared the usefulness of the KFC to the Charlson Index. These studies have been reviewed in previous sections. The study population was a series of patients with squamous cell carcinoma under 45 years of age and the outcomes were survival and complications of treatments. They found that the KFC, when grouped into levels of 0 and 1 vs. 2 and 3, stratified patients on survival on univariate, Kaplan-Meier and Cox regression analysis. The relative risk of death was 2.29, controlling for stage, for their more severe group. They redefined the Charlson Index into four Grades using cutoffs of 0, 1-2, 3-4 and >4, and then combined the Grades into two groups of Grades 0-2 and >2. For these 2 groups of Grades, the Kaplan-Meier curves were statistically distinct (log rank, p=.08) and the relative risk of death was 2.35 for the more severe group. They concluded that there was little difference in the survival information by the KFC vs the Charlson for the increased work involved in the KFC, but did find that the KFC was a better predictor of complications. These studies, as noted previously, have significant selection bias.

Singh et al. ⁷¹ in a study of 200 consecutive patients with Head and Neck cancer, who had major surgery with free tissue transfer reconstruction, investigated the use of the Charlson Index to predict major post-operative complications. They grouped the scores into three levels (0, 1-2 and >2). They found on univariate analysis, that the only predictors of complications were pre-operative radiotherapy, surgical time >10 hours, history of smoking, and Charlson score >2. Using logistic regression models, the only variable that was significant was the Charlson Index. The occurrence of major complications was also associated with an increased average length of stay from 15 to 27 days.

Sabin et al. ⁷² validated the Charlson Index using a series of 152 patients with squamous cell cancer of the larynx treated at a single institution from 1984 to 1994 and compared it to a combined age plus Charlson score (CCI). The Charlson scores were divided into four Grades (0, 1-2, 3-4 and >4) similar to Singh et al. One point was then added per patient per decade of life over the mean of the population, although it is not clear if the one point was added to the original Charlson score or to the assigned Grade. The final scores for regular Charlson and the CCI were then divided into low and high risk (Grades 1-2 vs. Grades 3-4). They found using these two groups that both the original Charlson and the CCI predicted survival by univariate, Kaplan-Meier and Cox regression analysis. The age adjusted index did not outperform the standard index. The limitations of this study are the method of age adjustment (Charlson used one point for every 10 years over 40), the short follow-up time (25% of patients were followed for < 8 months) and the failure to define how they calculated the second index.

Piccirillo’s article entitled, “Importance of Comorbidity in Head and Neck Cancer” has been
noted above in Section 2.2.5. This study analysed the impact of the KFC on survival across many cancer sites, including Head and Neck. He used an updated version of the KFC (mKFC) and renamed the KFC Grades 0-3 using none, mild, moderate and severe comorbidity. This was a single academic institution study and used prospective data. For the 341 patients with Head and Neck cancer, the score distribution was 55%, 24%, 16%, and 5% for the four levels. Two year survival analysis is reported for 203 of the patients. Mortality increased with increasing scores, and the Kaplan-Meier curves for none plus mild, moderate and severe split into separate curves. Using a logistic regression model, controlling for age, sex, race, site and stage, the OR of death increased with increasing mKFC scores except for the most severe level that had only 11 patients. The 95% confidence limits overlapped. The problems with this study, as acknowledged by the author, are the short follow-up time and the incomplete follow-up. The inclusion and exclusion criteria are not stated. The results however are clear - comorbidity is a predictive factor in survival.

In summary, the nine studies noted above set the stage for further research into the application of comorbidity indexes in Head and Neck cancer; all authors conclude with a similar statement. These studies suggest that both the Charlson and the KFC can be used to stratify patients on survival, as well as other outcomes such as complications and length of stay. These articles provide some distributions for comparisons, examples of different groupings within each index, and the Sabin study also tested the combination of age with the traditional Charlson Index. All of the publications have limitations. The ideal study would determine which index works the best for the patient population, would have sufficient sample size with complete follow-up to capture enough events to be statistically relevant, would be free of selection bias, and would report the results in a format similar to the original publications instead of attempting to redefine grades and levels.

2.5 Summary of the Five Indexes

In the preceding four sections, each index was presented, reviewed and appraised. The studies that compared indexes were reviewed and the present knowledge of comorbidity within the setting of Head and Neck Oncology was outlined. In this section, the findings of all the preceding sections are combined across the indexes to create a unique ranking system for the setting of Head and Neck Cancer. A scale representing the overall appraisal of an index in each of the five categories (content validity, face validity, reliability, feasibility and construct validity) and comparing each index to the others in those categories has been created and is presented on Table 2.8. For this scale, excellent is represented by ++++, good is ++ and fair is +. I made these judgements based on the 40 publications in the literature review. For reliability, the results of the reported statistical tests are found on a separate Table 2.9.

Content Validity: Content validity refers to the completeness and accuracy of an index to describe what it is intending to measure. The Charlson Index ranks ahead of all of the others, as it was devised
using statistical methodology and has had numerous validations. The CDS is also excellent, as the judgmental approach was done using a conducted consensus panel. The Miller version or the CIRS-G used a consensus process and is still robust. The ICED and the KFC are both weaker in content validity as they were made up by an individual without item generation or reduction. All the indexes seem to have used all the possible diseases and conditions to create the scales.

**Face Validity:** The CIRS and the CDS have the best face validity as enlightened common sense suggests they are the good scales. The CIRS takes the sum of all possible comorbidities and the CDS is based on a prescription drugs that should reflects chronic illness. The Charlson Index has problems in face validity and has been described by many authors as too narrow or restrictive. Common sense suggests that diseases are missing (ie hypertension), that categories may be too broad to be valid in every case (ie solid tumors) and that the skewed data with a very small number of patients in the severe group makes analysis difficult. The ICED is weak in face validity due to the unjustified combination of scores and the use of 2 different severity scales. The indexes such as the Charlson, the KFC and the ICED that are explicit for diseases can become out of date.

**Reliability:** The reliability of the these indexes has not been extensively investigated or reported, especially in oncology studies. In Waite's study of readmissions, the Charlson, the KFC and the ICED were compared. The Charlson was the most reliable and the ICED was the least reliable. The Imamura study that reported poor reliability of the ICED cannot be ignored. Newschaffer found the Charlson to have better reliability than the KFC. These results are summarized on Table 2.9. The Charlson has the best reliability, probably because it is explicit and has the fewest number of items. There are no studies on the reliability of the CDS as it has no been used in a clinical records based study.

**Construct Validity:** Construct validity has been defined as “how well the index does its job in describing its construct” 25. The Charlson Index, the KFC, the CIRS and the ICED have been reported in studies that stratify patients on survival and only the CIRS has not been reported with outcome data in an oncology study. The literature that compares performance does suggest that the Charlson may be better. The Charlson is a continuous scale that is skewed far right. Increasing Charlson scores correlate with survival and the score is usually divided into cut-points, although the actual cutpoints vary from paper to paper. The KFC describes comorbidity using a Grade 0-3 scale and the distributions, when reported, suggest the fewest patients in the highest group. In univariate studies, increasing KFC is associated with increasing mortality. Using bivariate and multiple variable analysis on survival, at least two of the Grades have to be combined to have statistically significantly different strata. The CIRS describes comorbidity as a continuous variable with a normal distribution, but has not been used extensively in oncology and has not been as robust as the Charlson and KFC. Increasing scores are associated with increasing mortality when used as either a continuous variable or when using cut-points. The ICED stratifies comorbidity on a 1-4 ordinal scale. The distribution of scores tends to a normal pattern, but the distribution varies with the study population. In survival studies using Kaplan-Meier or multiple variable regression, most studies found that the strata overlapped, except the Albertson study
<table>
<thead>
<tr>
<th></th>
<th>content validity</th>
<th>face validity</th>
<th>reliability (see Table 2.9)</th>
<th>construct validity</th>
<th>feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS</td>
<td>++</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>KFC</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Charlson Index</td>
<td>+++</td>
<td>+</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>ICED</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>CDS</td>
<td>+++</td>
<td>+++</td>
<td>n/a</td>
<td>n/a</td>
<td>+++</td>
</tr>
</tbody>
</table>
### Table 2.9
Summary of Reliability studies

<table>
<thead>
<tr>
<th>index</th>
<th>author (reference)</th>
<th>number of abstractors</th>
<th>kappa</th>
<th>% agreement</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS</td>
<td>Extermann(9) Miller(12)</td>
<td>2 3</td>
<td></td>
<td></td>
<td>.76 .88</td>
</tr>
<tr>
<td>KFC</td>
<td>Waite(10) Newschaffer(14)</td>
<td>5 2</td>
<td>.82</td>
<td>50.6</td>
<td>.82</td>
</tr>
<tr>
<td>CCI</td>
<td>Waite(10) Extermann(9) Newschaffer(14)</td>
<td>5 2 2</td>
<td>.94</td>
<td>58.0</td>
<td>.93 .74</td>
</tr>
<tr>
<td>ICED</td>
<td>Waite (10) Imamura(11) Krousel-Wood(13)</td>
<td>5 3 (2) 2</td>
<td>(.90)</td>
<td>52.0 53.0</td>
<td>.79 .57</td>
</tr>
</tbody>
</table>
which found that the ICED separated patients into four subjectively distinct curves. For the CDS, although no complete data is available on the distribution of scores, increasing mortality correlated with the increasing scores based on administrative data. All the indexes have been used in other settings outside oncology survival analysis and the Charlson has had the widest application. The CIRS has tended to be used in the geriatric population, the KFC in Oncology and the ICED in health policy studies. Examples of classic construct validity have been published with the CIRS.

**Feasibility:** The Charlson is the acknowledged to be the easiest to use and the CDS would be the easiest to use if a drug list was available. The ICED is the most complex, as reflected in the abstraction time data by Waite et al.

Extermann reviewed four of the indexes to assess their use in outcome studies of elderly cancer patients. In his review, each index was evaluated under the headings of "description, clinical experience, performance, ease of use, cross-compatibility and data preservation". Some of his relevant conclusions were:

1. "The Charlson Index is simple to use and highly suitable for vast cohort studies but may under-detect significant problems..."

2. "The CIRS gives a very accurate profile of comorbidity prevalence but may over-detect minor problems..."

3. "The KFC has shown reproducible results ... whilst remaining simple to use."

4. "The ICED can be very interesting for studies ... where a global 'host cofactor' measurement is sought"

5. "None of the indexes represents the ultimate and final word in comorbidity measurements".

6. "A good measurement and understanding of comorbidity ... will be the key to future progress in geriatric oncology".

In summary, each of the indexes is very different in methodology and each has a list of both strengths and weaknesses that have to be judged in the context of the setting for which it is to be used. The CIRS with the CDS has the best face validity, but the Charlson Index, with the CDS, has the best content validity. The Charlson Index has had the widest exposure and, if that is the criteria, might be considered the gold standard by some investigators. The Charlson Index has been reported in Head and Neck cancer setting. The KFC has the most exposure in Oncology and specifically in Head and Neck Oncology, but is not as robust as others in face validity or content validity. Formal reliability testing has been reported rarely except for studies on the ICED which found it lacking. The ICED and the Charlson Index have problems with face validity. From this, and as summarized on Table 2.8, it is concluded in comparison with the other indexes, that the ICED is not as valid, that the CDS might be if it had any applications, and that the Charlson has better overall validity than the KFC and the CIRS.
Chapter 3  Methods

The objectives of this chapter are:
i. to describe the patient population, the data sources, the study population, and the variables;
ii. to describe the development of a PC-based data abstraction/entry module for the five indexes;
iii. to provide an example of data abstraction using the five indexes;
iv. to outline the steps taken for the comorbidity data abstraction;
v. to outline the plan for data analysis;
vi. to acknowledge the ethics approval

The design of this project is described in detail in the sections below and is summarized in Figure 3-1. Four hundred consecutive outpatient charts were selected from a prospective database of over 1000 consecutive patients from the Head and Neck Oncology Clinic at the Kingston Regional Cancer Center. Exclusion criteria were used to reduce the study population to 379. An electronic data abstraction and entry module for the five indexes was developed for the project. Comorbidity data was extracted from the charts of the selected patients and linked to a prospective clinical database. The best index(es) would stratify patients into three or more statistically distinct comorbidity strata (low, medium and high) on survival (all cause death). This will be determined using bivariate (Kaplan-Meier) analysis and multivariable regression (Cox Proportional Hazards Model). In the bivariate section, adjacent Kaplan Meier curves will be combined or subtracted until the 3 most separate curves are obtained. These grouping will be tested using the log-rank test. The Cox Regression will use the same 3 comorbidity strata for each index in models using indicator variables to determine the independent effect of each strata. The best index will be the one that one that performs the best in both bivariate and multivariable analysis.

3.1  Data Collection

This section will describe the patient population, the reference database used to obtain relevant clinical data, the variables extracted, and the processes used in both the data collection and the data cleansing.

3.1.1. Patient Population:

The Kingston Regional Cancer Center (KRCC) is the regional referral center for the assessment and treatment of patients with Head and Neck cancer for South Eastern Ontario, Canada. The multidisciplinary Head and Neck clinic evaluates approximately 70 new patients per year from an estimated population of 400,000; squamous cell carcinoma represents approximately 80% of cases. Due to the complexity of cancer treatments in Head and Neck cancer, essentially all patients from the region are assessed at the KRCC.
Figure 3-1  The Study Design

400 patients with newly diagnosed squamous cell ca of the Head and Neck

379 patients after exclusion criteria

'OTL' file in Medlog clinical database

Comorbidity Indexes:
- Cumulative Illness rating Scale
- Kaplan Feinstein Classification
- Charlson Index
- Index of Coexistent Disease
- Chronic Disease Scale

KRCC outpatient medical record

data abstraction

PC data abstraction & entry module

patient ID number

analysis
3.1.2. Reference Database:

Since 1985, all new patients with Head and Neck cancer registered at the KRCC for assessment and or treatment have been entered prospectively into the MEDLOG Clinical Data Management System\(^7\). This clinical data management system was designed specifically for the entry, storage, retrieval, manipulation, export and analysis of patient information by medical personnel. In 1985, with Dr David Ginsburg, (CEO of the KRCC, 1985), I designed the data collection system for the Head and Neck Clinic and have been responsible for it since then. Each patient has a unique patient identifier in the system and the OTL database includes 51 variables across the three fields consisting of presentation, treatments and follow-up(Appendix 6). The initial data include KRCC chart number, date of registration, demographic information (age, sex, city) and cancer information (site of the index cancer, subsite of the index cancer, TNM stage, histology, date of diagnosis). Treatment plans and treatment descriptions of radiotherapy, surgery, and chemotherapy are entered after treatment is completed. At all follow-up visits, the patient status is recorded, as are any subsequent treatments. Mortality data are constantly updated using known follow-up information, and supplemented by information from families, doctors or telephone contacts. Patient information is recorded onto data entry sheets by the surgical or radiation oncologists at the end of each clinic and is subsequently entered into MEDLOG by specifically trained clerical personnel. The OTL database is stored on the KRCC mainframe and the only copy outside of the clinic is on my PC. The MEDLOG system is password protected, as is the KRCC mainframe, to protect patient confidentiality. MEDLOG can be programmed for rapid scanning of longitudinal information to identify errors and omissions, and the data-set has been checked at frequent intervals, often in preparation for data analysis. Missing information, such as mortality, has been updated yearly by contacting family doctors. The analysed data in the OTL database has been used in seven publications\(^ {13,16,74-76} \)

3.1.3 The Study Population:

Four hundred (400) consecutive charts of patients with newly diagnosed invasive squamous cell carcinoma of the Head and Neck were identified between November 1990 and September 1996. This cohort was selected to assure minimum three year follow-up. All patients were included in the inception cohort regardless of treatment because treatment is determined and biased by factors such as stage, site and comorbidity.

The exclusion criteria were:

1. Patients with two or more synchronous primary cancers of the Head and Neck at presentation were excluded because it would be impossible to (i) designate the true primary index tumor, (ii) record the stage for analysis, and (iii) know the origin of a relapse. These patients are typically excluded in Head and Neck Oncology research
2. Patients presenting with a second primary Head and Neck cancer were excluded because (i) the treatment as well as prognosis of the new cancer could be compromised by the treatment of the previous cancer and (ii) the site of origin of a failure would not be known. These patients are typically excluded in Head and Neck Oncology research (n=9).

3. Insufficient information in the outpatient chart (n=3)

4. Lost to follow-up within 3 years (3 months, 7 months, 16 months, 34 months) (n=4)

3.1.4 Variables

The variables extracted from the OTL database were:

1. age
2. sex
3. TNM T category (T1, T2, T3, T4 and T? [unknown primary site])
4. TNM N category (N0, N1, N2a, N2b, N2c, N3)
5. index cancer site (oral cavity, oropharynx, larynx, hypopharynx, nose [nasopharynx and paranasal sinuses], unknown primary site)
6. intent of treatment (cure, palliation)
7. initial treatment (radiation, surgery, both, non-curative treatment)
8. date of diagnosis
9. date of last entry
10. status at last entry : <alive and well>, <alive with disease>, <dead with/of cancer>, <dead with cancer, but of an unrelated cause>, <dead of another disease>

3.1.5 Data Completeness:

A recent medical school graduate was trained on MEDLOG to examine the data quality on the 379 patients by checking for completeness and comparing it to the outpatient medical record for accuracy. Every chart was examined. If a patient was recorded at last visit as "alive with disease", the status was investigated and corrected by contacting the family doctor (occasionally the family). The status of all patients with status "alive and well" whose last entry was over one year ago was investigated and corrected if further information was obtained. Updated information on patients on two, three, four, and six month follow-up schedules was added after appointments. Patients known to be deteriorating were entered with the latest possible information available until January 1, 2000, the cut-off date for follow-up clinic visit information. I reviewed new or corrected information before entry. With this strategy, follow-up information was as complete and as accurate as possible to Jan 1 2000.
3.2 Development of Abstraction and Entry Module

This section describes how the comorbidity indexes were incorporated into a computer-based data-entry module, how the module was tested, and how the abstraction was performed. A sample fictitious patient is presented in Section 3.3 where score calculations are demonstrated.

The objective of this process was to create a PC-based module for the rapid and accurate entry of extracted comorbidity data. In order to improve efficiency, it was decided that the module should enter comorbidity data by disease system (domain) across the indexes, rather than by each individual index. This would allow more focussed and more rapid input and would allow analysis by domains if needed. The process consisted of eight steps:

Step 1. Each of the five indexes divide systems and illnesses into different groups. These groupings were combined and/or reduced into 13 separate domains, similar to the KFC and CIRS groupings for the module. These domains were cardiac, respiratory, hypertension, peripheral vascular disease, diabetes, gastrointestinal (both upper and lower), hepato-biliary, other cancers, neurological/psychiatric, renal, genital-urinary, extended musculo-skeletal, and a miscellaneous group. These 13 domains represent the wide range of medical conditions covered by the five indexes and include the known comorbid conditions related to smoking and alcohol.

Step 2. Each index was reviewed under each of our new domains, and the criterion for each severity level was recorded to create the first version of the combinations.

Step 3. Each of the 13 domains was carefully examined to identify conflicts in the five indexes that did not match our chosen 13 domain structure. For example, the CIRS places cancer within each major system, whereas the other scales have separate domains for cancer. This example of cancer is further complicated, as the CIRS places melanoma in the musculoskeletal system and has a group including thyroid cancer, breast cancer, and leukemia in its miscellaneous group. Another example is hypertension, which is listed in the cardiac domain in ICED and in the vascular domain in CIRS. Diabetes is its own domain in KFC but is part of the miscellaneous group in the CIRS. Finally, a large number of co-existent illnesses which appeared in only one or two scales (collagen vascular disease, deafness, AIDS, alcoholism, obesity, anemia) had to be incorporated into a domain. All of these differences were incorporated by creating clearly marked pathways for the abstractor to follow if our domain did not match the index. For instance, if the chart identified a patient with diabetes, entry was made to the KFC index under Diabetes, but a pathway pointed to Miscellaneous for the CIRS index. For breast cancer, a patient would be entered under our domain of cancer for the Charlson Index, ICED and KFC, but a pathway directed the abstractor to the Miscellaneous domain of the CIRS. In Chapter 2 it is noted that the CDS has a potential problem with the drug family of ACE inhibitors as they are not included in the antihypertensive list but are commonly prescribed for this condition. For this project, it was decided to modify the CDS to include this
but only count ACE inhibitors once. Therefore a patient using those drugs for both heart
disease and hypertension was recorded for the heart disease only. For hypertension a score of
2 was assigned. The same applies to beta-blockers that have different scores for different
systems and can only be counted once.
Step 4. Each index has a series of general guidelines or statements about the severity scale.
These general guidelines for the severity scales for each index were reviewed within each
domain to assure consistency.
Step 5. Two essential definitions were established across all five indexes. First, a comorbidity
had to precede treatment to be included. This follows the concept of time zero as defined by
Feinstein 79, that we set as the date of treatment (or the case of palliation as the date of
treatment decision). The abstractor was instructed not to include any information on new
comorbid illnesses recorded after time zero. Second, a “current illness” had to be active within
12 months of diagnosis to be included and an “active drug” had to be used in the 12 months
before the diagnosis. These definitions were not clear in some of the indexes.
Step 6. A final paper version of the module was printed to be used in the abstraction process
(Appendix 7). This document included the rules of each scale within each domain for the
abstractor.
Step 7. A research assistant was hired to convert the 13 domains into electronic ‘pages’. We
selected MedQuest for the PC module. MedQuest 80 is a clinical data collection design system
created by/for the Health Care Financing Administration, the organization that administers
Medicare, Medicaid and the Children’s Health Insurance Programs across the USA. MedQuest
was designed as a platform to create, manage, store, export and analyse custom-programmed
data collections, and it is available as freeware on the Internet 81.
Step 8. Eleven electronic pages were created from the 13 domains to enter abstracted
comorbidity information by index within each domain (Appendix 8). Also included in the data
entry was the unique MEDLOG patient ID number, the KRCC medical record number, patient
name, date of birth, and a complete drug list. Each index within each domain has a series of
check boxes or number entry spaces. We decided that a blank would be used instead of ‘zero’
or ‘no entry’ to indicate “no comorbid illness” in an index within a domain. This was done to
improve the efficiency of input (only positives are entered) and to reduce the size of the module,
recognizing that true missed field errors could not be identified.

3.3 An Example of Calculation of the Index Scores

A typical patient might present with the history below. It is essential to note that the scoring is
of the comorbid or co-existing disease and does NOT include the disease of interest. The scoring
follows the descriptions found in Chapter 2 and Appendixes 1-5.
A 65 year old male presents with a T3 N0 squamous cell ca of the larynx. He had a documented myocardial infarction (MI) three yrs ago, repair of an Abdominal Aortic Aneurysm (AAA) and is taking a Beta Blocker for hypertension (controlled). He has no angina. He is a life-long smoker, and although he has a daily productive cough, he has no respiratory symptoms. He takes Ranitidine for dyspepsia and a past 'ulcer'.

The CIRS score is 10 (range 0-56) for the MI (CIRS=3), AAA surgery (CIRS=3), smoking (CIRS=2), peptic ulcer (CIRS=2). The hypertensive medication is ignored in the Vascular item as the AAA has a greater score.

The KFC is 1 (range 0-3) for any one of the MI, the AAA, the hypertension or the ulcer

The Charlson Index is 3 (range 0-31) for the sum of the MI, the AAA and the presumed ulcer.

Hypertension is not one of the 16 items in the Charlson Index.

The CED 'Individual Disease Severity' score is 2 for the MI, but, as he has no functional limitations his ICED score is 1 (range 0-3).

The CDS score is 2 (range 0-36): 1 for ranitidine and 1 for beta-blocker for hypertension

In the past, to abstract the information on the example patient, the abstractor would have had to flip through each index, checking for items and for the definitions for each of the five listed comorbid illnesses for this hypothetical patient. Each index score would then be individually calculated and recorded. In the electronic version developed for this project, only five of our 11 pages need to be opened (cardiac, respiratory, hypertension, GI and peripheral vascular disease). The appropriate boxes are checked and the scores are generated (and stored) by the computer.

### 3.4 Abstraction

An experienced medical record technician from the KRCC was hired for the project. This person was specifically chosen, because she was familiar with the outpatient charts, had experience with MEDLOG data entry in the past, and had previously done chart abstractions from the KRCC outpatient charts for other research. The goals of the project, the process to date and the paper version were carefully reviewed with the abstractor in a series of meeting with the module developer and myself. The abstractor was taught how to use MedQuest and the PC module was intensively reviewed. At a meeting of the group consisting of the abstractor, the module developer, and myself, the first ten charts were abstracted and entered. All abstracted information was entered onto the paper version first and then into the PC version for all 379 patients. This double entry was performed to allow for checking of the integrity of the electronic version. After 40 charts and again after 60 charts, the information on the paper version was compared with the PC version. During the abstraction of the remaining 369 charts, I met weekly with the abstractor to assist in problem solving for unclear criteria or difficult matching of patient information to criteria. In unclear situations on the interpretation of chart information, I made the final decisions. Some of the problems/solutions/decisions included:
1. Obesity was occasionally noted in the medical record, but no weight criteria were used in any of the indexes. It was ignored unless the record specifically stated gross obesity (CIRS only);

2. If a patient was taking digitalis but no information was available about the actual type of heart disease, it was assumed he/she either had either congestive heart failure or an arrhythmia;

3. We defined 'heavy' smoking as > 40 pk years;

4. We defined 'heavy' drinking as > 5 drinks per day;

Finally, thirty-nine charts (10% of the 379), randomly selected by SAS using the MEDLOG patient identifier, were re-abstracted for error-rate analysis and reliability testing.

3.5 Plan for Analysis

Stage 1. Frequency distributions and/or tables were created for age, sex, T and N stage category, site, initial treatment, status and the scores of the comorbidity indexes. Age was divided into <55, 56-65, 66-75, and >75. The TNM Stages I-IV groups were not utilized for the analysis, as it has been demonstrated that for the population of Head and Neck cancer patients the Stages are not heterogeneous between, or homogeneous within Stage groupings. Instead, a previously developed clinical staging system of five prognostic levels, using combinations of the T and N categories was used.

Stage 2. Kaplan-Meier survival curves and the log-rank test were used for bivariate survival analysis. Followup began at the date of diagnosis and the outcome of interest was all-cause death. Survival analysis by T category was performed to demonstrate stratification by severity of illness. For the comorbidity indexes, the objective of the Kaplan-Meier analysis was to identify at least three statistically different combinations of scores (KFC, ICED) or cut-points of scores (Charlson, CIRS, CDS) to create three levels or strata within each index. Using the unique patient identifier in MEDLOG, a random 2/3 of patients were selected for a training set with the remaining 121 for a test or validation set. The survival curves of each level of each index were examined and adjacent similar curves were then combined or removed until the best three stable separate curves were obtained for each index. These combinations were then confirmed using the validation set. Finally, the log-rank test for all curves, and between all adjacent curves, was used to determine if these best curves were statistically distinct using all 379 patients. The correlation adjusted Bonferroni correction was used to adjust alpha (.05) downward to correct for chance in multiple tests.

Stage 3. The Cox Proportional Hazards Model was used to investigate the independent effects of the indexes on survival. The objective of the multivariable analysis was to test the best combinations of curves from the Kaplan Meier analysis when controlling for the known risk factors of age, sex, site and stage. The Cox model assumes that the relationship between survival and the values of the covariate
being tested are constant over time. This assumption of proportional hazards is tested by plotting the log-log survival function of the SAS Proc Lifetest\textsuperscript{83}. If the curves do not intersect or are “roughly parallel”, the hazards are proportional. The results of the Cox regressions are reported as estimates of the relative risk of all cause death (with 95\% confidence limits and p values) for each strata of each index. The reference group for the model was male, age <55, T1, N0, oral cavity, lowest level comorbidity. When a patient had a primary cancer site of unknown location, the same information is found in both the site and Tstage variables, and therefore 14 patients were removed for the model (N=363). Treatment was not included in the model, as treatment is determined primarily by T stage and site. The indexes were tested for non-proportionality. The index scores were tested for linearity by plotting the values of both the coefficients and the relative risk estimates from the SAS ProcPhreg output.

**Stage 4.** Receiver Operating Characteristic (ROC) curve analysis was performed to examine the accuracy or predictive ability of each index to predict five year survival. The ROC curve uses the sensitivity and specificity of the test (index) to isolate its overall impact on the outcome and is a useful method of univariate comparison of tests. The area under the ROC curve is equivalent to the probability of a random patient who dies having a higher score than one who survives. The ROC Analyser \textsuperscript{84}. The analysis included ROC curve plots and calculations of area under ROC curves with 95\% confidence intervals

**Stage 5.** The proportion of variance explained (pve) was calculated using the Likelihood Ratios from the Cox model output following the method of Schemper for censored data \textsuperscript{85}. The pve for the Cox model is similar to the R squared of Linear regression models and quantifies the proportion of the variance explained by the significant prognostic factors. It provides a method of ranking factors that appear to be similar in relative risk estimates and has been used in Oncology research publications \textsuperscript{86}. For this analysis the baseline pve was calculated using a model with the variables age, sex, site and stage, with each index separately inserted into the Cox model. The partial pve is the difference between the baseline pve and the model pve with each index included and is a measure of the variance added to the model by the index. The partial pve of other prognostic variables are reported for comparison.

**Stage 6.** The indexes were tested for stability over different follow-up times using both Kaplan-Meier curves and the Cox Proportional Hazards Model. The data-set was analyzed after 24 months follow-up, 36 months follow-up and 60 months follow-up.

**Stage 7.** The influence of the comorbidity on disease-specific survival was tested for each index using Kaplan-Meier method. Disease-specific death was defined as death related to the cancer or its treatment. Non-cancer related deaths were censored.

**Stage 8.** Evaluate the performance of the indexes by treatment was performed to be sure that they were valid in order to subsequently compare treatments. Patients were divided into groups of radiotherapy and surgery (included surgery or surgery combined with radiotherapy). Kaplan Meier
analysis was performed including the log rank tests for all three curves for each index. The Cox model was used, controlling for age, sex, site, and stage.

Stage 9. The performance of the indexes was analysed for the cohort of patients who had curative treatment, to be sure that the results based on all patients were not biased by patients who had non-curative treatment, and to be sure that the findings would apply to studies that only included patients treated for cure. This analysis was done using the Kaplan-Meier method and Cox models controlling for age, sex, site, and stage.

Stage 10. The influence of age on survival and the potential interaction of age with the indexes was examined using a series of Cox models with age, the age categories and an interaction term (age*index with age as a continuous variable). If age and an index were correlated, the p value of the relative risk estimate of the interaction term would be more significant than the p value of the individual terms.

Stage 11. The scores of the best 3 comorbidity levels for each index were correlated using the Pearson correlation coefficient.

Stage 12. To be sure the findings comparing indexes were not biased by the decision to establish 3 levels of comorbidity for each index, the Kaplan Meier analysis and Cox models were rerun by creating the best 2 survival groups using the same method as Stage 2.

Stage 13. The PC module abstraction error rate was estimated by comparing the raw scores of all 84 variables extracted from the re-abstraction set of 39 charts. The abstraction re-abstraction error rate was 1.2%. A total of 38 errors were identified. When these errors were examined, one chart had been extracted as the wrong MEDLOG patient identifier and as it was the wrong patient accounted for 14 of the 38 errors. When that chart error was removed, the abstraction error rate was .08% (99.2% error free abstraction).

Stage 14. Intra-rater reliability of the overall data abstraction process for each index was tested by comparing the index scores from the 39 random re-abstracted charts to the original abstraction of the 39 charts. The Pearson correlation coefficient for CIRS, KFC, Charlson, ICED and CDS was .94, .92, .95, .96 and .99 respectively.

3.6 Ethics Approval

This project was reviewed and approved by the Ethics Review Board of Queen's University (Feb 1, 2000).
Chapter 4 Results

The objective of this chapter is to report the results of the data analysis following the outline in Chapter 3 under the following headings:

i. Demographic data and Clinical Information;
ii. Kaplan-Meier Analysis;
iii. Multivariable Analysis;
iv. Receiver Operating Characteristics;
v. Percent Variance Explained;
vi. Effect of Follow-up Time;
vii. Effect of Disease-specific Survival
viii. Effect of Treatments;
ix. Effect of Intent of Treatment;
x. Effect of Age;
xii. Correlation of Index Scores
xiii. Further analysis
xiv. Comments of the Abstractor.

4.1 Demographic Data and Clinical Information

The distributions of the demographic and clinical variables on 379 patients are presented in Table 4.1. The mean age was 65.03 years (SD 10.5, median 66.01, range 30 - 94). Seventy-one percent of patients were male. Follow-up data was complete for 39 months. The mean follow-up for all patients was 41.5 months (SD 29.4, range 0.1 - 110.1 months).

Figures 4.1, 4.2, and 4.3 present the frequency distributions of the five (5) comorbidity indexes extracted and calculated as described in Chapter 2. The expected skewed right distribution was found for the Charlson Index. The CIRS has a normal distribution with a median of 5. Both the KFC and the ICED have non-normal distributions. The CDS distribution demonstrates that 56% of patients were not taking prescription drugs.

4.2 Kaplan-Meier Analysis

The Kaplan-Meier survival curve for all patients (all cause death) is found on Figure 4.4. The probability of survival for three years was 57% and for five years was 41%. The survival curves by T stage are presented in Figure 4.5, demonstrating that increasing T stage is associated with increasing mortality and that T stage stratifies patients on survival. Similar results were seen for N stage and for the five Prognostic Levels (see method for definition).

Figures 4.6a - 4.6d present the four sets of survival curves for the best combinations of curves for the CIRS, KFC, Charlson, and ICED using the training set of 258 patients. The four sets of curves are placed on the same page to simplify comparison between curves and between data sets. Each of the Charlson, KFC, and CIRS could be combined into three prognostic groups that separated survival into levels of decreasing survival with increasing comorbidity. The original KFC has four Grades but
levels 0 and 1 overlapped, and therefore three groups were created. For the CIRS, the cut-points of 0-2, 3-8, and >8 produced the best separation of three curves. For the Charlson Index, the curves for scores 1 and 2 overlapped, and were combined to produce three levels (0, 1 and 2, >2). For the ICED, the lowest three of the original four index levels could not be separated. In order to create three levels, similar to the other indexes, the curves with scores 1 and 2 were combined as they were by Krousel-Wood. The curves of the CDS produced a non-interpretable result. Patients taking no drugs did worse than patients taking multiple drugs and the curves could not be separated into groups of adjacent similar levels. The CDS curves are not reported.

Figures 4.7a - 4.7d present the survival curves for the four indexes from the validation set of 121 patients. The curves are similar in each case to the curves in Figure 4.6 using the test set.

The Kaplan-Meier curves for the CIRS, KFC, Charlson, and ICED, using the complete dataset of all 379 patients, are presented in Figures 4.8a - 4.8d. The curves for the complete dataset are similar to the training set. By simple observation the Charlson Index appears to create the best 3 curves. However the Charlson Index curves 2 and 3 overlap in the first 10 months as do the curves 2 and 3 of the CIRS. The KFC does not have any overlap and at 3-5 years follow-up has 3 distinct curves.

Table 4.2 presents both the distribution of patients and the univariate survival for each level of each index. This demonstrates the difference in the sensitivity of the levels created by the study method to both the highest and lowest measurable comorbidities, as well as the range of survival to which each level is sensitive. Each of the indexes stratifies patients into groups of different size; the lack of balance between the created levels for all the indexes is evident. Both the CIRS and the ICED have normal distributions and the Charlson and the KFC have skewed distributions. When using the Charlson Index, 75% of patients are in the lowest comorbidity group whereas the CIRS has 17% in the lowest comorbidity group. The opposite is seen with the highest comorbidity groups; 8% of the patients are in the highest Charlson group and 22% are in the highest ICED group. The range of the survival strata also varies, as the Charlson Index has the narrowest range (19.4%) while the ICED has the widest (31.6%). This table confirms that the indexes measure different aspects of comorbidity as the 85 patients with the worst prognosis using the ICED, have much poorer survival than the 77 worst prognosis patients using the CIRS and the 32 worst prognosis patients using the Charlson Index.

Log-rank tests were performed to determine the statistical significance of the separation of adjacent curves using the data-set of all 379 patients. The log rank test p values are found in Table 4.3. The mean correlation of the indexes was 0.49 (Table 4.6 and Section XI) and therefore the correlation adjusted Bonferroni corrected alpha = 0.014. None of the indexes created three statistically different curves of survival. The CIRS did create statistically separate curves 1-2 and 1-3, but not for 2-3. The KFC did not separate 1-2 (p=.09). The Charlson did not separate either 1-2 or 2-3. The ICED, like the KFC did not separate 1-3 although the magnitude of this was greater than the KFC as expected from the figures. Based on these findings and the fact that the KFC was the only index without overlapping curves, the KFC came the closest to creating three separate curves.

To test for non-proportionality in preparation for the Cox regression analysis, log(-log) plots were
generated for T stage, N stage, and the four indexes. The curves in each case were consistent with proportionality of hazards. The log(-log) survival plots for the KFC are presented in Figure 4.9.

4.3 Multiple Variable Regression Analysis

Table 4.4 presents the results of the Cox Proportional Hazards regression model for the variables of age, sex, site, and T and N category, with the four indexes inserted separately into the model. The relative risk of death increases with age, varies by site and increases with T stage. As noted in other publications, there is no significant difference in the relative risk estimate between N0 and N1. When the T and N stages are combined into five prognostic levels, the relative risks are 1.42, 2.38, 5.65, and 21.59, with the lowest prognostic level as the reference group. Patients with cancer of the larynx have a better prognosis than all other sites and cancer of the nose/sinuses/nasopharynx has the worst prognosis. Sex was not significant in the model. All four indexes show progressive increases in relative risk of death with increases in comorbidity levels. The relative risk estimates for the Charlson and the CIRS have statistically significant p values, the KFC has marginal significance between levels 1 and 2 with a p value of 0.055, and the ICED is the only index that does not reach statistical significance for both levels. Comparison of the relative risk point estimates and the 95% confidence limits is best appreciated using a high-low plot (Figure 4.10). For all four indexes, the relative risk of level 2 comorbidity is extremely close to the relative risk of the reference group, and both the KFC and the ICED separate levels 2 and 3. The final results for the ICED and the KFC using the Cox model are similar. The CDS was tested using many combinations of cut-points including those reported in the original study, but no statistically significant strata could be created.

Linearity was tested by plotting the relative risk point estimates and the parameter estimates from the Cox output (Figure 4.11). The Cox models were then rerun using each of the indexes as a continuous variable. The relative risks were 1.36, 1.73, 1.66 and 1.44 for the CIRS, KFC, Charlson, and ICED. All were statistically significant and the 95% confidence limits did not cross 1.0. The KFC had a 73% increase in the relative risk of all cause death per level increase, controlling for age, stage and site.

4.4 Receiver Operating Characteristics

The ROC curves are presented on Figures 4.14 and 4.15. The areas under the ROC curves for the indexes were CIRS (0.57 +/- 0.029), KFC (0.62 +/- 0.028), Charlson (0.58 +/- 0.029), and ICED (0.61 +/- 0.028). The areas under the ROC curves for the KFC and the CIRS were the most different but there was no statistical difference between those areas (p=0.08). The ROC study shows that using this form of univariate analysis, all the indexes predicted survival and were all similar. ROC curve scores in this range suggest that comorbidity contributes a modest but significant portion of the information determining survival.
4.5 Percent Variance Explained

The Percent Variance Explained (pve) was calculated for each index using the output from the Cox model. The results are found in Table 4.5. All indexes increase the pve of the baseline model. The KFC and the ICED produced the greatest increases in explained variance. This increase is similar to the increase seen when N stage is added to the model of age, sex, site and T stage. The partial pve for site was 14.0 and for T stage was 17.0. To Oncologists treating patients, an increase in pve similar to that of N stage represents a very important contribution.

4.6 Effect of Follow-up Time

The influence of follow-up time on the levels of the four comorbidity indexes was examined by analysing the data for 24 months, 36 months and 60 months follow-up. The number of events (all cause death) was 141 at 24 months, 160 at 36 months and 203 at 60 months. The overall number of events was 228. With increasing follow-up time, the relative risk point estimates were more stable for all indexes and the lower 95% confidence limits tended to increase steadily (data not included).

4.7 Effect on Disease-specific Survival

The possible influence of comorbidity on disease-specific survival was investigated using the Kaplan-Meier analysis. Figure 4.12 presents the curves for the Charlson Index. There is overlap initially but eventually the 3 curves separate indicating that comorbidity, as expected, does have some effect on disease-specific survival (the impact of treatment selection bias), but the curves are very different from those on Fig 4.8c, demonstrating the impact of adding non cancer death to the curves of survival. For the ICED, the curves for level 1 and 2 reversed position. The curves for the KFC overlapped for severity 1 and 2. The curves for the CIRS overlapped for levels 2 and 3. None of the indexes performed well in this, as expected because the all cause death reflects the impact of comorbid illness.

4.8 Effect of Treatment

The analysis was rerun using the initial treatment groups of radiation or surgery (as defined in the method section 3.5, Stage 8) to be sure the findings applied across treatments. There were 243 patients treated with radiotherapy. In the Kaplan-Meier analysis, the curves of the radiotherapy patients were similar to the curves using all 379 patients, except that both the CIRS and the ICED had poorer separation. The overall log-rank test for the three CIRS curves were not statistically significant for the radiotherapy patients. Using multivariable analysis, the CIRS had different relative risks compared to
the analysis using all patients and was the only index that did not have one level with a significant p value. The results for the other indexes were comparable to the cohort of all patients. In the surgery group (n=83), the Charlson did not perform well; the relative risks for levels 2 and 3 were similar and were very high (relative risk=3.2), but this is probably a reflection of the small sample size of surgical patients with higher Charlson scores. Both the CIRS and the KFC had one level with a non-significant p value, and the ICED had relative risks of 2.9 and 6.6 for levels 2 and 3 respectively, both of which were significant. The ICED performed best across the treatment types. The results by treatment have to considered with caution due to the small numbers of patients.

4.9 Effect of Intent of Treatment

The data on the 334 patients treated "for cure" was analysed to be sure that the results were stable and not biased by including the sickest patients. Using the Kaplan-Meier method and the log rank tests, the curves were all similar to the curves using all patients and the log rank tests were unchanged for all curves. The results of the Cox model, reported as relative risk of all cause death by index on a high-low graph, are found in Figure 4.13. There is no difference in these results compared to Figure 4.10 which reported all 363 patients.

4.10 Effect of Age

The influence of age on survival was explored using Cox modeling. Using the age categories as indicator variables, age produced a slight progressive increase in relative risk without statistical significance (Table 4.4). When the four categories were entered as a continuous variable, the relative risk was 1.22 (p=.002). When age was entered as a continuous variable, the relative risk was 1.02 (p=.0001). These findings suggest age is a weak, but significant factor in survival. The Cox model was rerun using a series of interaction terms with age and the indexes as continuous variables. The p value of the interaction term was never significant confirming that age and comorbidity are independent variables.

4.11 Correlation of the Index Scores

The results of the Pearson correlation between the 4 indexes using the 3 best curves to create groups is presented on Table 4.6. The CIRS did not correlate with the other indexes. The KFC and the ICED best correlated. These results are not surprising (an reassuring) given the curves the Kaplan-Meier analysis and the distribution of patients by level for the indexes. This is an example of construct validity for the ICED and KFC.
4.12 Further analysis

To be sure that the methodology of selecting the best 3 curves from each index did not bias the results of each index as well as the comparisons of the indexes, the analysis was rerun. First the best 2 curves for each index were created using the same previously described method. This method was chosen to isolate the worst prognosis patients within each index. The distribution of scores per level, the p values of the log rank tests, the relative risk point estimates for all cause death (continuous variables) and the partial pves are found on Table 4.7. All the indexes created a group of 50-100 worst prognosis patients. The log rank tests were all significant. The relative risk estimates were similar to the 3 level risk estimates for the CIRS and the Charlson but more significant, and similar, for the KFC and the ICED. The pve for the model without the indexes was 35%. The partial pve for the indexes were 0, 2.4, 4.1 and 5.3 for the CIRS, the Charlson, the KFC and the ICED respectively. These results are similar to those obtained with the 3 levels of severity and support the finding that the ICED and the KFC out-perform the CIRS and the Charlson in the multivariable analysis.

4.13 Comments of the Abstractor

No formal tests of feasibility were performed in this study, but at the end of the data abstraction, the abstractor was asked for comments about the indexes.

"The Charlson didn’t catch much"
"The KFC was too specific"
"The CIRS was complete with a logical range of illness severity. Having each cancer item within each system made it easier"
"The ICED was too variable across items by severity"
"The CDS has drugs that are of no importance such as drugs for acne, migraine, glaucoma and gout."
### Table 4.1
Clinical information on 379 patients with squamous cell carcinoma of the Head and Neck

<table>
<thead>
<tr>
<th>Location</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>oral cavity</td>
<td>99</td>
<td>26.1</td>
</tr>
<tr>
<td>oropharynx</td>
<td>78</td>
<td>20.6</td>
</tr>
<tr>
<td>larynx</td>
<td>135</td>
<td>35.6</td>
</tr>
<tr>
<td>hypopharynx</td>
<td>33</td>
<td>8.7</td>
</tr>
<tr>
<td>nose/nasopharynx etc</td>
<td>18</td>
<td>4.7</td>
</tr>
<tr>
<td>unknown primary</td>
<td>16</td>
<td>4.2</td>
</tr>
<tr>
<td>T1</td>
<td>112</td>
<td>29.6</td>
</tr>
<tr>
<td>T2</td>
<td>130</td>
<td>34.3</td>
</tr>
<tr>
<td>T3</td>
<td>57</td>
<td>15.0</td>
</tr>
<tr>
<td>T4</td>
<td>64</td>
<td>16.9</td>
</tr>
<tr>
<td>T? (unknown primary)</td>
<td>16</td>
<td>4.2</td>
</tr>
<tr>
<td>N0</td>
<td>269</td>
<td>71.0</td>
</tr>
<tr>
<td>N1</td>
<td>44</td>
<td>11.6</td>
</tr>
<tr>
<td>N2</td>
<td>55</td>
<td>14.4</td>
</tr>
<tr>
<td>N3</td>
<td>11</td>
<td>2.9</td>
</tr>
<tr>
<td>prognostic level 1</td>
<td>98</td>
<td>25.8</td>
</tr>
<tr>
<td>prognostic level 2</td>
<td>106</td>
<td>27.9</td>
</tr>
<tr>
<td>prognostic level 3</td>
<td>64</td>
<td>16.8</td>
</tr>
<tr>
<td>prognostic level 4</td>
<td>82</td>
<td>21.6</td>
</tr>
<tr>
<td>prognostic level 5</td>
<td>13</td>
<td>3.4</td>
</tr>
<tr>
<td>radiotherapy</td>
<td>256</td>
<td>67.5</td>
</tr>
<tr>
<td>surgery</td>
<td>64</td>
<td>16.9</td>
</tr>
<tr>
<td>both</td>
<td>24</td>
<td>6.3</td>
</tr>
<tr>
<td>non curative treatment</td>
<td>35</td>
<td>9.2</td>
</tr>
<tr>
<td>alive and well</td>
<td>141</td>
<td>37.2</td>
</tr>
<tr>
<td>dead of disease</td>
<td>137</td>
<td>36.1</td>
</tr>
<tr>
<td>dead of another cause</td>
<td>90</td>
<td>23.7</td>
</tr>
<tr>
<td>dead with disease of another cause</td>
<td>11</td>
<td>2.9</td>
</tr>
</tbody>
</table>
Table 4.2
The best 3 levels of each index: frequency distributions and survival* (n=379)

<table>
<thead>
<tr>
<th>index</th>
<th>level</th>
<th>scores</th>
<th>n</th>
<th>% survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS</td>
<td>1</td>
<td>0-2</td>
<td>56</td>
<td>57.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3-8</td>
<td>246</td>
<td>35.7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>&gt; 8</td>
<td>77</td>
<td>29.8</td>
</tr>
<tr>
<td>KFC</td>
<td>1</td>
<td>0, 1</td>
<td>204</td>
<td>45.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>102</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>73</td>
<td>19.1</td>
</tr>
<tr>
<td>Charlson</td>
<td>1</td>
<td>0, 1</td>
<td>286</td>
<td>41.2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>61</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>&gt; 2</td>
<td>32</td>
<td>21.8</td>
</tr>
<tr>
<td>ICED</td>
<td>1</td>
<td>1</td>
<td>104</td>
<td>48.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2 &amp; 3</td>
<td>190</td>
<td>40.5</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>85</td>
<td>16.4</td>
</tr>
</tbody>
</table>

* all cause death and overall survival at Jan 2000
Table 4.3

The p values of the log rank tests* for best 3 survival curves for each index (n=379) (see Figures 4.8a-4.8d)

<table>
<thead>
<tr>
<th>index</th>
<th>curve 1-2</th>
<th>curve 2-3</th>
<th>curve 1-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS</td>
<td>.009</td>
<td>.25</td>
<td>.002</td>
</tr>
<tr>
<td>KFC</td>
<td>.099</td>
<td>.001</td>
<td>.0001</td>
</tr>
<tr>
<td>Charlson</td>
<td>.02</td>
<td>.117</td>
<td>.0001</td>
</tr>
<tr>
<td>ICED</td>
<td>.508</td>
<td>.0001</td>
<td>.0002</td>
</tr>
</tbody>
</table>

* all cause death
the results of the cox proportional hazards regression** for 363 patients with Head and Neck Cancer

<table>
<thead>
<tr>
<th>variable</th>
<th>n</th>
<th>relative risk estimate</th>
<th>95% confidence limits</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;56 *</td>
<td>47</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56-65</td>
<td>109</td>
<td>1.22</td>
<td>0.86-1.73</td>
<td>.245</td>
</tr>
<tr>
<td>66-75</td>
<td>135</td>
<td>1.26</td>
<td>0.93-1.70</td>
<td>.128</td>
</tr>
<tr>
<td>&gt;75</td>
<td>72</td>
<td>1.48</td>
<td>1.04-2.11</td>
<td>.028</td>
</tr>
<tr>
<td>sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>male *</td>
<td>258</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>female</td>
<td>115</td>
<td>0.85</td>
<td>0.62-1.15</td>
<td>.308</td>
</tr>
<tr>
<td>site</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oral cavity*</td>
<td>99</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oropharynx</td>
<td>78</td>
<td>0.91</td>
<td>0.62-1.32</td>
<td>.629</td>
</tr>
<tr>
<td>hypopharynx</td>
<td>33</td>
<td>1.12</td>
<td>0.69-1.82</td>
<td>.63</td>
</tr>
<tr>
<td>larynx</td>
<td>135</td>
<td>0.40</td>
<td>0.28-0.58</td>
<td>.0001</td>
</tr>
<tr>
<td>nose, nasopharynx</td>
<td>18</td>
<td>2.07</td>
<td>1.09-3.69</td>
<td>.013</td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 *</td>
<td>112</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>130</td>
<td>1.58</td>
<td>1.09-2.28</td>
<td>.014</td>
</tr>
<tr>
<td>T3</td>
<td>57</td>
<td>3.49</td>
<td>2.29-5.33</td>
<td>.0001</td>
</tr>
<tr>
<td>T4</td>
<td>64</td>
<td>5.39</td>
<td>3.55-8.18</td>
<td>.0001</td>
</tr>
<tr>
<td>N stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0 *</td>
<td>269</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>38</td>
<td>0.83</td>
<td>0.5-1.32</td>
<td>.439</td>
</tr>
<tr>
<td>N2</td>
<td>49</td>
<td>2.30</td>
<td>1.5-3.309</td>
<td>.0001</td>
</tr>
<tr>
<td>N3</td>
<td>7</td>
<td>3.53</td>
<td>1.52-8.18</td>
<td>.003</td>
</tr>
<tr>
<td>CIRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 *(0-2)</td>
<td>55</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (3-8)</td>
<td>238</td>
<td>1.66</td>
<td>1.06-2.56</td>
<td>.024</td>
</tr>
<tr>
<td>3 (&gt;8)</td>
<td>70</td>
<td>2.18</td>
<td>1.31-3.63</td>
<td>.002</td>
</tr>
<tr>
<td>KFC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 *(0, 1)</td>
<td>194</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (2)</td>
<td>97</td>
<td>1.38</td>
<td>0.99-1.93</td>
<td>.055</td>
</tr>
<tr>
<td>3 (3)</td>
<td>72</td>
<td>3.03</td>
<td>2.12-4.33</td>
<td>.0001</td>
</tr>
<tr>
<td>Charlson</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 *(0, 1)</td>
<td>273</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (2)</td>
<td>59</td>
<td>1.70</td>
<td>1.20-2.42</td>
<td>.002</td>
</tr>
<tr>
<td>3 (3)</td>
<td>31</td>
<td>2.82</td>
<td>1.77-4.48</td>
<td>.0001</td>
</tr>
<tr>
<td>ICED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 *(1)</td>
<td>98</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (2 &amp; 3)</td>
<td>184</td>
<td>1.37</td>
<td>0.96-1.95</td>
<td>.080</td>
</tr>
<tr>
<td>3 (4)</td>
<td>81</td>
<td>3.17</td>
<td>2.13-4.72</td>
<td>.0001</td>
</tr>
</tbody>
</table>

* reference group  ** all cause death
<table>
<thead>
<tr>
<th></th>
<th>model without indexes</th>
<th>CIRS</th>
<th>KFC</th>
<th>Charlson</th>
<th>ICED</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent variance explained</td>
<td>35.6</td>
<td>37.3</td>
<td>41.4</td>
<td>39.4</td>
<td>41.6</td>
</tr>
<tr>
<td>partial percent variance explained</td>
<td>1.7</td>
<td>5.8</td>
<td>3.8</td>
<td>6.0</td>
<td></td>
</tr>
</tbody>
</table>

* controlling for age, sex, site, stage
Table 4.6

Pearson correlation of Comorbidity Index Scores

<table>
<thead>
<tr>
<th></th>
<th>CIRS</th>
<th>KFC</th>
<th>Charlson</th>
</tr>
</thead>
<tbody>
<tr>
<td>KFC</td>
<td>0.367</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charlson</td>
<td>0.301</td>
<td>0.552</td>
<td></td>
</tr>
<tr>
<td>ICED</td>
<td>0.436</td>
<td>0.615</td>
<td>0.499</td>
</tr>
</tbody>
</table>
Table 4.7

Summary of analysis using the 2 survival curves.

<table>
<thead>
<tr>
<th>index</th>
<th>raw scores</th>
<th>n</th>
<th>log rank p value</th>
<th>relative risk estimate</th>
<th>partial pve</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS</td>
<td>0,1,2</td>
<td>56</td>
<td>0.004</td>
<td>1.31</td>
<td>0.5%</td>
</tr>
<tr>
<td></td>
<td>&gt;3</td>
<td>323</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KFC</td>
<td>0,1,2</td>
<td>306</td>
<td>0.0001</td>
<td>2.47</td>
<td>4.1%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charlson</td>
<td>0,1</td>
<td>286</td>
<td>0.0003</td>
<td>1.9</td>
<td>2.4%</td>
</tr>
<tr>
<td></td>
<td>&gt;1</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICED</td>
<td>0,1,2</td>
<td>294</td>
<td>0.0001</td>
<td>2.59</td>
<td>5.3%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.1
the distribution of the comorbidity scores for the Charlson Index and the Chronic Disease Scale (n=379)
Figure 4.2
the distribution of the comorbidity scores for the Kaplan-Feinstein Classification and the Index of Coexistent Disease (n=379)
Figure 4.3
the distribution of the comorbidity scores for the Cumulative Illness Rating Scale
(n=379)
Figure 4.4

the probability of survival* for all patients (n=379)

this figure demonstrates 80% 1 year survival, 57% 3 year survival and 41% 5 year survival

* all cause death
Figure 4.5

the probability of survival* by T stage (n=363)**

This figure demonstrates that increasing T stage is associated with increasing mortality.

* all cause death
** excludes unknown primary
Figure 4.6.a

the probability of survival * by the Cumulative Illness Rating Scale for the training set (n=258)

This figure demonstrates the best 3 strata of increasing comorbidity
* all cause death

Figure 4.6.b

the probability of survival * by the Kaplan Feinstein Classification for the training set (n=258)

This figure demonstrates the best 3 strata of increasing comorbidity
* all cause death

Figure 4.6.c

the probability of survival * by the Charlson Index for the training set (n=258)

This figure demonstrates the best 3 strata of increasing comorbidity
* all cause death

Figure 4.6.d

the probability of survival * by the Index of Coexistent Illness for the training set (n=258)

This figure demonstrates the 3 best strata of increasing comorbidity
for the ICED using the training set of 258 patients
* all cause death
The probability of survival * by the Cumulative Illness Rating Scale for the test set (n=121)

This figure confirms the 3 best strata of increasing comorbidity for the CIRS using the test set of 121 patients
* all cause death

The probability of survival * by the Kaplan-Feinsten Classification for the test set (n=121)

This figure confirms the 3 best strata of increasing comorbidity for the KFC using the test set of 121 patients
* all cause death

The probability of survival * by the Charlson Index for the test set (n=121)

This figure confirms the 3 best strata of increasing comorbidity for the Charlson index using the test set of 121 patients
* all cause death

The probability of survival * by the Index of Coexistent Disease for the test set (n=121)

This figure confirms the best 3 strata of increasing comorbidity for the ICED using the test set of 121 patients
* all cause death
**Figure 4.8a**
the probability of survival* by the Cumulative Illness Rating Scale for all patients (n=379)

*This figure demonstrates the best 3 strata of increasing comorbidity for the CIRS when all patients are included and is similar to 4.6a and 4.7a
* all cause death

**Figure 4.8b**
the probability of survival* by the Kaplan Feinstein Classification for all patients (n=379)

*This figure demonstrates the best 3 strata of increasing comorbidity for the KFC when all patients are included and is similar to 4.6b and 4.7b
* all cause death

**Figure 4.8c**
the probability of survival* by the Charlson index for all patients (n=379)

This figure demonstrates the best 3 strata of increasing comorbidity for the Charlson index when all patients are included and is similar to 4.6c and 4.7c
* all cause death

**Figure 4.8d**
the probability of survival* by the Index of Coexistent Disease for all patients (n=379)

This figure demonstrates the best 3 strata of increasing comorbidity for the ICED when all patients are included and is similar to 4.6d and 4.7d
* all cause death
Figure 4.9

log(- log) plot * for the Kaplan Feinstein Classification using all patients (n=379)

the figure confirms the proportionality of survival hazard for the KFC as the curves are "roughly parallel

* all cause death
Figure 4.10
High-low graph of the relative risk point estimates * and 95% confidence limits from the Cox model ** for the 3 levels of each index (n=363)

* controlling for age, sex, site and stage
** all cause death
relative risk point estimates* and parameter estimates from the Cox model** for the 3 levels of each index (n=363)

* all cause death
** controlling for age, sex, site and stage
Figure 4.12

the probability of survival* by the Charlson Index for all patients (n=379)

* disease specific survival
Figure 4.13
High-low graph of relative risk point estimates* and 95% confidence limits from the Cox model ** for the 3 levels of each index (n=334, patients treated for cure)

* all cause death
** controlling for age, sex, site and stage
Figure 4.14
ROC curves for Kaplan- Feinstein Classification and the Cumulative Illness Rating Scale

![ROC curves for Kaplan- Feinstein Classification and the Cumulative Illness Rating Scale](image-url)
Figure 4.15
ROC curves for Index of Co-Existence Disease and the Charlson Index
The objectives of this chapter are:

i. to use the results of the data analysis, in association with the findings in the Literature Review, to determine the best index for retrospective survival studies in Head and Neck oncology;

ii. to use the results of the data analysis, in association with the Literature Review to compare the summative scales to the scales using the most severe illness;

iii. to compare the results of the pharmacy based scale (CDS) to the other scales;

iv. to present some clinical implications of the findings and methodology;

v. to describe the strengths and weaknesses of the study;

vi. to outline some ideas for future research that lead from this thesis;

vii. to present the final conclusions on the study.

5.1 What is the best comorbidity index in this patient population?

"In clinical practice, the prognostic influences of age and co-morbidity are well recognized, and these influences usually receive careful consideration during the diverse decisions of clinical judgement. In statistical studies, however, the effects of co-morbidity are generally ignored." Kaplan & Feinstein

The purpose of this study was to identify the best comorbidity index to be used in future survival studies of Head and Neck cancer. The best index would be valid in the literature and would stratify patients into at least three levels of severity using both Kaplan-Meier and multivariate regression analysis. This study, unlike studies reported in the literature, uses a population of unselected patients with complete follow-up and reports the scales in their original validated formats. Prospective data (including diagnosis, treatment and survival) were linked to a retrospective chart abstraction for five comorbidity indexes. The CIRS, the Charlson, the KFC and the ICED all stratified patients on comorbidity but no index created 3 levels of comorbidity that were statistically significant in both bivariate and multivariable analysis. The Kaplan-Feinstein Classification (KFC) came the closest - it produced three survival curves that approached statistical significance using the log-rank test. These curves also had the second widest overall range. Using the p values of the Cox multivariable regression analysis, and as shown in the high-low graph (Figure 4.10), the three levels of the KFC were the closest to being statistically different based on relative risk estimates and 95% confidence limits. The KFC, with the ICED, had the highest partial percent variance explained, that increased the variance explained of the reference model from 35.6% to 41.4%. The KFC had the highest relative risk estimate of all the indexes when used as a continuous variable and was stable over different follow-up times. The KFC stratified patients in both the surgery and radiotherapy treatment groups. In this study, the original KFC Grades 0 and 1 were combined based on Kaplan-Meier analysis to reduce the number of strata from four to three. In the literature, both Singh and Clemens used this combination of KFC
Grades to stratify patients on survival. In the critical appraisal sections of Chapter 2, the KFC was not as strong as the Charlson Index. The KFC has good feasibility, face validity, content validity and reliability, and has been used in Oncology including Head and Neck Oncology. In summary, the KFC, based on both performance and validity, is the best index currently available for survival studies in Head and Neck Oncology.

The Charlson Index is the most frequently used comorbidity index in the literature, but it did not stratify patients as well as the KFC in this study. The three best survival curves did separate on Kaplan-Meier analysis despite overlap in the first year of follow-up. Using the log-rank test, the highest two comorbidity strata were not statistically different (p=.17) noting that the small numbers of patients would influence this finding. The Charlson Index had the narrowest range of survival curves suggesting that it was not as sensitive to the comorbidities of this patient group. Using the Cox model, the three levels were statistically significantly different from the reference group based on the point estimates of the relative risk of death and p values, but the 95% confidence limits for the upper two levels overlapped considerably. The Charlson Index did not create independent strata. The percent variance explained by the model including the Charlson Index was lower than the model with the KFC, a finding that is similar to the relative risk estimates when the Charlson Index was used as a continuous variable. The Charlson Index did not perform well for surgical patients but that finding may be due to small numbers as well. In this study the cut-points of 0,1 and 2, >2 were found to create the best Kaplan-Meier curves. These cut-points have been reported previously, including the Head and Neck study by Singh. These results have to be balanced with the findings of the critical appraisal such as superior content validity, construct validity and feasibility, and fair face validity, compared to the other indexes. The Charlson Index, despite its validity in the literature, is not recommended for survival studies in Head and Neck Oncology because it did not stratify patients in either the Kaplan-Meier or Cox analysis for this patient population.

The CIRS did not perform well in this patient population. Despite a normal distribution of scores, it did not stratify patients into three statistically distinct survival curves using bivariate analysis. Using the Cox model, controlling for age, sex, site and stage, the 95% confidence limits of the relative risk estimates overlapped. When the CIRS was used as a continuous variable, the relative risk was the lowest of the four indexes, as was the percent variance explained. The three levels of the CIRS were stable over different follow-up times, but it did not perform well in the two treatment groups. Cut-points of 0-2, 3-8, and >8 were found to create the best survival curves. In the literature the two studies that used similar cut-points were Katz et al. (0-2, 3-5, >5) and Rochon ( <8, 8, >8). These findings are disappointing, as the CIRS has superior face validity. It is not recommended for use in this population.

The ICED had a mixed performance compared to the other indexes. Three distinct survival curves could not be created as the levels 0, 1 and 2 of the original 4 levels overlapped in the training set, and 2 of the 3 severity levels created for this project, based on the literature, overlapped when all 379 patients were used. The ICED produced the greatest separation of curves and it identified the largest number of the worst prognosis patients suggesting that it was most sensitive to the spectrum of
comorbid illnesses that influenced survival. The ICED did perform better in multivariable analysis. The 95% confidence limits did not overlap for levels 2 and 3, but, using the p values of the Cox model, the level 2 was not statistically significantly different from the reference group when controlling for other known predictors. When used as a continuous variable, the relative risk point estimate for all cause death for the ICED was lower than the point estimate for the KFC. The partial percent variance explained was high, similar to the KFC, and was slightly higher when only 2 levels of severity were considered. The ICED performed well in the surgical group, unlike some of the other indexes such as the CIRS. Based on the criteria for this study which was that the best index would create 3 strata of comorbidity in both bivariate and multivariable analysis, the ICED cannot be recommended. If a study wanted only 2 prognostic groups or was only going to use multivariable analysis, the ICED could be considered, noting the problems identified in the literature review including poorer content validity, face validity, reliability (inter-rater) and feasibility compared to the other indexes.

The CDS scores showed an unusual distribution of patients. Only 56% of patients were using the prescription drugs on the CDS list and the number of patients with the other CDS scores was almost equal. The Kaplan-Meier survival curves and the Cox modelling did not separate into levels of increasing mortality with increasing CDS scores. The literature and the critical appraisal suggested that the CDS might be useful, but it was not effective in this population. (see section 5.3)

5.2 How does a summative scale perform compared to a scale that uses the single most severe illness?

All five indexes use slightly different information to assess comorbidity within disease systems, but the fundamental difference between the indexes is how the final score is calculated. The CIRS, the Charlson and the CDS are summative indexes of multiple items, unlike the ICED that uses the single highest score from all the items. Common sense suggests that these different methodologies would result in different score distributions. An index using the single greatest score might underestimate the total medical burden and an index that summed all diseases might overestimate by including conditions that have no relative effect on the outcome. This difference in methodologies would most likely affect the medium strata of the scales when reported as low, medium and high severity. As seen in Figures 4.8a-d, in both the KFC and the ICED there was overlapping of the lowest two comorbidity curves. In fact, in both cases, the original index has four levels and in this study the three least severe levels had some overlap. Conversely, when using the CIRS there was overlap of the two more severe levels of comorbidity. (Figure 4.8) The curve for the mild and for the most severe comorbidity groups using the CIRS had better survival than the mild and the most severe strata for the other indexes and the range of survival across the CIRS strata was the smallest. When using multivariate analysis, the two summative indexes had overlap of the 95% confidence intervals for the upper two strata, and the two indexes using single illness had overlap between the lowest two strata (Figure 4.10). Finally, the two summative indexes had the lowest scores for percent variance explained. These results suggest that the
summative method may over-estimate the middle level of comorbidity while under-estimating the upper level, and that the using the greatest illness method may under-estimate the middle level of comorbidity. This hypothesis is supported by the survival curves for the KFC, the index that combines the 2 methods.

5.3 How does a scale based on pharmacy data perform compared to a scale based on the clinical record?

The CDS index did not perform well in this population - the use of prescription drugs weighted by severity of illness did not stratify patients on survival or improve models of survival. Forty-four percent of patients were taking at least one of the prescription drugs which correlated with the 47% of patients with KFC Grade 2 or greater (definition: impaired vital system). One explanation might be that the patterns of comorbid illness for patients with Head and Neck cancer are very different from those of the general population of the State of Washington that the CDS was based on.

5.4 Clinical and Investigative Implications of this study

There are four important implications from this study.

1. The KFC is a valid index of comorbidity for survival studies in Head and Oncology. The Charlson Index although popular, did not perform as well and combined with its problems of face validity, it is not recommended. The ICED performed well in multivariable analysis but as it has questionable validity in the literature compared to the other indexes. It should be used and interpreted with caution.

2. The partial percent variance explained of the KFC in the Cox model for all cause death controlling for sex, age, T stage and site was similar to the variance explained of N Stage. In the clinical setting this is an enormous relative impact that a clinician not familiar with advanced statistics can understand.

3. This project was designed to identify a scale for future research in the clinical setting of Head and Neck Oncology - a setting confounded by treatment selection bias as noted in the Introduction. When the data from three studies reported by Piccirillo 44 and Pugliano 46 47 are combined, 10.8% of 750 total patients had Grade 3 KFC comorbidity. This is compared to the 379 patients from the KRCC in this thesis that had an incidence of 19.3% Grade 3 KFC comorbidity. Piccirillo 12 has recently published comorbidity data and survival data on all patients with Head and Neck cancer from a large midwest US university center for the year 1995. He used a modification of the KFC, as noted above in Chapter 2.2.2. He reported that the distribution of 341 patients by the modified KFC was 55% Grade 0, 24% Grade 1, 18% Grade 2, and 5% Grade 3. The KFC data from this study was 31% Grade 0, 23% Grade 1, 26% Grade 2 and 19% Grade 3. These findings are statistically different (Chi square p=0.001). It is clear that either Piccirillo's modified KFC is very different from the KFC, or the comorbidity case mix of patients undergoing treatments in his center is less severe that the case mix of patients at the Kingston
Regional Cancer Center. The KFC can be used to demonstrate treatment selection bias.

4. I have added the KFC to the Medlog prospective data collection system for the Head and Neck Clinic at the Kingston Regional Cancer Center effective Jan 2001.

5.5 Strengths and Limitations

As stated in Chapter 1, three components were required to assemble these data to answer the questions in this thesis. First, a high quality, accessible, and consistent patient record had to be identified as a data source. Second, a process for rapid and accurate data abstraction and data entry needed to be established. Third, all baseline demographic data, treatment information, and follow-up status had to be available. The outpatient chart of the KRCC is of high and consistent quality, all charts were available, and a single abstractor with experience with those charts was used. A PC-based module was designed to minimize time, reduce cost, and reduce transfer error. A complete prospective dataset from a series of unselected consecutive patients from a regional cancer center was used and follow-up was as complete as possible.

There are, as in all studies, limitations to this study. Four are itemized below:

1. First, these indexes were designed to be used with inpatient charts to allow for incorporation of test information, along with a more complete medical record. It is possible that using the outpatient chart to collect comorbidity data may produce an overall under-estimation of the scores. However, for these patients, the treatment decisions are almost always made at the outpatient clinic and the reasons for any therapeutic decisions that varied from the clinic treatment policies would be recorded on the chart. Furthermore, using the outpatient chart provided a consistent standard of information as over the time period of the study, the oncologists at the KRCC remained constant. Less than 50% of patients in this study were admitted for treatment, and if information from the inpatient charts was used, the data would be inconsistent and biased. The decision to use the outpatient chart was made, because of its quality, accepting the potential for underestimation.

If there was an underestimation of the comorbidity by using the outpatient chart, the scales most likely affected would be the CIRS and the CDS, the two scales that include data on less severe diseases. The KFC, the Charlson and the ICED rely on more severe diseases that would more likely be used in any therapeutic decisions.

2. Second, in the literature there are no criteria for determining the “best” index. It was decided a priori that both bivariate and multivariate analysis would have to stratify patients if the results were to be acceptable to both clinicians and researchers. If the Cox model determined that one index stratified patients into three distinct groups or created either the highest relative risk estimates or the highest partial percent variance explained, but the Kaplan-Meier curves overlapped, the index would not be acceptable (ie the ICED). Furthermore, there would have to be at a minimum 3 levels of severity (low, medium and high) to be of clinical value. In this study, no index met the strict criteria, but the KFC was
the closest (the p value for one of the log-rank tests was 0.55 and the lower 95% confidence limit of the relative risk of death between level 1 and level 2 was 0.99).

3. **Third**, this study is from one treatment center and therefore includes the treatment selection bias of the Oncologists at the KRCC. The involvement of another center was considered but an accessible complete dataset was not available. Incorporation of data from other centers will be part of the subsequent planned projects (Section 5.6).

4. This study could be improved if more than one outcome were used to test the indexes. The possibility of using admission to the hospital during radiotherapy and length of stay for those admissions was explored. There were only 90 events for 265 radiotherapy patients, and therefore statistical significance would be unlikely, especially using multivariable regression. This analysis could be considered for further investigations.

### 5.6 Future Directions

The main objective of this study was to determine the best index to be used to control for comorbidity in another retrospective study. Squamous cell carcinoma of the hypopharynx is the least common of all the major Head and Neck sites (8%) and has the worst prognosis. The treatments for this disease are very difficult for patients and selecting the best treatment for patients is constant source of disagreement between surgical and radiotherapy oncologists\(^\text{18,19}\) (all the studies are flawed by selection bias). Furthermore the treatment of this disease has changed in the past 18 months to include concomitant chemotherapy with radiotherapy. Further analysis of the data from this study examining the incidence and severity of comorbidity by site also showed patients with hypopharyngeal cancer to also have the highest levels of comorbidity. Because cancer of hypopharynx is uncommon and because of the problems of multiple prognostic factors (4 T Stages, 4 N stages, and 3 subsites), it is unlikely that a clinical trial could be performed that would provide evidence for the most effective treatments by stage. An effectiveness study of the treatments of this disease is therefore planned that will not only describe treatments but compare treatment outcomes. This study will utilize the Ontario Radiotherapy Database located at the Radiation Oncology Research Unit at Queen’s University. This database has linked the Ontario Cancer Registry and CIHI for all cancer patients in Ontario since 1982. The planned study will identify all patients from all centers in Ontario from 1990 to 1997, and obtain the treatment and survival information from the linked database. The outpatient charts of a random selection of these patients will be then examined at each of the Cancercare Ontario Centers to abstract the clinical stage and comorbidity information that are missing from the linked data. The analysis will compare the incidence of cancer of the hypopharynx and its treatments by region, and then compare the results of treatment types. This study, in the absence of data from randomized clinical trials, controlling for the known prognostic factors of stage and comorbidity will report the results of treatments by stage for this disease, and establish the background data for the new chemotherapy/radiotherapy protocols.
The second objective of the study was to validate an index to use to further my investigation in institution selection bias and treatment selection biases between centers. Recently we have been involved in a project that compared treatments and outcomes in Kingston and Southeastern Norway\(^6\). In a subsequent project, Dr Piccirillo will be contacted to attempt a similar comparison between Kingston and St Louis. The plan would be to insert his modified version of the KFC into the module used for the data collection in this thesis and compare treatments controlling for comorbidity (as well as Stage) between the centers.

The unexpected finding from this thesis is the need for reliability studies for these indexes and in particular for the KFC. Four new data abstractors will be selected and trained over two days by the original abstractor and myself. Ten random charts will be used for training, and then a further 40 random charts will be abstracted by each technician. Data input will utilize the PC module developed for this thesis. The Intra class Correlation Coefficient will be used to establish inter-rater reliability of each index.

In order to be sure that the best version of each index is being used, further examination of the indexes is required and a subsequent data analysis is planned. Each of the indexes will be modified and retested. For example, the KFC is ordinal because it combines two Grade 2 comorbidities to make a Grade 3, but the KFC can be simply be summed to form a continuous variable. The CIRS has been reported and validated in other forms including using the greatest single comorbidity, the number of items with scores greater than 2 and the mean score of the positive items. The Charlson needs to be tested in the 'age plus' format, as described by Charlson\(^5\) in this population. The ICED has been reported with its Individual Disease Score only and its items could be summed to create a continuous variable. This would avoid the confusion and face validity problems of Greenfield's condensation of scores. Figure 4.11 demonstrates the linearity of the index scores using the relative risk estimates and the coefficients from the Cox output. Some of the indexes in the graph plotting relative risk estimates have a curvilinear or exponential appearance. This could be explored in subsequent data analysis. All of these modifications need to be tested, using the methodology of a training/validation sets, but this was considered to be outside of the scope of this thesis.

The data from this project provides an opportunity to examine closely the distribution of illnesses that are present in our patient population and in particular to identify the spectrum of disease that most influences survival. This project would examine the diseases and severity of the diseases present in the most severe comorbidity groups in each index. An analysis such as this would also provide the opportunity to compare how the indexes perform within each major disease system.

As noted in the limitations section, further validation of the findings using another construct would be useful. The data is available to test the indexes on their ability to predict the need for and length of stay for hospitalization during radiotherapy. Data was obtained by linking the identified patients in MEDLOG to the Cancercare Ontario patient care database (OPIS) using the available field <admission to hospital>. A logistic regression model, controlling for age, sex, site and T stage, is planned.
5.7 Conclusions

Comorbidities are diseases or conditions that co-exist with a disease of interest. The importance of comorbidity in Head and Neck Oncology is that it can create treatment selection bias that combined with institution selection bias, can confound comparisons of the results of treatments. Comorbidity was found to have the same overall influence on survival as N stage in this study. When randomized clinical trials are not feasible, a valid index is required to control for the effect of comorbid illness in multiple variable analysis. The objective of this thesis was to critically appraise and then test the common validated indexes in the literature to determine the most suitable index. Using the criteria of the study, the Kaplan-Feinstein Classification is the best index for survival studies in Head and Neck cancer.
References


55. Polanczyk CA, Rohde LEP, Pholbin EA, Disalvo TG. A new case adjustment index for hospital
mortality among patients with congestive heart failure. Medical Care 1998; 36:1489-
1499.
56. Keife CI, Funkhouser E, Fouad M, May DS. Chronic disease as a barrier to breast and cervical
57. Gabriel SE, Crowson CS, O'Fallon WM. A comparison of two comorbidity instruments in arthritis.
60. Greenfield S, Aronow HU, Elashof RM, Watanabe O. Flaws in Mortality data: the hazards of
61. Greenfield S, Apalone G, McNeil BJ, Cleary PD. The importance of co-existent disease in the
ocurrence of post operative complications and one year recovery in patients undergoing
62. Imamura K, McKinnon M, Middleton R, Black N. Reliability of a comorbidity measure - the Index of
63. Imamura K, Black N. Does comorbidity affect the outcome of surgery? Total hip replacement in
64. Nicolucci A, Cubasso D, Labbrozi D, Mari Eea. Effect of coexistent diseases on survival of
patients undergoing dialysis. ASAI0 Journal 1992; M291-M295
65. Chen AY, Daley J, Thibault GE. Angina patient's rating of current health and health without
cohort: adaption of the total illness burden index to prostate cancer. Urology 1999;
54:424-429.
67. Clark DO, VonKorff M, Saunders K, Baluch WM. A Chronic Disease Score with empirically
derived weights. Medical Care 1995; 33:783-795.
68. Johnson RE, Hombrook MC, Nichols G.A. Replicating the Chronic Disease Score from
69. Tapin SH, Barlow W, Urban N, Mandelson MT. Stage, age, comorbidity and direct costs of colon,
complications in microvascular reconstruction of Head and Neck defects. Plast and
72. Sabin SL, Rosenfeld RM, Sundaram K, Gady H. The impact of comorbidity age age on survival
73. Medlog Clinical Data Management System. Information Analysis Corporation. 2000:


84. ROC Analyzer version 5.0. Centor R. 1988;


86. Krongrad A, Hong L, Shenghan L. Variation in prostate cancer survival explained by significant prognostic factors. J Urol 1997; 57:1490
Appendix 1

The Cumulative Illness Rating Scale (G version)

Objective: to measure total medical burden

<table>
<thead>
<tr>
<th>Items</th>
<th>Severity Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. cardiac system</td>
<td>1= mild = current mild or past significant problem</td>
</tr>
<tr>
<td>2. vascular system</td>
<td>2= mod = requires first line therapy</td>
</tr>
<tr>
<td>3. respiratory system</td>
<td>3= severe = conditions uncontrolled by first line therapy</td>
</tr>
<tr>
<td>4. ear, nose, throat, eyes</td>
<td>4= extreme severe = acute severe or chronic end organ failure with severe impairment</td>
</tr>
<tr>
<td>5. Upper Gastrointestinal</td>
<td></td>
</tr>
<tr>
<td>6. Lower Gastrointestinal</td>
<td></td>
</tr>
<tr>
<td>7. liver</td>
<td></td>
</tr>
<tr>
<td>8. renal</td>
<td></td>
</tr>
<tr>
<td>9. genital urinary systems</td>
<td></td>
</tr>
<tr>
<td>10. musculoskeletal and skin</td>
<td></td>
</tr>
<tr>
<td>11. central nervous system</td>
<td></td>
</tr>
<tr>
<td>12. psychiatric</td>
<td></td>
</tr>
<tr>
<td>13. endocrine system</td>
<td></td>
</tr>
<tr>
<td>14. misc</td>
<td></td>
</tr>
</tbody>
</table>
examples of items in the Cumulative Illness Rating Scale (G)

1. cardiac
   1= past MI, prn GTN
   2= CHF, AF, CAD with meds
   3= MI<5yrs, abn stress test, CABG, angioplasty
   4= unstable angina, intractable CHF

2. CNS
   1= migraine, TIA
   2= severe headaches, past resolved stroke, mild Parkinsons
   3= past stroke with mild deficit, mod Parkinsons, CNS surgery
   4= functional deficit, severe Parkinsons

3. respiratory
   1= recurrent bronchitis, inhalers prn for asthma, smoker <20
   2= abn xray, recurrent pneumonia, daily inhalers, smoker <40
   3= restrictive dyspnea, steriods, smoker>40
   4= oxygen, past ventilation
Appendix 2

The Kaplan - Feinstein Classification

objective: to measure cogent comorbidity that might be expected to impair long term survival

<table>
<thead>
<tr>
<th>Items</th>
<th>Severity Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. hypertension</td>
<td>Grade 1 = slight decompensation of vital system</td>
</tr>
<tr>
<td>2. cardiac system</td>
<td>Grade 2 = impaired vital system</td>
</tr>
<tr>
<td>3. central nervous system</td>
<td>Grade 3 = recent full decompensation</td>
</tr>
<tr>
<td>4. respiratory system</td>
<td></td>
</tr>
<tr>
<td>5. renal</td>
<td></td>
</tr>
<tr>
<td>6. hepatic</td>
<td></td>
</tr>
<tr>
<td>7. Gastrointestinal system</td>
<td></td>
</tr>
<tr>
<td>8. peripheral vascular disease</td>
<td></td>
</tr>
<tr>
<td>9. cancer</td>
<td></td>
</tr>
<tr>
<td>10. locomotor system</td>
<td></td>
</tr>
<tr>
<td>11. alcohol</td>
<td></td>
</tr>
<tr>
<td>12. misc</td>
<td></td>
</tr>
</tbody>
</table>
examples of items in the Kaplan-Feinstein Classification

1. cardiac
   - Grade 1 = Atrial Fib or MI >6mos
   - Grade 2 = CHF or angina
   - Grade 3 = < 6 mos of CHF, MI, arrhythmia, hospitalization for angina

2. alcohol
   - Grade 1 = drinking problem
   - Grade 2 = complications related to alcohol
   - Grade 3 = past DTs or rumfits

3. respiratory
   - Grade 1 = mild dyspnea, TB, asthma, mild COPD with abn xray or PFTs
   - Grade 2 = moderate dyspnea, recurrent infections, recurrent asthma
   - Grade 3 = pulmonary insufficiency, severe asthma
The Charlson Index

objective: a prognostic taxonomy for conditions which might alter the risk of short term mortality

weighted items

1 = myocardial infarction
   congestive failure
   peripheral vascular disease
   cerebral vascular disease
   dementia
   chronic pulmonary disease
   connective tissue disease
   ulcer disease
   mild liver failure
   diabetes

2 = hemiplegia
   moderate or severe renal disease
   diabetes with end organ damage
   any tumor
   leukemia
   lymphoma

3 = moderate to severe liver disease

6 = metastatic solid tumor
   AIDS
examples of item definitions in the Charlson Index

1. MI = >1 definite or probable event, hospitalization with enzyme +/- ECG changes, but not simple ECG changes
2. Diabetes
   1. mild= oral or insulin but excludes diet rx
   2. mod= hospitalized, juvenile or brittle
   3. severe= an 'opathy
3. tumor = any solid tumor without mets within 5 years
4. respiratory system
   1. = chronic pulmonary disease
Appendix 4

The Index of Co-existent Disease

objective: to measure coexistent disease that might influence outcome

Items (Individual Disease Scale only) Severity Scale

1. organic heart disease
2. ischemic heart disease
3. arrhythmia
4. Congestive Heart failure
5. hypertension
6. cerebral vascular disease
7. peripheral vascular disease
8. diabetes
9. respiratory system
10. cancer
11. liver
12. renal
13. arthritis
14. Gastrointestinal system

0 = no coexistent disease
1 = asymptomatic to mild
2 = mild to moderate with symptoms and requiring Rx
3 = mod to severe, uncontrolled condition and on RX
4 = severe and uncontrolled
examples of items in the Index of Co-existent Disease

1. Ischemic heart disease
   1. = abn ECG or mild exertional angina
   2. = past MI or CABG but stable, daily living exertional angina
   3. = MI< 6 mos, CHF, restrictive angina
   4. = acute MI, Pulm edema, arrest

2. Respiratory system
   1. = chronic cough
   2. = productive cough, exertional dyspnea, mild abn PFTs
   3. = abn PFTs, dyspnea at rest, recurrent infections
   4. = resp failure
### Appendix 5

The Chronic Disease Scale

<table>
<thead>
<tr>
<th>Chronic Disease</th>
<th>Medication Level</th>
<th>Scoring Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart disease</td>
<td>1. Anticoagulants</td>
<td>One class = 3</td>
</tr>
<tr>
<td></td>
<td>2. Cardiac agents, ACE inhibitors</td>
<td>Two classes = 4</td>
</tr>
<tr>
<td></td>
<td>3. Loop diuretics</td>
<td>Three classes = 5</td>
</tr>
<tr>
<td>Respiratory disease</td>
<td>1. Isoproterenol</td>
<td>One class = 2</td>
</tr>
<tr>
<td></td>
<td>2. Betaadrenergics</td>
<td>Two or more classes = 3</td>
</tr>
<tr>
<td>Asthma, rheumatism</td>
<td>Xanthines</td>
<td>3</td>
</tr>
<tr>
<td>Rheumatiod arthritis</td>
<td>Bronchodilators &amp; Misc</td>
<td>3</td>
</tr>
<tr>
<td>Cancer</td>
<td>Antineoplastics</td>
<td>3</td>
</tr>
<tr>
<td>Parkinson's disease</td>
<td>L-Dopa</td>
<td>3</td>
</tr>
<tr>
<td>Hypertension</td>
<td>1. Antihypertensives (except ACE)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2. Beta Blockers or Diuretics</td>
<td>1 (unless 2 classes)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>Insulin</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Oral agents</td>
<td>2</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>Anticonvulsants</td>
<td>2</td>
</tr>
<tr>
<td>Asthma or rhinitis</td>
<td>Cromolyn</td>
<td>2</td>
</tr>
<tr>
<td>Acne</td>
<td>RX</td>
<td>1</td>
</tr>
<tr>
<td>Upper GI ulcers</td>
<td>Cimetidine</td>
<td>1</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>Miotic Drops</td>
<td>1</td>
</tr>
<tr>
<td>Gout</td>
<td>Allopurinol</td>
<td>1</td>
</tr>
<tr>
<td>High cholesterol</td>
<td>Antilipemics</td>
<td>1</td>
</tr>
<tr>
<td>Migraine</td>
<td>Ergots</td>
<td>1</td>
</tr>
<tr>
<td>TB</td>
<td>Antitubercular Drugs</td>
<td>1</td>
</tr>
</tbody>
</table>
### ESTABLISH A NEW PATIENT

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATE</td>
<td></td>
</tr>
<tr>
<td>LAST NAME</td>
<td></td>
</tr>
<tr>
<td>FIRST NAME</td>
<td></td>
</tr>
<tr>
<td>MED REC</td>
<td></td>
</tr>
<tr>
<td>DOB</td>
<td></td>
</tr>
<tr>
<td>ENT</td>
<td></td>
</tr>
</tbody>
</table>

### PL BASELINE DATA

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>city</td>
<td>Kingston, Trenton, Belleville, Brockville, Smiths Falls, Cornwall, Perth, Picton, Napanee, Gananoque, SharbotLake New York, Oshawa, Cobourg, Peterborough, Madoc, Campbellford, other</td>
</tr>
<tr>
<td>sex</td>
<td>male, female</td>
</tr>
<tr>
<td>nuprim</td>
<td>number of primary</td>
</tr>
<tr>
<td>site</td>
<td>oral, oropharynx, larynx, hypopharynx, nose etc, chemodactoma, major salivary, minor salivary, unknown</td>
</tr>
<tr>
<td>stororal</td>
<td>lip, floor mouth, tongue, buccal mucosa, alveolus, hard palate</td>
</tr>
<tr>
<td>storophra</td>
<td>base tongue, tonsil pillar, tonsil fossa, soft palate, lateral wall, posterior wall</td>
</tr>
<tr>
<td>stthypo</td>
<td>post wall, lateral wall, pyriform fossa, postcricoid</td>
</tr>
<tr>
<td>stlarynx</td>
<td>suprahypopharynx, infrahypopharynx, false cord, aryepi, vocal cord, subglottic</td>
</tr>
<tr>
<td>stnasal</td>
<td>nasopharynx, vestibule, cavity, max sinus, ethmoid sinus, sphenoid sinus</td>
</tr>
<tr>
<td>stsaliva</td>
<td>parotid, submandibular, minor salivary</td>
</tr>
<tr>
<td>Tstage</td>
<td>TX, T0, TIS, T1, T2, T3, T4, T?</td>
</tr>
<tr>
<td>Mstage</td>
<td>NX, NO, N1, N2a, N2b, N2c, N3</td>
</tr>
<tr>
<td>metasta</td>
<td>M0, M1</td>
</tr>
<tr>
<td>histolog</td>
<td>squamous, lymphoepithelial, spindle, verrucous, lymphoma, adenoidcystic, mucous, plasmacytoma, melanoma, chondroid, carcinoid, undiff, basal cell, oat cell</td>
</tr>
<tr>
<td>fstseen</td>
<td>date first in clinic</td>
</tr>
<tr>
<td>fstdct</td>
<td>date of diagnosis</td>
</tr>
<tr>
<td>sxplace</td>
<td>Kingston, elsewhere</td>
</tr>
</tbody>
</table>

(version March 1994)
P2 = THERAPY

- date
- name
- ent #
- prim #
- policy
  none exists, followed, not followed, individualized, altered
- poinot
  did not want radiotherapy, no treatment, did not want Sx, refused treatment
- therinit
  surgery, radiotherapy, comb sx&Rt, comb rt&Sx, chemo, no active, comb chemo&Rt
- substher
  surgery, radiotherapy, comb Sx&Rt, comb Rt&Sx, chemo
- intent
  cure, salvage, palliation, no treatment
- probabil
  cure rate
- visit
  consult, surgery, radiotherapy, chemotherapy, ther flup, routine flup, specific problem, info, office
- RT
  none, start, continue, delay, change, end, discontinued
- RTprim
  implant, local, local & neck, no primary, neck only;
  implant & primary, implant & primary & neck
- RTneck
  none, ipsi, contra, both
- RTdose
- RTcomplic
  necrosis bone, necrosis skin, CNS damage, edema,
  necrosis mucosa, poor healing, dysphagia, hospitalized
- SXprim
  none, incis biopsy, excis biopsy, local resn, wide
- SXneck
  radical, modified
- SXneckco
  radical, modified
- trach
  yes, no
- surgeon
  KENT, KGS, KPS, JENT, OGS, OPS, other
- CT(chemo)
  none, start, continue, delay, change, end

P3 = follow up

- response
  initial, no change, PR, CR, progression, relapse
- relapse
  no, yes, equivocal
- RLPSTST
  primary, neck, distant, primary & neck, primary & distant
- RLPSTRTD
  radiation, surgery, both, none
- current
  initial, A&W, AWD, DWD, DWOD, DWOD-AC, DWD-AC, lost
- another
  lung, esophagus, breast, GU, myeloma, gyn, rectal,
  lymphoma, renal, CNS, colon, melanoma

(version March 1994)
Form for comorbidity extraction project

Pat Key Number: ____________________

KRCC Chart Number: ________________

D.O.B. (dd/mmm/yyyy): _______________

Patient's Name: _______________________

Was there any mention of comorbid diseases in chart: Yes/No ___

If yes, please complete the following:

<table>
<thead>
<tr>
<th>Item 1: The Cardiac system</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIRS * includes all heart disorders</td>
</tr>
<tr>
<td>cvs</td>
</tr>
<tr>
<td>0 1= remote MI, occ angina</td>
</tr>
<tr>
<td>2= rx for CHF or angina or AF or arrhy</td>
</tr>
<tr>
<td>3= MI within 5 yrs, abn stress test, angioplasty, Cabg(coronary artery bypass graft)</td>
</tr>
<tr>
<td>4= unstable or intractable</td>
</tr>
<tr>
<td>Score ___</td>
</tr>
<tr>
<td>ICED-IDS * each system is separate item but only if independent disease</td>
</tr>
<tr>
<td>ischemic</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1= occ angina</td>
</tr>
<tr>
<td>2= past MI, past cabg, angina daily</td>
</tr>
<tr>
<td>3= MI within 6 mos, angina, restricted, CHF</td>
</tr>
<tr>
<td>4= acute MI, arrest, acute Pulm edema</td>
</tr>
<tr>
<td>Score ___</td>
</tr>
<tr>
<td>arrhy</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1= no rx</td>
</tr>
<tr>
<td>2= rx or pacer</td>
</tr>
<tr>
<td>3= uncontrolled</td>
</tr>
<tr>
<td>4= arrest</td>
</tr>
<tr>
<td>Score ___</td>
</tr>
<tr>
<td>CHF</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1= past event</td>
</tr>
<tr>
<td>2= on rx or mild symptoms</td>
</tr>
<tr>
<td>Score ___</td>
</tr>
<tr>
<td>hypertension</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1= abn EKG or echo</td>
</tr>
<tr>
<td>2= on rx and stable, CHF</td>
</tr>
<tr>
<td>3= unstable</td>
</tr>
<tr>
<td>Score ___</td>
</tr>
</tbody>
</table>

organic( valves, myopathy) 0 1= abn EKG or echo 2= on rx and stable, CHF 3= unstable

Score ___
ICED-FS
Circulation
0
1 = SOBOE, chest pain, dizzy, pacer, assisted walking
2 = heart failure, bedridden

Score ____

Kaplan Feinstein
cvs
0
1 = MI > 6 mos, abn EKG, AF
2 = stable angina or CHF > 6 mos
3 = < 6 mos MI, CHF, chest pain, arryth

Score ____

CDS
group 1 = anticoag
group 2 = cvs drugs including ACE
group 3 = loop diuretic
cds1

Score ____

Charison
MI = 1 (definite or probable in hosp)
CHF = 1 (past or present +/- rx)

Score ____

Item 2: The Respiratory system
CIRS • includes cancer
resp
0
1 = recurrent bronchitis, prn inhalers, <20 pack yrs
2 = recurrent pneumonia, daily meds or inhalers, >40 pack years, abn XRAY
3 = dyspnea, steriods, >40 pack years
4 = home oxygen, resp failure requiring admission/vent

Score ____

If lung cancer:
1 = in the remote past and clear
2 = within 5 years but clear
3 = major treatment in past 5 years
4 = recurrent, failed, palliative

cirs1

Score ____

ICED-IDS
resp
0
1 = chronic cough
2 = morning prod cough, SOBOE
3 = dyspnea, abn PFTs
4 = resp failure

Score ____
ICED-FS
resp 0
1=SOB, chronic cough, 1 block dyspnea
2=trach, home oxygen, respirator

Kaplan Feinstein
resp 0
1=recurrent asthma, abn xray, abn pfts
2= recurrent pneumonia, COPD, mild soboe
3= severe COPD or asthma

CDS
respiratory: any oral rx or inhaler =2
any oral tx and inhaler, asthma: steroid = 3
antituberculosis = 1

Score

Charlson COPD=1 (any degree)

Item 3: Hypertension

ICED-IDS • see cardiac
ICED-FS n/a

Kaplan Feinstein
hypert 1=mild diastolic, 1st line tx or 1 drug
2= moderate diastolic or complics such as headache, 2nd line tx or 2 drugs
3= severe, uncontrolled with tx, 3 or more drugs

Score

CDS
B blocker or diuretic=1
anti hyper = 2
ACE or Ca channel = 4

Score

Charlson n/a

Item 4: The Peripheral Vascular system

ICRS • includes hypertension
pvd 0
1=hypertension without rx
2= one rx for bp, claudication, small AA
3= >1 rx for bp, mod pvd
4= prev vascular surg or AA> 4 cm

Score
ICED-IDS
pvd 0
1 = past DVT or vasc surg
2 = claudication
3 = past PE, ulcer, venous obstruction, rest pain
4 = gangrene

Score
ICED-FS n/a

Kaplan Feinstein
pvd 0
1 = claudication, prev amputation
2 = recent amputation or gangrene
3 = n/a

Score
CDS n/a

Charison
pvd = 1 (claudication, postop or AA)
Score
Item 5: Diabetes Mellitis
CIRS * see misc item

ICED-IDS
dm 0
1 = chemical
2 = controlled
3 = uncontrolled and end organ damage
4 = end stage systemic

Score
ICED-FS n/a

Kaplan Feinstein
Dm 1 = controlled with meds
2 = unstable, juvenile or brittle
3 = uncontrolled, end organ failure

Score

CDS any rx = 2
Score

Charison
diabetes = 1 (must be on rx)
end organ diabetes = 2
Score
Item 6: The gastro-intestinal system

CIRS
- includes cancer
- both Upper GI and Lower GI are separate items
  upper GI
  0
  1= Hiatus hernia / heartburn with prn Rx
  2= H2 blocker or previous ulcer
  3= ulcer, dysphagia, +ve stool for blood
  4= previous perf ulcer or GI bleed

Score____
If upper GI cancer:
  1= in the remote past and clear
  2= within 5 years but clear
  3= major treatment in past 5 years
  4= recurrent, failed, palliative

Score____
lower GI
  0
  1= occ constipation, hemmorhoids, previous hernia
  2= constipation with meds, diverticular disease, hernia
  3= constipation problem, quiet diverticulitis
  4= lower GI bleed, active diverticulitis, prev bowel obstruction,

Score____
If lower GI cancer
  1= in the remote past and clear
  2= within 5 years but clear
  3= major treatment in past 5 years
  4= recurrent, failed, palliative

ICED-IDS
GI
  0
  1= past ulcer; mild diverticular disease, gastritis, irritable bowel
  2= active ulcer, mild ulcerative colitis, previous polyp, reflux
  3= abdominal complication of GI disease
  4= systemic complication of GI disease

Score____
ICED-FS
* each one is separate
  feeding
  0
  1= slight problem such as food cut
  2= paralysis, cannot feed, cannot eat, G tube, anorexia

Score____
  fecal
  0
  1= chronic disorder, pain, occ incontinence, any ostomy
  2= incontinent

Score____
Kaplan Feinstein
GI
  0
  1= +ve stool blood, symptomatic gall stones, ulcer, pancreatitis
  2= prev GI bleed with transfusion, recent pancreatitis, malabsorption
  3= major GI bleed (over 6 units)

Score____
CDS
  Ranitidine = 1
  0
Charlson
ulcer disease = 1 (on treatment or past GI bleed)

Score ______

Item 7: The Hepato-Biliary system

CIRS
- includes alcohol
- includes cancer

Hep 0
1 = past chole or old hepatitis, present daily limited alcohol or past (not recent) heavy drinker
2 = gall stones, prev hepatitis, daily or heavy alcohol/heavy drinker
3 = abn LFTs, chronic pancreatitis, complications of alcohol ie. hospitalizations, DTs, etc.
4 = hepatitis, pancreatitis

Score ______
If Hepato—Biliary cancer:
1 = in the remote past and clear
2 = within 5 years but clear
3 = major treatment in past 5 years
4 = recurrent, failed, palliative

Score ______

ICED-IDS
Hep 0
1 = prev hepatitis, mild cirrhosis
2 = recent hepatitis, duct obstruction
3 = chronic hepatitis or portal hypertension
4 = liver failure

Score ______

ICED-FS n/a

Kaplan Feinstein
- for alcohol see misc
Hep 0
1 = chronic liver disease (lab)
2 = liver disease (clinical)
3 = complication of liver failure

Score ______

CDS n/a

Charlson
mild liver disease = 1 (cirrhosis or chronic hep)
mod to severe liver disease = 3 (complications of liver failure)

Score ______

Item 8: Other Cancer sites
If other cancer, list cancer site: ________
If more than one prior cancer, list second site: ________
CIRS: * see each system
* for melanoma, see misc.
ICED * any site excluding skin
1=>5 yrs ago
2=<5 yrs ago
3=present, rx within the year
4= terminal

Score

Kaplan Feinstein
2= controlled
3= uncontrolled

Score

CDS chemotherapy=3

Score

Charlson
any tumor = 2 (any solid without mets)
leukemia/lymphoma =2
metastatic =6

Score

Item 9: The neurological system

CIRS * are separate items
neuro
0
1= headache prn rx, possible TIA
2= chronic headache, TIA, cva without deficit, mild degenerative disease
3= cva with mild deficit, post op neurosurg, mod degen
4= cva with deficit, severe degen

Score

psych
0
1= prev problem, occ rx, mild early dementia
2= prev major depression, mild dementia (established), prev admission,
3= depression requiring daily rx, substance abuse, moderate dementia
4= current active treatment or institutionalized

Score

ICED-IDS
* is CVA only and excludes psych
neuro
0
1= 1 TIA
2= TIAs, cva no deficit
3= frequent TIAs, cva with deficit, Sub arachnoid hemor.
4= coma

Score

ICED-FS * are separate items
neuro
0
1=dizzy, numbness, controlled seizures, syncope
2= ataxia, paralysis, uncontrolled seizures, bedridden

Score

mental
0
1= mild depression, irrational thinking, suicide, forgetfulness
2= confused, psychotic, depression, dementia

Score

char8
cdf8
cirsns
cirseps
celd9
icelden
icedef9m
speech 0
1= minor problem, slurring, prosthesis
2= cannot speak or be understood

Kaplan Feinstein
* includes psych
neuro 0
1=past cva or TIA, seizures
2=past cva with deficit, recent TIA's, status epilepticus
3= recent cva, suicidal, coma

Score ______

CDS * excluded antidepressants/antipsychotics etc
L-dopa = 3
anticonvulsants = 2
migraines (ergot) = 1

Score ______

Charlson
* includes psych
  cva=1 (no deficit)
  hemiplegia=2
  dementia=1

Score ______

CIRS
* includes cancer
renal 0
1= past/asymp kidney stone or pyelo
2= mild failure
3= renal failure with rx
4= dialysis

Score ______

If renal cancer:
1= in the remote past and clear
2= within 5 years but clear
3= major treatment in past 5 years
4= recurrent, failed, palliative

Score ______

ICED-IDS
renal 0
1= past infections or stone
2= chronic infections, stones
3= failure, dialysis, obstruction, transplant
4= end stage failure

Score ______

ICED-FS n/a

Kaplan Feinstein
renal 0
1= recurrent lower tract infections, stones or proteinuria
2= renal obstruction (hydronephrosis), mild kidney failure, recurrent renal failure
3= uremia with complications (+ or - renal failure)

Score ______
**CDS** n/a

**Charlson**
mod to severe renal = 2 (dialysis, transplants, or lab)

**Score**

**CIRS**
* includes cancer

- gu 0
  1 = stress incontinence, past hyster, BPH asymt
  2 = abn pap, recurrent infection, incontinence, BPH +/- TURP
  3 = Carcinoma In Situ of cervix or prostate, hematuria,
  4 = obstruction

**Score**
If GU Cancer:
1 = in the remote past and clear
2 = within 5 years but clear
3 = major treatment in past 5 years
4 = recurrent, failed, palliative

**CIRS**
* includes locomotion
- includes melanoma
+ includes cancer
- excludes vasculitis

**m/s** 0
- 1 = prn rx for arthritis, non melanoma cancer in past
- 2 = daily rx, assistive devices, past node neg melanoma
- 3 = steroids for arthritis, complicated osteoporosis, devices
- 4 = wheel chair or severe impaired, osteomyelitis, metastatic melanoma

**Score**

**Item 11: The GU system**

**Item 12: The extended Musculo skeletal system**
If bone cancer:

1 = in the remote past and clear
2 = within 5 years but clear
3 = major treatment in past 5 years
4 = recurrent, failed, palliative

Score

ICED-IDS
- includes vasculitis, locomotion
  m/s  0
  1 = mild/mod arthritis
  2 = recent total joint replacement, systemic Rheumatoid
  3 = major deformity or severe motion limitation
  4 = complex vasculitis

Score

ICED-FS * are separate items
  ambulation 0
  1 = walking with assistance (cane or walker)
  2 = bedridden, wheelchair

Score

transfer 0
  1 = out of bed with difficulty
  2 = bedridden

Score

Kaplan Feinstein
- is locomotion only
- see misc for vasculitis

m/s  0
  1 = slight impaired mobility
  2 = confined mobility
  3 = bed to chair existence

Score

CDS
- steroids=3
- gold=3
- gout rx=1

Score

Item 13: Miscellaneous

Charlson: connective tissue disease = 1

Score

CIRS
* all are separate items
* some include cancer

m/s  0
  1 = chronic sinusitis, mild hearing loss
  2 = corrected vision, hearing aid, chronic sinusitis rx, dizzy rx
  3 = problems despite glasses or hearing aid
  4 = blind, deaf, laryngectomy, inner ear surgery

Score
blood
0
1=mild anemia
2=mod anemia or low white count
3=worse
4=

Score
If leukemia or lymphoma:
1=in the remote past and clear
2=within 5 years but clear
3= major treatment in past 5 years
4=recurrent, failed, palliative

Score
metabolic
0
1=DM without rx, thyroid replacement
2=dm on rx, Fibrocystic breast disease
3=electrolyte problem requiring admission, obesity
4=poorly controlled dm, adrenal insuff, systemic infection (AIDS)

Score
If breast or thyroid cancer:
1=in the remote past and clear
2=within 5 years but clear
3= major treatment in past 5 years
4=recurrent, failed, palliative

ICED-IDS n/a
ICED-FS
• all are separate items
  - vision 0
    1=mild problem
    2= major problem

Score
hearing 0
1=hard of hearing, 1 hearing ear but not just 1 aid
2=deafness

Score
Kaplan Feinstein
• includes vasculitis
• separate items
  - misc 0
    1=chronic infection, HIV +ve
    2=collagen vasc disease controlled
    3=collagen vasc disease uncontrolled, AIDS

Score
Alcohol
1=a drinking problem
2=complication (pancreatitis, dts, psych problems)
3=severe (dts, seizures)

CDS
• acne rx=1
  • glaucoma=1
  • cholesterol=1

Score
Charlson AIDS=6 char13
Score

Comments:


Form for comorbidity extraction project

Pat Key Number: ________________

KRCC Chart Number: ________________

D.O.B. (dd/mmm/yyyy): ________________

Patient's Name: ________________________________

Verbatim description of all comorbid diseases and drugs (include dose) until beginning of treatment:

__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
__________________________________________________________________________
The PC-based data abstraction/entry module