A COMPARATIVE ANALYSIS OF FIVE INSTRUMENTS TO MEASURE CONTROL OF ASTHMA

By

DAWN TRAVERSE

A Project

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (1997) McMaster University

(Statistics) Hamilton, Ontario

TITLE: A Comparative Analysis of Five Instruments to Measure Control of Asthma

AUTHOR: Dawn Traverse

SUPERVISOR: Dr. S.D. Walter

NUMBER OF PAGES: xii, 77

Abstract

The control of asthma is a very important part of an asthmatic's life. Decreasing control can lead to asthma attacks, which can be fatal. For this reason, the researchers have set out to create an instrument to measure control of asthma, and have suggested five possible instruments.

Before an instrument can be used, it must be shown to be reliable, valid, and responsive. Reliability will be shown using various intraclass correlation coefficients, depending on the model being used for the data. Construct validity will be shown by how well the instrument's correlation coefficients with other instruments correspond to a priori predictions. Responsiveness will be shown by three methods, t tests comparing the change in changers and stable subjects, a responsiveness index, and Receiver Operating Characteristic (ROC) curves.

All five instruments are shown to be highly reliable. No conclusions can be drawn as of yet about the validity, as the a priori predictions have yet to made.

For all three methods of assessing responsiveness, the five instruments were ranked identically.

In choosing the best instrument, no final decisions can be made, as validity has yet to be shown, but at this time it appears as though the simplest instrument (involving only five questions on the patient's asthma symptoms asked at a visit to a clinic) is also the best as it has high reliability and is highly responsive.

Acknowledgements

Many thanks go out to my supervisor, Dr. S.D. Walter for his patience, guidance and suggestions throughout this project.

Thanks also go to Professor E. Juniper for all her insights into the world of asthma and for letting me analyze the data.

Thanks also go to Lauren Griffith for her help, and for the time she spent answering my many questions.

I would also like to thank my examiners, Dr. P.D.M. Macdonald and Dr. H. Shannon for their helpful comments and suggestions.

Finally, a great big thank you to my friends, my sisters, and my wonderful parents for their support and encouragement through all my years of school, but mostly for their love through all my life.

Contents

1 Introduction **1**

1.1 Asthma 1

1.1.1 Some Definitions 2

1.1.2 Control of Asthma 3

1.2 Instruments 4

1.2.1 Diary and Questionnaire 4

1.2.2 Questions 5

1.2.3 The Different Instruments 7

1.2.4 Summarizing the Instruments 9

1.2.5 Clinician's Global Rating of Change 9

1.3 Criteria for Instruments 9

1.3.1 Reliability 10

1.3.2 Validity 10

1.3.3 Responsiveness 10

2 Reliability **12**

2.1 General Concepts 12

2.2 Statistical Background 14

2.2.1 Model 1 14

2.2.2 Model 2 18

3 Validity **26**

3.1 General Concepts 26

3.1.1 Construct Validity 27

4 Responsiveness **28**

4.1 General Concepts 28

4.2 Statistical Background 30

4.2.1 Responsiveness Index 30

4.2.2 ROC Curve 30

5 Methods **33**

5.1 Subjects 33

5.2 Study Design 34
5.3 Data Analysis 35
5.3.1 Reliability 36
5.3.2 Validity 36
5.3.3 Responsiveness 36

6 Results **38**

6.1 General Description of the Data 39
6.2 Reliability 39
6.3 Validity 39
6.4 Responsiveness 39
6.5 Tables and Figures 40

7 Discussion **52**

7.1 Methods 52
7.1.1 Subjects and Study Design 52
7.1.2 Data Analysis 53
7.2 Results 55
7.2.1 Reliability 55
7.2.2 Validity 56

7.2.3 Responsiveness 56

7.3 Conclusions 57

7.4 Further Study 59

A Glossary **61**

B Diary **65**

C Questionnaire **68**

D Coding **71**

Bibliography **75**

List of Figures

6.1 95% Confidence Intervals for the Intraclass Correlation Coefficients for the Instruments 44

6.2 95% Confidence Intervals for the Responsiveness Index for the Instruments 49

6.3 95% Confidence Intervals for the Area Under the Curve for the Instruments 50

6.4 ROC Curves for Instruments 51

List of Tables

1.1 Summary of the Questions on the Diary and the Questionnaire 7

1.2 Summary of the Five Instruments 8

6.1 Characteristics of the Subjects (n=50) 40

6.2 Characteristics of the Subjects' Asthma 41

6.3 Mean (Standard Deviation) of the Scores on the Instruments 42

6.4 Mean (Standard Deviation) of the Change in Scores on the Instruments 42

6.5 Intraclass Correlation Coefficients for the Instruments 43

6.6 95% Confidence Intervals for the Intraclass Correlation Coefficients for the Instruments 43

6.7 Correlation Coefficients of the Scores on the Instruments and Other Instruments 45

6.8 Correlation Coefficients of the Change in Scores on the Instruments and Other Instruments 46

6.9 Mean (Standard Deviation) of Change and t Tests on the Instruments 47

6.10 Responsiveness Index and 95% Confidence Interval for the Responsiveness Index for the Instruments 48

6.11 Area Under the ROC Curve and 95% Confidence Interval for the Area for the Instruments 48

Chapter 1

Introduction

1.1 Asthma

Although asthma* [1] is a very common disease, there is no set definition. The most useful definition of asthma is that it is 'a disease characterized by wide variations over short periods of time in resistance to flow in intrapulmonary airways' [2]. This resistance to flow is due to inflammation of the air passages, for a variety of reasons. Ongoing narrowing of the air passages and muscle spasms cause repeated episodes of breathing difficulty that most people recognize as asthma. Treatment of asthma must consider two facets of the disease: the severity and the control. In contrast

[1] Terms marked with an * are included in the glossary in Appendix A.

to most diseases, control of asthma is more important than its severity, i.e. well controlled severe asthma is not as serious a problem as is poorly controlled mild asthma.

1.1.1 Some Definitions

bronchodilator* A drug, used by most asthmatics, which works by relaxing the contracted bronchial muscles in the patient's air passages. The drug is most often taken by use of an inhaler, and is taken as needed.

peak expiratory flow rate (PEFR)* A measurement of the speed with which air can be forcibly blown out of the lungs taken as 'the maximum rate of expiration, maintained for 10 milliseconds, that occurs within the first second of forced expiration' [9]. The reading takes only a minute and reliably measures the narrowing of the air passages. The test is done using a peak flow meter; an inexpensive, portable and easy to use instrument, which most asthmatics are encouraged to buy. The test is done by having the patient fill their lungs with as much air as possible, and then blow out all the air as fast as they can. This process is repeated three times and the highest value is recorded as the patient's PEFR reading. The patient's reading is compared to their predicted PEFR value, which is calculated based on their age, sex and height.

forced expired volume in one second (FEV_1)* A measurement of 'the volume of air that can be forcibly exhaled in one second after a full breath has been taken' [8]. The reading is a reliable measure of how narrowed the air passages are. The test is done using a spirometer at a clinic, by having the patient breathe in deeply and then blow out as quickly as they can. This process is repeated three times and the highest value is recorded as the patient's FEV_1 reading. The patient's reading is compared against their predicted FEV_1 value, which is calculated based on their age, sex, and height.

1.1.2 Control of Asthma

Most acute* asthma attacks* occur after a period of worsening control of the patient's asthma. Since these attacks can lead to death, it is desirable to be aware of a decrease in the control of a patient's asthma, so that by intervening with suitable treatment, the attack may be avoided. Many asthmatics will notice, on their own, that their asthma is becoming more difficult to control, although about 20% of asthmatics have 'a very poor perception of moderate changes in the state of their asthma' [2].

The most common signs of the worsening of control of a patient's asthma are: sleep disturbance, an increase in symptoms upon waking up, shortness of breath, wheezing, limitations in activities, an increasing need to use their bronchodilator,

and a fall in PEFR and FEV_1 [2, 8, 9, 17].

1.2 Instruments

An instrument* is any piece of equipment (for example, a machine, or a questionnaire) that measures and provides information about something.

Since the control of asthma is an important aspect of the daily lives of asthmatics, and there are no instruments existing which attempt to measure it, the creation of such an instrument is an important step in the control of asthma.

The researchers: Prof. Elizabeth Juniper (principal investigator,) Dr. Gordon Guyatt, Dr. Paul O'Byrne, and Penelope Ferrie are interested in creating such an instrument.

1.2.1 Diary and Questionnaire

Two possible methods of administering such an instrument were considered by the researchers. The first was a week-long diary (see Appendix B) filled out each morning and night at home, while the second was a questionnaire (see Appendix C) filled out weekly at a clinic.

1.2.2 Questions

Both methods contain five symptom questions, one question on the use of the patient's bronchodilator, and one question on the patient's PEFR reading. In addition, the questionnaire includes one question on the patient's FEV_1 reading.

Symptom Questions

The five symptom questions were:

- How often the patient was woken during the night by their asthma
- How bad the patient's asthma symptoms were when they woke up in the morning
- How limited the patient was in their activities because of their asthma
- How much shortness of breath did the patient experience because of their asthma
- How much of the time did the patient wheeze

Each symptom question, for both the diary and the questionnaire methods, was rated on a seven point scale ranging from 0 (no symptoms) to 6 (very severe

symptoms.) The daily diary responses were averaged over the previous week, while the questionnaire was phrased to ask for the patient's average over the past week.

Bronchodilator Use

The question on bronchodilator use was also scored on a seven point scale ranging from 0 (no puffs per day) to 6 (more than 16 puffs per day.) The diary averaged the daily answers over the previous week, while the questionnaire asked about the use, on average, over the past week.

PEFR Reading

The PEFR reading, for the diary, was found by averaging the patient's daily PEFR readings over the previous week and then expressing it as a percent of the patient's predicted PEFR. The percent was then converted to a seven point scale ranging from 0 (greater than 95%) to 6 (less than 50%.) The questionnaire PEFR question was defined as above, but was based on the patient's reading taken at the clinic.

FEV_1 Reading

The FEV_1 reading for the questionnaire was expressed as a percent of the patient's predicted FEV_1. The percent was then converted to a seven point scale ranging from

0 (greater than 95%) to 6 (less than 50%.)

Table 1.1: Summary of the Questions on the Diary and the Questionnaire

Questions	Diary	Questionnaire
5 symptom questions	√	√
1 question on bronchodilator use	√	√
1 question on PEFR reading	√	√
1 question on FEV_1 reading		√

1.2.3 The Different Instruments

The two different methods of administering the instuments considered here are a daily diary, filled out each morning and night, and a weekly questionnaire. These two methods are being considered to see if it is neccessary for an asthmatic to fill out a daily diary for a week, or if a simple assessment in the clinic is sufficient. Daily diaries are often considered to be a hassle, and likely to be forgotten.

Since many under-developed countries do not have the facilities for PEFR and FEV_1 readings at their clinics, let alone at the asthmatic's home, nor do asthmatics have access to bronchodilators, two forms of the diary and questionnaire were con-

sidered, the first with only the five symptom questions, and the second also including the bronchodilator use and the PEFR questions.

It was felt that a question on bronchodilator use and a measure of the narrowing of the lung passages where neccessary, and PEFR is easier to measure, and can be measured at home. A third possible form of the questionnaire, with the five symptom questions, the bronchodilator use question, the PEFR question and the FEV_1 question was also considered. This extra alternative form of the questionnaire added on the question on the subject's FEV_1 reading, to examine if this improved the instrument.

Table 1.2: Summary of the Five Instruments

	Diary		Questionnaire		
Questions	Symptoms	All	Symptoms	All #1	All #2
5 Symptom Questions	√	√	√	√	√
Bronchodilator Use		√		√	√
PEFR Reading		√		√	√
FEV_1 Reading					√

1.2.4 Summarizing the Instruments

The response for each of the five instruments was summarized as the mean of the questions included (see Appendix D.) Since the questions are all scored on a seven point scale from 0 (best situation) to 6 (worst situation,) smaller scores represent better control of asthma.

1.2.5 Clinician's Global Rating of Change

A clinician also gave a rating of the change in the subjects control of their asthma (clinician's global rating of change) at the visits to the clinic (except for the first visit) relative to the subject's prior visit. The subject's control of their asthma was rated as having deteriorated, stayed stable or improved since their previous visit. This rating was based on various objective measures of the subject's asthma.

1.3 Criteria for Instruments

Before a new instrument is to be used or commonly accepted, it must meet certain criteria. The most common criteria are those of being reliable, valid, and, if the instrument is to be used to measure change over time, responsive.

1.3.1 Reliability

An instrument must be shown to be reliable*, that is, it must be shown to be measuring something in a reproducible way [15]. For example, test-retest reliability* concerns getting the same result on the same subject on different occasions, when we assume that the subject's score stayed stable.

1.3.2 Validity

Validity* is concerned with making sure that the instrument is measuring what was intended [16]. One can compare the instrument with other instruments already in use which are assumed to give the truth (criterion validity*.) If there are no such instruments available, one can test hypotheses that certain groups of people will score higher/lower/different than other groups of people on the instrument, or one can compare the new instrument to other instruments to see how correlated they are, and compare these correlations to a priori predictions* (construct validity*.)

1.3.3 Responsiveness

If the change in score, over a time period, of an instrument is to be used, then the instrument must also be shown to be responsive*. Note that the change in scores should be used (in preference to using only post-scores) only if the intraclass

correlation coefficient for the instrument exceeds 0.5, so that the variance between subjects is larger than the variance within subjects [15]. Responsiveness is concerned with the ability of an instrument to detect clinically important changes over time, even if the change is small [6].

Chapter 2

Reliability

2.1 General Concepts

Reliability, also called reproducibility, examines whether an instrument is giving the same score on different occasions. Here we are interested in test-retest reliability, examining scores on the instrument at different times when the score should, theoretically, remain the same.

The Pearson correlation coefficient, ρ, could be used to measure reliability, but it is not adequate since it only measures the relationship between the two scores, and not their agreement. A better way of measuring agreement of scores is to use the intraclass correlation coefficient, ρ_I.

Using the intraclass correlation coefficient, reliability is given as a 'ratio of the variability between individuals to the total variability in the scores' [15]. This ratio can range from 0 to 1, with 0 indicating no reliability, and 1 indicating perfect reliability.

Many authors have given guidelines on the interpretation of the intraclass correlation coefficient. One such set of guidelines is [10]:

$\rho_I < 0.4$ — poor reliability

$0.4 \leq \rho_I < 0.75$ — fair to good reliability

$\rho_I \geq 0.75$ — excellent reliability.

Another set of guidelines is [4]:

0.00-0.20 — slight reliability

0.21-0.40 — fair reliability

0.41-0.60 — moderate reliability

0.61-0.80 — substantial reliability

0.81-1.00 — almost perfect reliability.

2.2 Statistical Background

Depending on whether or not we are to consider the effect of the visit, there are two models to be considered. The first considers the visits as replicates, and thus considers their effect as part of the error of measurement. The second considers the visit as a term in the model, with a possible effect.

2.2.1 Model 1

Consider the following model of an individual's score on an instrument:

$$X_{ij} = \mu + \pi_i + \eta_{ij} \quad , i = 1, 2, \ldots, n$$

$$j = 1, 2, \ldots, k$$

where X_{ij} = observed measurement on person i at visit j.

μ = overall mean.

π_i = effect of person i.

η_{ij} = error of measurement.

π_i are independent N(0,σ_π^2) random variables.

η_{ij} are independent N(0,σ_η^2) random variables.

π_i and η_{ij} are independent.

The variance of X_{ij} can be represented as:

$$\begin{aligned} V(X_{ij}) &= V(\mu) + V(\pi_i) + V(\eta_{ij}) \\ &= 0 + \sigma_\pi^2 + \sigma_\eta^2 \\ &= \sigma_\pi^2 + \sigma_\eta^2 \end{aligned}$$

Using the rules for deriving the expected mean squares, as described in Winer, p.371-373 [18], we have:

$$\begin{aligned} E(\text{MSW}) &= \sigma_\eta^2 \\ E(\text{MSB}) &= k\sigma_\pi^2 + \sigma_\eta^2 \end{aligned}$$

We can then find the following unbiased estimators of σ_π^2 and σ_η^2 [5]:

$$\begin{aligned} \hat{\sigma}_\eta^2 &= \text{MSW} \\ \hat{\sigma}_\pi^2 &= \frac{\text{MSB} - \text{MSW}}{k} \end{aligned}$$

If we define the intraclass correlation coefficient as:

$$\rho_I = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\eta^2}$$

and define θ as:

$$\theta = \frac{\sigma_\pi^2}{\sigma_\eta^2}$$

we have that

$$\rho_I = \frac{\theta}{1+\theta} .$$

Biased Estimate

A simple, but biased, estimate of the intraclass correlation coefficient is found by using the estimator

$$\begin{aligned}\hat{\theta}_b &= \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\eta^2} \\ &= \frac{\frac{\text{MSB}-\text{MSW}}{k}}{\text{MSW}} \\ &= \frac{\text{MSB}-\text{MSW}}{k\text{MSW}}\end{aligned}$$

therefore we have the following estimate of the intraclass correlation coefficient [18] p248:

$$\begin{aligned}\hat{\rho}_{I_b} &= \frac{\hat{\theta}_b}{1+\hat{\theta}_b} \\ &= \frac{\frac{\text{MSB}-\text{MSW}}{k\text{MSW}}}{1+\frac{\text{MSB}-\text{MSW}}{k\text{MSW}}} \\ &= \frac{\text{MSB}-\text{MSW}}{\text{MSB}+(k-1)\text{MSW}} \ .\end{aligned}$$

This is a biased estimate of the intraclass correlation coefficient because the ratio of 2 unbiased estimators is not neccessarily an unbiased estimator of the ratio [18] p248. i.e.

$$E\left[\frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\eta^2}\right] \neq \frac{\sigma_\pi^2}{\sigma_\eta^2} \ .$$

Unbiased Estimate

To find an unbiased estimator of ρ_I, we must first find an unbiased estimator of θ.

We have that

$$\frac{(n-1)\text{MSB}}{\sigma_\eta^2+k\sigma_\pi^2} \sim \chi^2_{n-1}$$

$$\frac{n(k-1)\text{MSW}}{\sigma_\eta^2} \sim \chi^2_{n(k-1)}$$

since these two distributions are independent,

$$\frac{\text{MSB}}{\text{MSW}} \sim \frac{\sigma_\eta^2+k\sigma_\pi^2}{\sigma_\eta^2}\frac{\chi^2_{n-1}/n-1}{\chi^2_{n(k-1)}/n(k-1)}$$

Therefore, we have [18] p246:

$$\frac{\text{MSB}}{\text{MSW}} \sim \frac{\sigma_\eta^2+k\sigma_\pi^2}{\sigma_\eta^2}F_{[(n-1),n(k-1)]}$$

Since we defined θ as

$$\theta = \frac{\sigma_\pi^2}{\sigma_\eta^2}$$

We have:

$$\frac{\text{MSB}}{\text{MSW}} \sim (1+k\theta)F_{[(n-1),n(k-1)]}$$

Since $E[F_{\nu_1,\nu_2}] = \frac{\nu_2}{\nu_2-2}$

$$E\left[\frac{\text{MSB}}{\text{MSW}}\right] = E[1+k\theta]\frac{n(k-1)}{n(k-1)-2}$$

Solving for $E[\theta]$:

$$\begin{aligned} 1+E[k\theta] &= E\left[\frac{\text{MSB}}{\text{MSW}}\right]\frac{n(k-1)-2}{n(k-1)} \\ kE[\theta] &= E\left[\frac{\text{MSB}}{\text{MSW}}\right]\frac{n(k-1)-2}{n(k-1)} - 1 \\ E[\theta] &= \frac{1}{k}\left[E\left[\frac{\text{MSB}}{\text{MSW}}\right]\frac{n(k-1)-2}{n(k-1)} - 1\right] \end{aligned}$$

Let $E[\theta] = \hat{\theta}_u$ and solve for $\hat{\theta}_u$:

$$\begin{aligned}\hat{\theta}_u &= \frac{[n(k-1)-2]\hat{\text{MSB}} - n(k-1)\hat{\text{MSW}}}{kn(k-1)\hat{\text{MSW}}} \\ &= \frac{\hat{\text{MSB}} - \frac{n(k-1)}{n(k-1)-2}\hat{\text{MSW}}}{k\frac{n(k-1)}{n(k-1)-2}\hat{\text{MSW}}} \\ &= \frac{\hat{\text{MSB}} - m\hat{\text{MSW}}}{km\hat{\text{MSW}}}\ , \ where\ m = \frac{n(k-1)}{n(k-1)-2}\end{aligned}$$

Now, substituting to solve for $\hat{\rho}_I$ [18] p287:

$$\begin{aligned}\hat{\rho}_{I_u} &= \frac{\hat{\theta}_u}{1+\hat{\theta}_u} \\ &= \frac{\frac{\hat{\text{MSB}}-m\hat{\text{MSW}}}{km\hat{\text{MSW}}}}{1+\frac{\hat{\text{MSB}}-m\hat{\text{MSW}}}{km\hat{\text{MSW}}}} \\ &= \frac{\hat{\text{MSB}} - m\hat{\text{MSW}}}{\hat{\text{MSB}} + (k-1)m\hat{\text{MSW}}}\ , \ where\ m = \frac{n(k-1)}{n(k-1)-2}\ .\end{aligned}$$

Confidence Interval for ρ_I

An approximate $(1-\alpha)100\%$ confidence interval for ρ_I is [10]:

$$\left(\frac{\frac{F}{F_{n-1,n(k-1),1-\frac{\alpha}{2}}}-1}{k+\frac{F}{F_{n-1,n(k-1),1-\frac{\alpha}{2}}}-1}\ ,\ \frac{\frac{F}{F_{n-1,n(k-1),\frac{\alpha}{2}}}-1}{k+\frac{F}{F_{n-1,n(k-1),\frac{\alpha}{2}}}-1}\right)\ ,where\ F = \frac{\hat{\text{MSB}}}{\hat{\text{MSW}}}\ .$$

2.2.2 Model 2

Consider the following model of an individual's score on an instrument:

$$X_{ij} = \mu + \pi_i + \alpha_j + \eta_{ij} \quad , i = 1, 2, \ldots, n$$

$$j = 1, 2, \ldots, k$$

where X_{ij} = observed measurement on person i at visit j.

μ = overall mean.

π_i = effect of person i.

α_j = effect of the visit j.

η_{ij} = error of measurement.

π_i are independent N(0,σ_π^2) random variables.

α_j are independent N(0,σ_α^2) random variables.

η_{ij} are independent N(0,σ_η^2) random variables.

π_i, α_j, and η_{ij} are pairwise independent.

The variance of X_{ij} can be represented as:

$$\begin{aligned} V(X_{ij}) &= V(\mu) + V(\pi_i) + V(\alpha_j) + V(\eta_{ij}) \\ &= 0 + \sigma_\pi^2 + \sigma_\alpha^2 + \sigma_\eta^2 \\ &= \sigma_\pi^2 + \sigma_\alpha^2 + \sigma_\eta^2 \end{aligned}$$

Using the rules for deriving the expected mean squares, as described in Winer p371-373 [18], we have:

$$E(\text{MSR}) = \sigma_\eta^2$$

$$E(\text{MSV}) = n\sigma_\alpha^2 + \sigma_\eta^2$$

$$E(\text{MSB}) = k\sigma_\pi^2 + \sigma_\eta^2$$

We can then find the following unbiased estimators of σ_η^2, σ_α^2 and σ_π^2 [5]:

$$\hat{\sigma}_\eta^2 = \text{MSR}$$

$$\hat{\sigma}_\alpha^2 = \frac{\text{MSV} - \text{MSR}}{n}$$

$$\hat{\sigma}_\pi^2 = \frac{\text{MSB} - \text{MSR}}{k}$$

Since the model is now more complicated, there are various forms of the intraclass correlation coefficient that can be considered. Recall that reliablility expressed as an intraclass correlation coefficient is a 'ratio of the variability between individuals to the total variability in the scores' [15]. Depending on whether or not we want to consider the variation due to the visit as part of the total variation in the scores there are two possibilities for ρ_I. The first does not consider the variability due to visits as part of the total variability in the scores, while the second does. The decision not to include the variability due to visits as part of the total variability in scores would be applicable here if, when the instrument will be used, a correction factor will be applied to an individuals score according to which visit the score is from. If no such correction factor will be used in the eventual application of the instrument, then the variability due to visits should be included as part of the total variability in scores.

$\rho_{I(1)}$

If we define the intraclass correlation coefficient as [18]:

$$\rho_{I(1)} = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\eta^2}$$

and define θ as:

$$\theta = \frac{\sigma_\pi^2}{\sigma_\eta^2}$$

we have that

$$\rho_{I(1)} = \frac{\theta}{1+\theta} .$$

Biased Estimate

A simple, but biased, estimate of the intraclass correlation coefficient is found by using the estimator

$$\begin{aligned} \hat{\theta}_{(1)_b} &= \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\eta^2} \\ &= \frac{\frac{\text{MSB}-\text{MSR}}{k}}{\text{MSR}} \\ &= \frac{\text{MSB} - \text{MSR}}{k\text{MSR}} \end{aligned}$$

therefore we have the following estimate of the intraclass correlation coefficient [18] p290:

$$\hat{\rho}_{I(1)_b} = \frac{\hat{\theta}_{(1)_b}}{1 + \hat{\theta}_{(1)_b}}$$

$$= \frac{\frac{\text{MSB}-\text{MSR}}{k\text{MSR}}}{1+\frac{\text{MSB}-\text{MSR}}{k\text{MSR}}}$$
$$= \frac{\text{MSB}-\text{MSR}}{\text{MSB}+(k-1)\text{MSR}}.$$

This is a biased estimate of the intraclass correlation coefficient because the ration of 2 unbiased estimators is not neccessarily an unbiased estimator of the ratio [18] p248. i.e.

$$E\left[\frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\eta^2}\right] \neq \frac{\sigma_\pi^2}{\sigma_\eta^2}$$

Unbiased Estimate

To find an unbiased estimator of $\rho_{I(1)}$, we must first find an unbiased estimator of θ.

We have that

$$\frac{(n-1)\text{MSB}}{\sigma_\eta^2+k\sigma_\pi^2} \sim \chi^2_{n-1}$$
$$\frac{(n-1)(k-1)\text{MSR}}{\sigma_\eta^2} \sim \chi^2_{(n-1)(k-1)}$$

since these two distributions are independent,

$$\frac{\text{MSB}}{\text{MSR}} \sim \frac{\sigma_\eta^2+k\sigma_\pi^2}{\sigma_\eta^2}\frac{\chi^2_{n-1}/n-1}{\chi^2_{(n-1)(k-1)}/(n-1)(k-1)}$$

Therefore, we have [18] p246:

$$\frac{\text{MSB}}{\text{MSR}} \sim \frac{\sigma_\eta^2+k\sigma_\pi^2}{\sigma_\eta^2}F_{[(n-1),(n-1)(k-1)]}$$

Since we defined θ as

$$\theta = \frac{\sigma_\pi^2}{\sigma_\eta^2}$$

We have:

$$\frac{\text{MSB}}{\text{MSR}} \sim (1 + k\theta) F_{[(n-1),(n-1)(k-)]}$$

Since $E[F_{\nu_1,\nu_2}] = \frac{\nu_2}{\nu_2 - 2}$

$$E\left[\frac{\text{MSB}}{\text{MSR}}\right] = E[1 + k\theta]\frac{(n-1)(k-1)}{(n-1)(k-1)-2}$$

Solving for $E[\theta]$:

$$\begin{aligned} 1 + E[k\theta] &= E\left[\frac{\text{MSB}}{\text{MSR}}\right]\frac{(n-1)(k-1)-2}{(n-1)(k-1)} \\ kE[\theta] &= E\left[\frac{\text{MSB}}{\text{MSR}}\right]\frac{(n-1)(k-1)-2}{(n-1)(k-1)} - 1 \\ E[\theta] &= \frac{1}{k}\left[E\left[\frac{\text{MSB}}{\text{MSR}}\right]\frac{(n-1)(k-1)-2}{(n-1)(k-1)} - 1\right] \end{aligned}$$

Let $E[\theta] = \hat{\theta}_{(1)_u}$ and solve for $\hat{\theta}_{(1)_u}$:

$$\begin{aligned} \hat{\theta}_{(1)_u} &= \frac{[(n-1)(k-1)-2]\hat{\text{MSB}} - (n-1)(k-1)\hat{\text{MSR}}}{k(n-1)(k-1)\hat{\text{MSR}}} \\ &= \frac{\hat{\text{MSB}} - \frac{(n-1)(k-1)}{(n-1)(k-1)-2}\hat{\text{MSR}}}{k\frac{(n-1)(k-1)}{(n-1)(k-1)-2}\hat{\text{MSR}}} \\ &= \frac{\hat{\text{MSB}} - m_{(1)}\hat{\text{MSR}}}{km_{(1)}\hat{\text{MSR}}}, \quad where\ m_{(1)} = \frac{(n-1)(k-1)}{(n-1)(k-1)-2} \end{aligned}$$

Now, substituting to solve for $\hat{\rho}_{I(1)}$ [18] p290:

$$\hat{\rho}_{I(1)_u} = \frac{\hat{\theta}_{(1)_u}}{1 + \hat{\theta}_{(1)_u}}$$

$$= \frac{\frac{\hat{\text{MSB}}-m_{(1)}\hat{\text{MSR}}}{km_{(1)}\hat{\text{MSR}}}}{1+\frac{\hat{\text{MSB}}-m_{(1)}\hat{\text{MSR}}}{km_{(1)}\hat{\text{MSR}}}}$$

$$= \frac{\hat{\text{MSB}}-m_{(1)}\hat{\text{MSR}}}{\hat{\text{MSB}}+(k-1)m_{(1)}\hat{\text{MSR}}}\ ,\ where\ m_{(1)}=\frac{(n-1)(k-1)}{(n-1)(k-1)-2}\ .$$

Confidence Interval for $\rho_{I(1)}$

In similar fashion to the confidence interval for ρ_I, an approximate $(1-\alpha)100\%$ confidence interval for $\rho_{I(1)}$ is given as:

$$\left(\frac{\frac{F}{F_{n-1,(n-1)(k-1),1-\frac{\alpha}{2}}}-1}{k+\frac{F}{F_{n-1,(n-1)(k-1),1-\frac{\alpha}{2}}}-1}\ ,\ \frac{\frac{F}{F_{n-1,(n-1)(k-1),\frac{\alpha}{2}}}-1}{k+\frac{F}{F_{n-1,(n-1)(k-1),\frac{\alpha}{2}}}-1}\right)\ ,where\ F=\frac{\hat{\text{MSB}}}{\hat{\text{MSR}}}\ .$$

$\rho_{I(2)}$

Secondly, consider the following definition of the intraclass correation coefficient [15]:

$$\rho_{I(2)}=\frac{\sigma_\pi^2}{\sigma_\pi^2+\sigma_\alpha^2+\sigma_\eta^2}.$$

Biased Estimate

A simple, but biased, estimate of $\rho_{I(2)}$ is [13]:

$$\hat{\rho}_{I(2)_b} = \frac{\hat{\sigma}_\pi^2}{\hat{\sigma}_\pi^2+\hat{\sigma}_\alpha^2+\hat{\sigma}_\eta^2}$$

$$= \frac{\frac{\hat{\text{MSB}}-\hat{\text{MSR}}}{k}}{\frac{\hat{\text{MSB}}-\hat{\text{MSR}}}{k}+\frac{\hat{\text{MSV}}-\hat{\text{MSR}}}{n}+\hat{\text{MSR}}}$$

$$= \frac{\hat{\text{MSB}}-\hat{\text{MSR}}}{\hat{\text{MSB}}+(k-1)\hat{\text{MSR}}+k\frac{\hat{\text{MSV}}-\hat{\text{MSR}}}{n}}\ .$$

Unbiased Estimate

No unbiased estimate of $\rho_{I(2)}$ was found in the literature.

Confidence Interval for $\rho_{I(2)}$

An approximate $(1-\alpha)100\%$ confidence interval for $\rho_{I(2)}$ is (c_1, c_2) [13], where:

$$
\begin{aligned}
c_1 &= \frac{n(\hat{\text{MSB}} - F_{n-1,\nu,1-\frac{\alpha}{2}}\hat{\text{MSR}})}{F_{n-1,\nu,1-\frac{\alpha}{2}}[k\hat{\text{MSV}} + (kn-k-n)\hat{\text{MSR}}] + n\hat{\text{MSB}}} \\
c_2 &= \frac{n(F_{\nu,n-1,1-\frac{\alpha}{2}}\hat{\text{MSB}} - \hat{\text{MSR}})}{k\hat{\text{MSV}} + (kn-k-n)\hat{\text{MSR}} + nF_{\nu,n-1,1-\frac{\alpha}{2}}\hat{\text{MSB}}} \\
\textit{where } \nu &= \frac{(k-1)(n-1)\{k\hat{\rho}_{I(2)}F_V + n[1+(k-1)\hat{\rho}_{I(2)}] - k\hat{\rho}_{I(2)}\}^2}{(n-1)k^2\hat{\rho}_{I(2)}^2F_V^2 + \{n[1+(k-1)\hat{\rho}_{I(2)}] - k\hat{\rho}_{I(2)}\}^2} \\
\textit{where } F_V &= \frac{\hat{\text{MSV}}}{\hat{\text{MSR}}} .
\end{aligned}
$$

Chapter 3

Validity

3.1 General Concepts

Validity examines whether the instrument is assessing what is intended. If a gold standard* (a gold standard is an instrument already in place that is assumed to tell the truth) exists, we would be interested in criterion validity, where we see how well the new instrument correlates with the 'gold standard.' If no such instrument exists (which is the situation here, as the instruments under study are the first instruments to attempt to measure control of asthma) we are interested in construct validity.

3.1.1 Construct Validity

Since a 'gold standard' does not exist, we have nothing to compare the new instrument against, so we create tests of hypotheses for which we believe we know the conclusion, and then test these hypotheses. For example, we may test if a certain group of people will score higher on the instrument than another group of people.

We can also compare the new instrument to instruments which measure other characteristics using the Pearson correlation coefficient, and compare the resulting correlations to a priori predictions made by asthma clinicians before the analysis was done.

Cross Sectional*

Cross sectional construct validity examines the data at single points in time, and calculates the correlation of these scores with the other instruments.

Longitudinal* (or Evaluative)

Longitudinal (or evaluative) construct validity examines the change in scores between two points in time, and looks at the correlation of these change scores with other instruments.

Chapter 4

Responsiveness

4.1 General Concepts

Responsiveness, also termed sensitivity to change, examines the ability of an instrument to detect clinically important changes. Here we are interested in whether the instruments we are considering are able to detect if a person's control of their asthma has stayed stable or has changed (either deteriorated or improved.) Alternatively, we are interested in whether or not the instruments we are considering are able to detect if a person's control of their asthma has deteriorated or not (either stayed stable or improved.)

Intuitively, it is desirable that the change scores in stable subjects are not

(clinically) significantly different from 0. Similarly, it is also desirable that the change scores in subjects who changed are (clinically) significantly different from 0.

There are various ways to measure an instrument's responsiveness. One of the most common is to compare, using a t-test, the change scores for subjects who stayed stable against the change scores for subjects who changed. Using this method, the instrument with the largest calculated t (or correspondingly, the smallest P-value) would be considered to be the most responsive [14]. Since, for this method, the P-values are influenced by the sample sizes, we would need equal sample sizes to compare instruments.

The responsiveness index, the absolute value of the ratio of the mean change in subjects who changed to the standard deviation of the change in scores for subjects who changed is another method to assess responsiveness [14].

Another method to assess an instrument's responsiveness is to consider the instrument as a diagnostic test* for detecting whether a person's control of their asthma has deteriorated or not, and, for different cut-off points, find the sensitivity* and specificity* and the corresponding receiver operating characteristic (ROC) curve* [3].

4.2 Statistical Background

4.2.1 Responsiveness Index

Define the responsiveness Index (RI):

$$RI = \text{absolute value}\left(\frac{\text{mean change in changers}}{\text{standard deviation of change in scores for changers}}\right).$$

To find the standard error of the responsiveness index, we must assume known variance for the change in changers:

$$\begin{aligned} SE(\hat{RI}) &= \sqrt{\frac{\frac{\sigma^2_{changers}}{n_{changers}}}{\sigma^2_{changers}}} \\ &= \frac{1}{\sqrt{n_{changers}}}. \end{aligned}$$

A $(1-\alpha)100\%$ confidence interval for the Responsiveness Index, using the t-distribution, is:

$$\left(\hat{RI} - t_{n_{changers}-1,1-\frac{\alpha}{2}} SE(\hat{RI}),\ \hat{RI} + t_{n_{changers}-1,1-\frac{\alpha}{2}} SE(\hat{RI})\right).$$

4.2.2 ROC Curve

Before considering the instrument to be a diagnostic test, we need some definitions. In the general case of a diagnostic test, we are interested in discriminating between patients with a condition and those without. Using a diagnostic test we make our

diagnosis, and then compare it to the truth. **Prevalence*** is the proportion of patients with the condition. A **false positive result** occurs when a subject is diagnosed as having the condition when they really do not. A **false negative result** occurs when a subject diagnosed as not having the condition really does. A **true positive result** occurs when a subject is diagnosed as having the condition and they really do. A **true negative result** occurs when a subject is diagnosed as not having the condition and they really do not. The proportion of subjects with the condition who are properly diagnosed is called **sensitivity**. The proportion of subjects without the condition who are properly diagnosed is called **specificity**.

If we now consider the instrument to be a diagnostic test to discriminate between subjects who deteriorated, and those who did not, using the clinician's global rating of change as the truth, then we can use ROC curves to study the responsiveness of the instruments.

If we calcluate the sensitivity and specificity of the diagnostic test for various cut-off points for determining whether a patient's control of asthma has deteriorated or not we can plot the ROC curve for each instrument by graphing the pairs of sensitivity and 1–specificity for the various cut–off points.

The most responsive instrument is the one with the largest area under its curve [3]. Once the instrument to be used is chosen, the ROC curve is helpful in determining

the best cut-off point for distinguishing between patients with the condition and those without. 'The point on an ROC curve that is closest to the upper left-hand corner is the 'best' cutoff in terms of making the fewest mistakes when prevalence is at around 50%' [11]. This is applicable as the 'best' cut-off point only if both false positives and false negatives are equally harmful.

The area under an ROC curve can be interpreted as the probability of correctly identifying the patient whose control of asthma deteriorated from randomly selected pairs of patients where one patient's control of their asthma deteriorated and the other's did not [7].

An estimate of the trapezoidal area under the ROC curve can be found by using the Mann–Whitney U [1]:

$$\hat{\text{Area}} = \frac{\text{"Mann} - \text{Whitney U"}}{n_{deteriorated}\, n_{not\ deteriorated}} .$$

Centor [1] gives a Visicalc spreadsheet program which calculates an estimate of the area under an ROC curve and its standard error.

A $(1-\alpha)100\%$ confidence interval for the area under an ROC curve, using the normal distribution [7], is:

$$\left(\hat{\text{Area}} - z_{1-\frac{\alpha}{2}} SE(\hat{\text{Area}}) \ , \ \hat{\text{Area}} + z_{1-\frac{\alpha}{2}} SE(\hat{\text{Area}})\right) .$$

Chapter 5

Methods

5.1 Subjects

From previous experience, the researchers aimed for a sample of 60 subjects, as this was the number of subjects who could usually be found within a reasonable time period and within the budget allocation. (Personal communication with Prof. Elizabeth Juniper, 1997) The subjects were volunteers recruited from previous research studies, notices in the local media, and referrals from asthma clinics in the city of Hamilton. All subjects had to have current asthma symptoms and be between 17 and 70 years of age.

Fifty-one volunteers were found, but one was dropped from the study as their

diagnosis of asthma was found to be suspicious. The study proceeded with 50 subjects.

5.2 Study Design

The intended schedule was for each subject to be interviewed at the clinic four times: at baseline, 1 week, 5 weeks, and 9 weeks. In addition, for the week prior to a visit to the clinic (except for the baseline visit) the subjects were to fill in the daily diary on control of asthma.

At the baseline visit, various demographic information was collected from each subject. This information included age, gender, height, weight, whether or not their asthma was atopic*, and if the subject was under the care of an asthma specialist.

At each of the visits to the clinic, the subjects filled in the control of asthma questionnaire and various other questionnaires: Asthma Quality of Life (QOL) Questionnaire, Additional Asthma Quality of Life Questions, Medical Outcomes Survey Short Form-36 (MOS SF-36), and additional asthma control questions. In addition, at the 1,5, and 9 week visits, the subjects were asked various questions about their asthma in relation to the previous visit.

A clinician also gave a rating of the change in the subjects' control of their

asthma (clinician's global rating of change) at the 1, 5, and 9 week visits relative to the prior visit. This rating was given on a 15 point scale ranging from -7 (a very great deal worse) to 7 (a very great deal better.) This scale was collapsed to a 3 point scale of worse (-7 to -2,) same (-1 to 1,) and better (2 to 7.)

5.3 Data Analysis

The data were analysed using the Statistical Analysis System (SAS.) The ROC curve calculations were done in Quattro Pro for Windows.

The change scores were calculated as the score at the later time less the score at the earlier time. A positive change score on any of the five control of asthma instruments in the study corresponds to a deterioration in the subject's control of asthma, while a negative change score corresponds to an improvement in the subject's control of asthma.

All subjects answered all the questions on both the diary and questionnaire to measure control of asthma. There were missing values on some of the other questionnaires: Asthma Quality of Life Questionnaire (6.8% missing), Additional Asthma Quality of Life Questions (10.7% missing), MOS SF-36 (0.6% missing), and the additional asthma control questions (0.7% missing), but these questionnaires all

have ways to deal with missing values in their analyses.

5.3.1 Reliability

In this study, there were 50 subjects, of whom 36 were judged to have stayed stable (using the clinician's global rating of change) for at least one of the two four-week long intervals. Considering only these subject's stable intervals, and for subjects who were stable for both of the intervals, randomly selecting one, for a balanced design, there were n=36 subjects, each observed at k=2 visits.

5.3.2 Validity

For the validity analysis, all the data will be used for both the cross-sectional and the longitudinal analysis.

5.3.3 Responsiveness

In this study, only two of the three intervals were four weeks long, and for the responsiveness section, we will consider only those two intervals. For the analyses comparing the changers and subjects who stayed stable, the change in scores on the five instruments in study was multiplied by -1 if the subject deteriorated, to adjust for the fact that the change scores of patients who improved will be mostly negative,

and that the change scores of patients who deteriorated will be mostly positive. For the ROC curve analysis, we will randomly select one of the two intervals for each subject.

Chapter 6

Results

Note that we have the following notation for the five instruments in the study:

Diary–Symptoms the five symptom questions on the diary

Diary–All the five symptom questions, the bronchodilator use question and the PEFR question on the diary

Questionnaire–Symptoms the five symptom questions on the questionnaire

Questionnaire–All–#1 the five symptom questions, the bronchodilator use question and the PEFR question on the questionnaire

Questionnaire–All–#2 the five symptom questions, the bronchodilator use question, the PEFR question and the FEV_1 question on the questionnaire

6.1 General Description of the Data

Descriptive statistics on the subjects and their asthma are in Tables 6.1 and 6.2. The means and standard deviations for the scores and the change in scores for the five instruments used in the study are in Tables 6.3 and 6.4 and Figure 6.1.

6.2 Reliability

The intraclass correlation coefficients and the corresponding 95% confidence intervals for the five instruments in the study are in Tables 6.5 and 6.6.

6.3 Validity

The correlation coefficients of the five instruments in the study with other instruments are in Tables 6.7 (cross-sectional) and 6.8 (longitudinal.)

6.4 Responsiveness

The results from the t tests comparing the mean change in changers and stable patients for the five instruments in the study are in Table 6.9. This table also includes the paired t tests to examine if the change in changers is different from 0

and if the change in stable patients is equal to 0.

The Responsiveness Index and the corresponding 95% confidence interval for the five instruments in the study are in Table 6.10 and Figure 6.2.

The Areas under the ROC curve and the corresponding 95% confidence interval for the five instruments in the study are in Table 6.11 and Figure 6.3. The ROC Curves are plotted in Figure 6.4.

6.5 Tables and Figures

Table 6.1: Characteristics of the Subjects (n=50)

Age	Mean (SD)	37.1 (13.1)
	Range	17–70
Gender	Female	64%
	Male	36%

Table 6.2: Characteristics of the Subjects' Asthma

Atopic Asthma	Yes	82%
	No	18%
Under Care of an	Yes	56%
Asthma Specialist	No	44%
Medication	Bronchodilator Only	24%
Use	Bronchodilator and Steroids	74%
	Other	2%
% of times subject's	Deteriorated	24%
control of asthma	Stayed stable	50%
(4-week intervals)	Improved	26%

Table 6.3: Mean (Standard Deviation) of the Scores on the Instruments

Instrument	Week #1	Week #5	Week #9
Diary–Symptoms	0.97544 (0.72401)	0.78424 (0.61049)	0.83360 (0.57329)
Diary–All	1.33103 (0.71392)	1.15731 (0.59908)	1.22400 (0.52354)
Questionnaire–Symptoms	1.23200 (0.75199)	1.19200 (0.77718)	1.12400 (0.65763)
Questionnaire–All–#1	1.46571 (0.70449)	1.40286 (0.69189)	1.37143 (0.59078)
Questionnaire–All–#2	1.61500 (0.70568)	1.54750 (0.69155)	1.51750 (0.62424)

Table 6.4: Mean (Standard Deviation) of the Change in Scores on the Instruments

Instrument	Week #5 - Week #1	Week #9 - Week #5
Diary–Symptoms	-0.19120 (0.62941)	0.04936 (0.53885)
Diary–All	-0.17371 (0.58580)	0.06669 (0.46791)
Questionnaire–Symptoms	-0.04000 (0.80509)	-0.06800 (0.75767)
Questionnaire–All–#1	-0.06286 (0.67078)	-0.03143 (0.64361)
Questionnaire–All–#2	-0.06750 (0.65231)	-0.03000 (0.63264)

Table 6.5: Intraclass Correlation Coefficients for the Instruments

Instrument	ρ_{I_b}	ρ_{I_u}	$\rho_{I(1)_b}$	$\rho_{I(1)_u}$	$\rho_{I(2)_b}$
Diary–Symptoms	0.83650	0.82771	0.83256	0.82331	0.83614
Diary–All	0.85856	0.85086	0.85673	0.84870	0.85842
Questionnaire–Symptoms	0.88244	0.87596	0.87944	0.87259	0.88226
Questionnaire–All–#1	0.92013	0.91563	0.91946	0.91479	0.92010
Questionnaire–All–#2	0.93657	0.93297	0.93484	0.93102	0.93652

Table 6.6: 95% Confidence Intervals for the Intraclass Correlation Coefficients for the Instruments

Instrument	95% CI for ρ_I	95% CI for $\rho_{I(1)}$	95% CI for $\rho_{I(2)}$
Diary–Symptoms	(0.70427 , 0.91281)	(0.69610 , 0.91095)	(0.70182 , 0.91300)
Diary–All	(0.74166 , 0.92499)	(0.73714 , 0.92427)	(0.74085 , 0.92506)
Q'naire–Symptoms	(0.78299 , 0.93803)	(0.77652 , 0.93664)	(0.78125 , 0.93819)
Q'naire–All–#1	(0.85001 , 0.95829)	(0.84795 , 0.95810)	(0.84984 , 0.95832)
Q'naire–All–#2	(0.87999 , 0.96702)	(0.87610 , 0.96623)	(0.87909 , 0.96713)

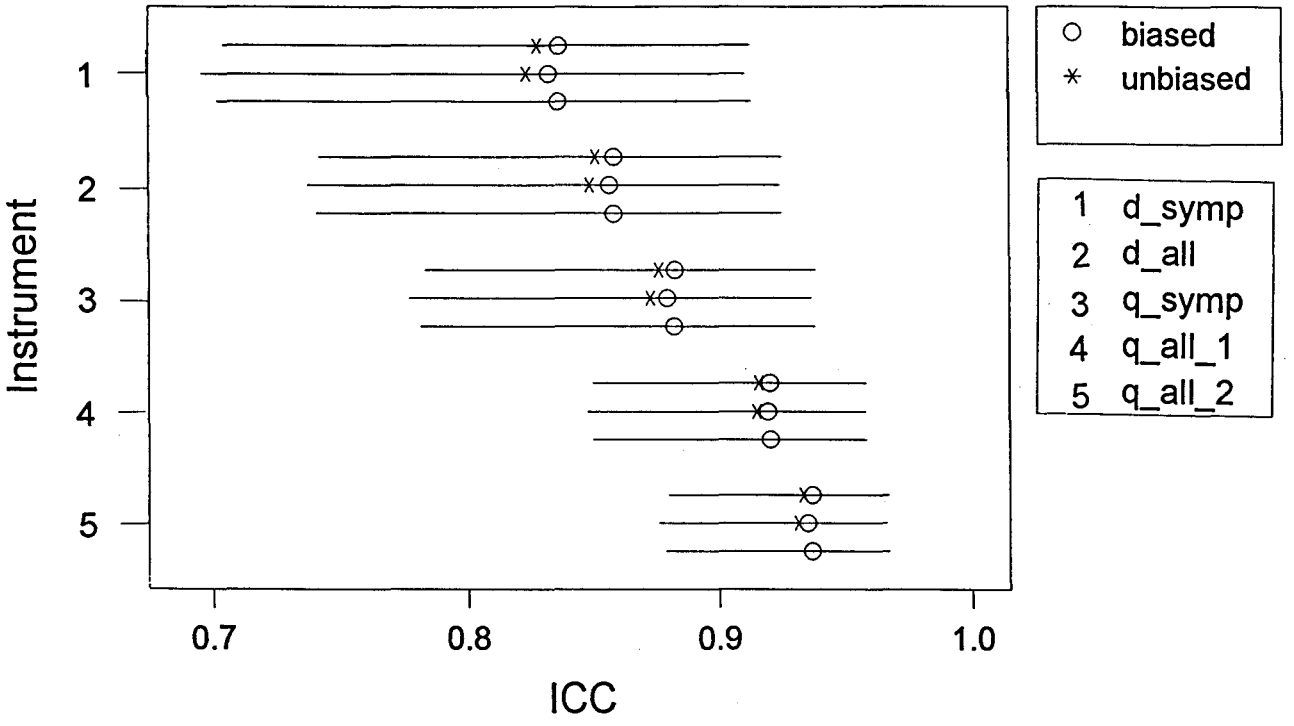

Figure 6.1: 95% Confidence Intervals for the Intraclass Correlation Coefficients for the Instruments

Table 6.7: Correlation Coefficients of the Scores on the Instruments and Other Instruments

	Diary		Questionnaire		
Instrument	Symptom	All	Symptom	All–#1	All–#2
Asthma–QOL–Limits	-0.70788	-0.63661	-0.71720	-0.67204	-0.59764
Asthma–QOL–Symptoms	-0.81991	-0.72708	-0.85925	-0.77620	-0.64835
Asthma–QOL–Emotions	-0.68028	-0.63209	-0.68679	-0.66730	-0.60189
Asthma–QOL–Exposures	-0.66385	-0.54362	-0.58682	-0.51003	-0.42499
Asthma–QOL–Overall	-0.82498	-0.73185	-0.82945	-0.76191	-0.65674
MOS–SF36–Physical	-0.40235	-0.41008	-0.43995	-0.44680	-0.43082
MOS–SF36–Mental	-0.47552	-0.36731	-0.40952	-0.33799	-0.27999
Add'l Asthma Control Q'ns	0.56622	0.46666	0.63223	0.53838	0.42641

Table 6.8: Correlation Coefficients of the Change in Scores on the Instruments and Other Instruments

	Diary		Questionnaire		
Instrument	Symptom	All	Symptom	All–#1	All–#2
Asthma–QOL–Limits	-0.64135	-0.63625	-0.62157	-0.65040	-0.61913
Asthma–QOL–Symptoms	-0.80159	-0.80061	-0.72820	-0.75849	-0.72070
Asthma–QOL–Emotions	-0.62569	-0.60776	-0.58669	-0.61410	-0.57013
Asthma–QOL–Exposures	-0.46702	-0.44499	-0.44839	-0.46908	-0.44534
Asthma–QOL–Overall	-0.77389	-0.76506	-0.73191	-0.76415	-0.72330
MOS–SF36–Physical	-0.06993	-0.11232	-0.11587	-0.16045	-0.17372
MOS–SF36–Mental	-0.31956	-0.30001	-0.17925	-0.18675	-0.16122
Add'l Asthma Control Q'ns	0.18354	0.15221	0.25697	0.26785	0.23477
Global–Control	-0.72289	-0.72031	-0.68094	-0.68281	-0.64455
Global–Quality	-0.65847	-0.65526	-0.55139	-0.56961	-0.54585
Global–Symptoms	-0.70751	-0.69940	-0.65113	-0.66523	-0.63355
Global–Activity	-0.65036	-0.65184	-0.55046	-0.56837	-0.54084
Global–Emotions	-0.28723	-0.26390	-0.33961	-0.32463	-0.32090
Global–Factors	-0.46166	-0.46503	-0.39630	-0.39119	-0.35670
Global–Tired	-0.48491	-0.49723	-0.41764	-0.45061	-0.43261

Table 6.9: Mean (Standard Deviation) of Change and t Tests on the Instruments

Instrument	Change in Changers	Change in Stable Subjects	T value (P–value)
Diary–Symptoms	-0.53504* (0.58723)	0.02680 (0.29771)	6.0342 (< 0.0001)
Diary–All	-0.49931* (0.52045)	0.03343 (0.26651)	6.4425 (< 0.0001)
Questionnaire–Symptoms	-0.88000* (0.59796)	0.00400 (0.28425)	9.4412 (< 0.0001)
Questionnaire–All–#1	-0.71429* (0.54072)	0.02000 (0.23626)	8.7991 (< 0.0001)
Questionnaire–All–#2	-0.65500* (0.58060)	0.00250 (0.23756)	7.4113 (< 0.0001)

An * represents that the paired t test P-value was less than 0.0001.

Table 6.10: Responsiveness Index and 95% Confidence Interval for the Responsiveness Index for the Instruments

Instrument	RI	95% CI for the RI
Diary–Symptoms	0.91113	(0.62687, 1.19538)
Diary–All	0.95939	(0.67513 , 1.24364)
Questionnaire–Symptoms	1.51391	(1.22965 , 1.79817)
Questionnaire–All–#1	1.35791	(1.07366 , 1.64217)
Questionniare–All–#2	1.16967	(0.88541 , 1.45393)

Table 6.11: Area Under the ROC Curve and 95% Confidence Interval for the Area for the Instruments

Instrument	Area Under the ROC Curve	95% CI for the Area
Diary–Symptoms	0.83143	(0.71759 , 0.94527)
Diary–All	0.84000	(0.73225 , 0.93918)
Questionnaire–Symptoms	0.95524	(0.90366 , 1.00000)
Questionnaire–All–#1	0.88265	(0.76250 , 1.00000)
Questionnaire–All–#2	0.84667	(0.70508 , 0.98826)

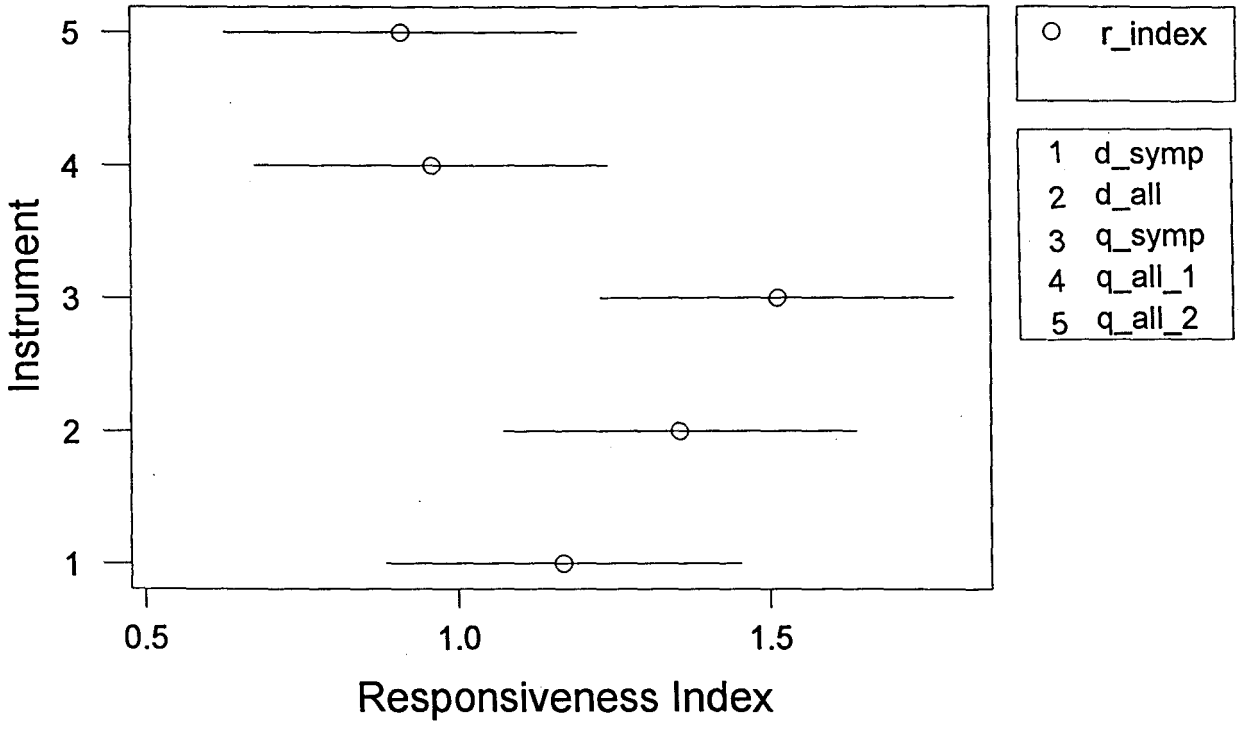

Figure 6.2: 95% Confidence Intervals for the Responsiveness Index for the Instruments

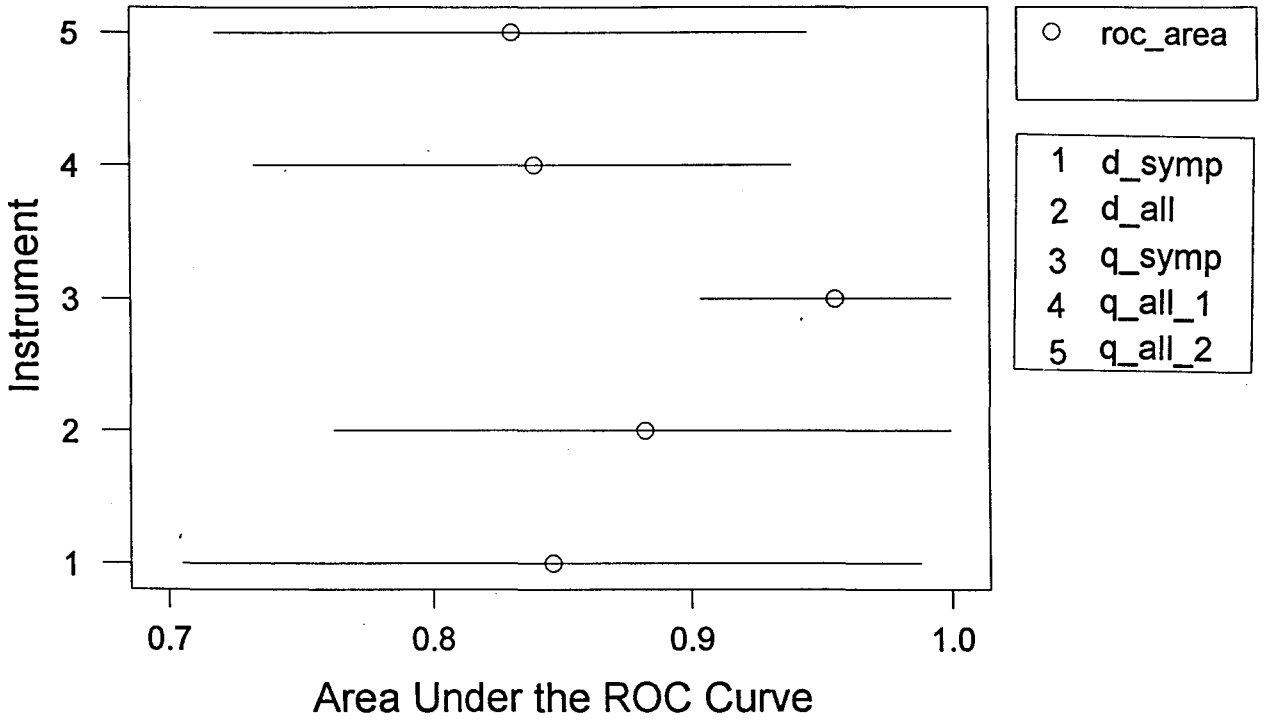

Figure 6.3: 95% Confidence Intervals for the Area Under the Curve for the Instruments

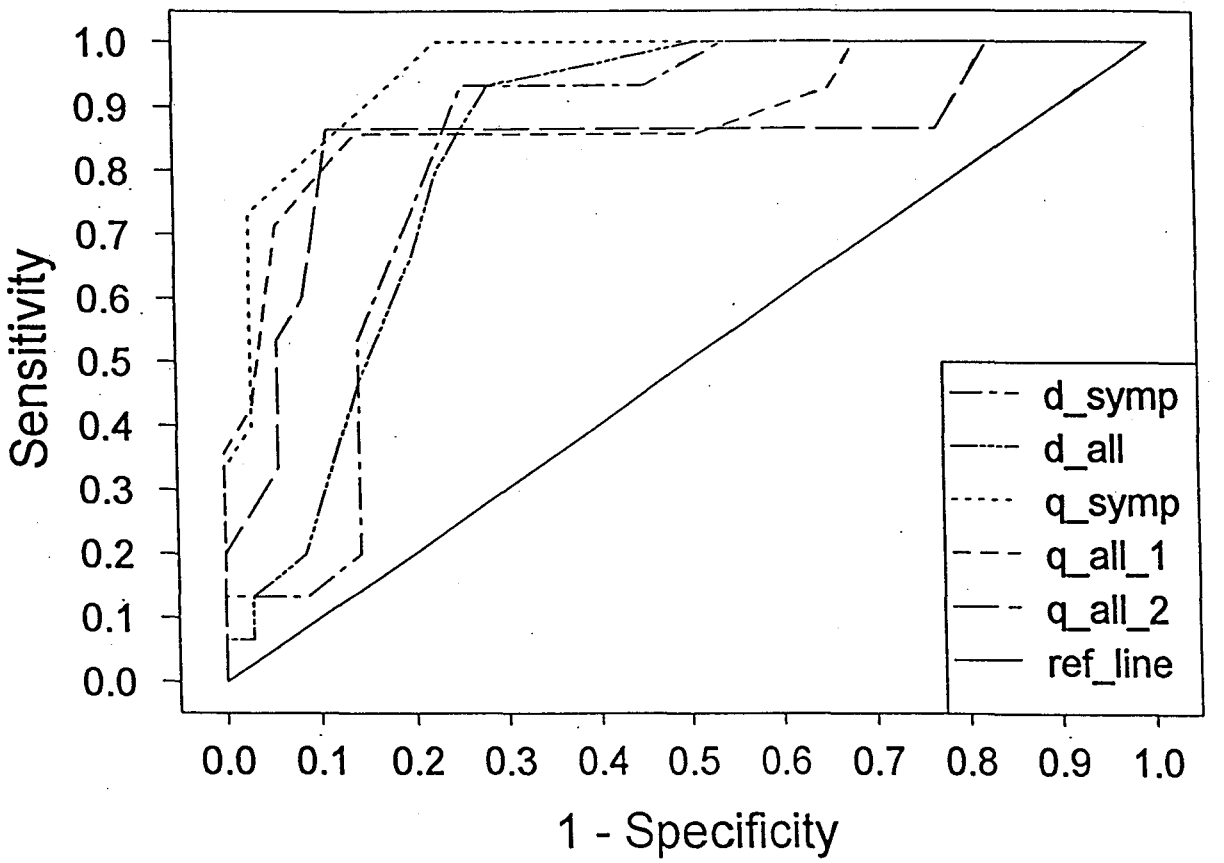

Figure 6.4: ROC Curves for Instruments

Chapter 7

Discussion

7.1 Methods

7.1.1 Subjects and Study Design

Volunteers were used as subjects for this study, and although this is the most common method of obtainning subjects for this type of research, it must be noted that these subjects may not be representative of all asthmatics in either their asthma or their habits, which may affect if they fill in the daily diary as it was intended, and their recall of their asthma symptoms during the previous week.

Some problems with the study design were that the sample size was not chosen by a statistical method, and thus the precision available with a sample of size 50 may

not be sufficient to base any important decisions upon. A further problem is that all subjects filled in both the diary and the questionnaire, and this may have affected the subjects' responses on the questionnaire, as their responses may be memorized from the diary. It would have been better if the subjects had been randomly assigned to either the diary or the questionnaire.

7.1.2 Data Analysis

Various decisions were made as to how much data to include in the analysis. For the reliability section, only the four-week long intervals were considered, and since we are interested in test–retest reliability, only stable subjects were considered. In addition, to preserve the assumption of independent observations in the analysis, if a subject was stable for both of the intervals, one was randomly selected to be used.

For the validity section, all the data was used for both the cross- sectional and longitudinal sections. The asthma clinicians who made the a priori predictions were aware of this and made their predictions correspondingly.

For the responsiveness section, one issue currently being discussed by the researchers is how much of the data to use in the analysis. One possibility is to use all three intervals, the problem with this is that one of the intervals is only a week long, and not of length for which the instrument will be used in practice, while the other

two intervals are four weeks long, so another possibility is to consider only the two four-week long intervals. The third possibility is to randomly chose one of these two intervals per subject to be included in the analysis, this would result in randomness of the observations, which is not a neccessary condition to study responsiveness, and would also result in some of the information collected being lost. In this analysis, the second possibility was used.

For the reliability sections, two models were discussed, the first with one intraclass correlation coefficient, the second with two possible intraclass correlation coefficients. The decision as to which model and, if applicable, which intraclass correlation coefficient is appropriate is a clinical one, requiring knowledge of whether or not the visit should be included as a term in the model, and how the instrument will be used in practice, i.e. with or without a correction factor for the visit.

The biased estimate of an intraclass correlation coefficient is larger than the unbiased estimate although the difference between the two will become smaller as the number of subjects and/or the number of visits increases.

For the responsiveness section, the clinician's global rating of change was used as the truth as to whether a subjects control of asthma deteriorated, stayed stable, or improved in this analysis. This method is not accepted by all, as there is controversy over whether a clinician can make such a judgement based on a few objective

measurements of the patients asthma. Also, this rating of change is not independent of the instruments in the study, as some of the objective measurements are used in both the instruments and for the rating of change, although this does not affect the instruments which included only the five symptom questions.

7.2 Results

The comparisons of the instruments for each of the sections was done heuristically. The results of these comparisons are not necessarily statistically significant unless stated.

7.2.1 Reliability

From Tables 6.5 and 6.6 and Figure 6.1, within each of the five instruments in study, the five intraclass correlation coefficients are equivalent. The confidence intervals within each instrument are also equivalent.

Since all five instruments have intraclass correlation coefficients in the excellent [10] or almost perfect [4] reliability range, there is no real need to compare the instruments. Some general trends may be noted though. The questionnaire based instruments have higher estimates of test–retest reliability than the diary based in-

struments. Within both the diary and questionnaire, the symptoms only alternatives have lower estimates of reliability than the other instruments.

7.2.2 Validity

At this time, the a priori predictions have yet to be made. Although they are usually made before the data is analysed, the asthma clinicians making the predictions have not, and will not see the results until the predictions are made.

Once the a priori predictions have been made, they will be compared against the actual correlation coefficients in Tables 6.7 and 6.8.

7.2.3 Responsiveness

From the paired t test results in Table 6.9, note that the change in changers is significantly different from 0, while the change in stable subjects is not significantly different from 0.

Also from Table 6.9, all five instruments had highly significant results when comparing the change in changers and stable subjects.

Table 6.10 and Figure 6.2 show the responsiveness index and its 95% confidence interval.

From Table 6.11, and Figures 6.3 and 6.4, none of the five instrument's 95%

confidence intervals for the area under their ROC curves included 0.5 (no discrimating ability) so all five instruments are significantly better at discriminating than chance alone.

Some general trends can be noted for all three methods of assessing responsiveness. The questionnaire based methods had better responsiveness than did the diary based methods. The two diary based methods were approximately equal, although the symptoms only alternative always had slightly lower responsiveness. Within the questionnaire based instruments, the symptoms only alternative was always the most responsive, followed by the all #1 alternative, and then by the all #2 alternative.

For the questionnaire based instruments, it seems odd that the more information added to the instrument, the less responsive the instrument became. This may be due to the fact that the two lung function measurement questions were based on a one time reading at the clinic, and perhaps were not representive of the subject's asthma control over the previous week.

7.3 Conclusions

No final conclusions can be made from the analysis so far as the validity sections has yet to be completed. Validity is concerned with making sure that the instrument

is measuring what we believe it to be, and as such, must be shown before any final decisions can be made.

A final decision on which instrument should be used, would be made on the basis of reliability, validity, responsiveness, and other factors, such as cost, simplicity, and plausibility of the instrument.

At this point in the analysis, we can make a recommendation based on the results so far. In the reliability section, we showed that all five instruments have high reliability, and that the questionnaire based instruments had better reliability than the diary based instruments. In the responsiveness section, we showed that the questionnaire based instruments had better responsiveness than the diary based instruments, and that the symptoms only questionnaire based instrument was consistently the most responsive. Since no one instrument stands out as a clear 'best', we can select the questionnaire–symptoms as the instrument of choice as it has excellent reliability, is the most responsive, and it is also the simplest of the five, involving only five symptom questions asked at a clinic.

Returning to the reasoning behind the different instruments in Section 1.2.3, using the results so far, we can attempt to answer the questions. Firstly, it appears that a simple assessment at a clinic is sufficient and that patients need not fill in daily diaries. Secondly, it appears that since questionnaire–symptoms is the instrument of

choice, it could be used in under-developed countries to measure control of asthma. Finally, it appears that the addition of the FEV_1 question is of no benefit.

7.4 Further Study

There are aspects of the data, methods, and concepts which deserve to be further explored.

Using the data, one could estimate the precision available with a sample of 50 subjects. One could also examine the concept of minimal clinically important difference using the subject's global rating of change.

From the method side, an unbiased estimate of $\rho_{I(2)}$ still remains to be found. For responsiveness, it would be interesting to see if other ways, for example, the analysis of variance of change scores method, Norman's S_{repeat} method, and the correlation coefficient method [14], of assessing responsiveness would rank the five instruments in study in the same order.

From the clinical aspect, one should check if the questionnaire based instruments would be as reliable and responsive (and possibly as valid) if the subjects had not filled in a daily diary for the week prior to their visit to the clinic, as part of the high reliability and responsiveness shown might be due to memorized responses. Also,

an analysis on which questions are necessary and which may be excluded should be done before any of the instruments are put into practice.

Appendix A

Glossary

A Priori Prediction predictions made by asthma clinicians before the analysis was undertaken as to what the magnitude of the correlation coefficient between the instruments in the study and instruments measuring other characteristics should be

Acute sudden and short term

Asthma 'a disease characterized by wide variations over short periods of time in resistance to flow in intrapulmonary airways' [2]

Asthma Attack acute narrowing of the air passages

Atopic Asthma asthma due to allergies

Bronchodilator a drug which works by relaxing the contracted bronchial muscles in the patient's air passages

Chronic persistent or recurring

Construct Validity the method of assessing validity when there is no gold standard available

Criterion Validity the method of assessing validity when there is a gold standard available

Cross Sectional Construct Validity construct validity for the data at single points in time

Diagnostic Test a test to classify patients as to whether or not they possess a certain condition

Forced Expired Volume In One Second (FEV_1) a measurement of 'the volume of air that can be forcibly exhaled in one second after a full breath has been taken' [8]

Gold Standard an instrument already in place that is assumed to tell the truth

Instrument any piece of equipment that measures and provides information about something

Longitudinal (Evaluative) Construct Validity construct validity for the change in scores between two points in time

Peak Experiatory Flow Rate (PEFR) a measurement of the speed with which aire can be forcibly blown out of the lungs taken as 'the maximum rate of expiration, maintained for 10 milliseconds, that occurs within the first second of forced expiration' [9]

Prevalence the proportion of patients with a certain condition

Receiver Operating Characteristics (ROC) Curve a graph of the pairs of sensitivity and 1–specificity for a diagnostic test

Reliability measurement property concerned with the instrument measuring something in a reproducible way [15]

Responsiveness measurement property concerned with the instrument being able to detect clinically important changes over time [6]

Sensitivity the proportion of subjects with a certain condition who are properly diagnosed

Specificity the proportion of subjects without a certain condition who are properly diagnosed

Test–Retest Reliability reliability concerned with getting the same scores on two different testing occassions

Validity measurement property concerned with the instrument measuring what was intended [16]

Appendix B

Diary

The diary was to be filled in, by the subject, each morning and bedtime for a week.

Morning Diary

Please do the breathing test and fill in the diary before you take your morning medication.

Write in the number that best describes how your asthma has been during the night and this morning.

Date							
Peak Flow Record the best of 3 blows before you take any medications							
How often were you woken by your asthma during the night? 0) Not woken at all 1) Once 2) Woken a few times 3) Woken several times 4) Woken many times 5) Woken a great many times 6) I was awake all night							
How bad were your asthma symptoms when you woke up this morning? 0) I had no symptoms 1) Very mild symptoms 2) Mild symptoms 3) Moderate symptoms 4) Quite severe symptoms 5) Severe symptoms 6) Very severe symptoms							

Bed Time Diary

Please write in the number that best describes how your asthma has been during the day today.

Date							
How limited were you in your activities today because of your asthma? 0) Not limited at all 1) Very slightly limited 2) Slightly limited 3) Moderately limited 4) Very limited 5) Extremely limited 6) Totally limited							
How much shortness of breath did you experience today? 0) None 1) Very little 2) A little 3) A moderate amount 4) Quite a lot 5) A great deal 6) A very great deal							
How much of the time did you wheeze today? 0) None of the time 1) Hardly any of the time 2) A little of the time 3) A moderate amount of the time 4) A lot of the time 5) Most of the time 6) All the time							
Please record the total number of puffs of bronchodilator (_____) you have used in the past 24 hours							

Appendix C

Questionnaire

The questionnaire was to be filled in, the questions by the subject, and the lung function readings by a researcher, at the visits to the clinic.

CLINICAL ASTHMA CONTROL QUESTIONNAIRE (CLINIC)

Please answer questions 1 - 6.

Circle the number of the response that best describes how you have been during the past week.

1. **On an average night**, during the past week, how often were you **woken by your asthma?**

   0 Never
   1 Hardly ever
   2 A few times
   3 Several times
   4 Many times
   5 A great many times
   6 Unable to sleep because of asthma

2. **On average**, during the past week, how **bad were your asthma symptoms when you woke up** in the morning?

   0 No symptoms
   1 Very mild symptoms
   2 Mild symptoms
   3 Moderate symptoms
   4 Quite severe symptoms
   5 Severe symptoms
   6 Very severe symptoms

3. **In general**, during the past week, how **limited were you in your activities** because of your asthma?

   0 Not limited at all
   1 Very slightly limited
   2 Slightly limited
   3 Moderately limited
   4 Very limited
   5 Extremely limited
   6 Totally limited

4. **In general**, during the past week, how much **shortness of breath** did you experience because of your asthma?

   0 None
   1 Very little
   2 A little
   3 A moderate amount
   4 Quite a lot
   5 A great deal
   6 A very great deal

5. **In general**, during the past week, how much of the time did you **wheeze**?

0 None of the time
1 Hardly any of the time
2 A little of the time
3 A moderate amount of the time
4 A lot of the time
5 Most of the time
6 All the time

6. **On average**, during the past week, how many **puffs of bronchodilator** (.....) have you used each day?

0 None
1 1 - 2 puffs most days
2 3 - 4 puffs most days
3 5 - 8 puffs most days
4 9 - 12 puffs most days
5 13 - 16 puffs most days
6 More than 16 puffs most days

7. FEV1 Pre-bronchodilator: .....

VC Pre-bronchodilator.....

FEV1 Predicted: .....

FEV1 % Predicted: .....

FEV1 Post:.....

VC Post:.....

0 >95 % predicted
1 95 - 90 %
2 89 - 80 %
3 79 - 70 %
4 69 - 60 %
5 59 - 50%
6 < 50 % predicted

8. Peak Expiratory Flow Rate:.....

PEFR Predicted: .....

PEFR % Predicted: .....

0 >95 % predicted
1 95 - 90 %
2 89 - 80 %
3 79 - 70 %
4 69 - 60 %
5 59 - 50%
6 < 50 % predicted

Appendix D

Coding

```
/*   ASTHMA CONTROL CODING   */

data asthma;

infile 'asthma.dat';
input id visit age sex speclist height weight atopic
      pefr_d woken_d morn_d limit_d short_d wheeze_d puff_d
      woken_q morn_q limit_q short_q wheeze_q puffs_q f_pre_q
      v_pre_q f_pred f_perc_q f_post_q v_post_q pefr_q p_pred p_perc_q
      change;

label id = 'subjects id number';
label visit = 'visit number';
label age = 'subjects age';
label sex = 'subjects sex';
label speclist = 'asthma specialist';
label height = 'subjects height';
label weight = 'subjects weight';
label atopic = 'atopic asthma';
```

```
label pefr_d = 'average PEFR reading, diary';
label woken_d = 'average score on woken during night, diary';
label morn_d = 'average score on asthma symptoms in morning, diary';
label limit_d = 'average score on how limited in activities, diary';
label short_d = 'average score on shortness of breath, diary';
label wheeze_d = 'average score on time wheezing, diary';
label puff_d = 'average number of bronchodilator puffs per day, diary';

label woken_q = 'score on woken during night, questionnaire';
label morn_q = 'score on asthma symptoms in the morning, questionnaire';
label limit_q = 'score on how limited in activities, questionnaire';
label short_q = 'score on shortness of breath, questionnaire';
label wheeze_q = 'score on time wheezing, questionnaire';
label puffs_q = 'score on number of bronchodilator puffs, questionnaire';
label f_pre_q = 'FEV1 reading pre medication, questionnaire';
label v_pre_q = 'VC reading pre medication, questionnaire';
label f_pred = 'predicted FEV1 reading';
label f_perc_q = 'FEV1 percent of predicted, questionnaire';
label f_post_q  = 'FEV1 reading post medication, questionnaire';
label v_post_q = 'VX reading post medication, questionnaire';
label pefr_q = 'PEFR reading, questionnaire';
label p_pred = 'predicted PEFR reading';
label p_perc_q = 'PEFR percent of predicted, questionnaire';
label change = 'clinicians global rating of change, -7 to +7';

/* recoding clinician's global rating of change to a 3 point scale */
if -2 >= change >= -7 then rating = -1;
   else if 1 >= change >=-1 then rating = 0;
   else if 7 >= change >= 2 then rating = 1;
label rating = 'clinicians global ratinf of change, -1 to +1';

/* recoding the average number of bronchodilator puffs per day, */
/* diary, to a 7 point scale                                    */
if puff_d = 0 then puffs_d = 0;
   else if 2.5 > puff_d > 0 then puffs_d = 1;
   else if 4.5 > puff_d >= 2.5 then puffs_d = 2;
   else if 8.5 > puff_d >= 4.5 then puffs_d = 3;
```

```
   else if 12.5 > puff_d >= 8.5 then puffs_d = 4;
   else if 16.5 > puff_d >= 12.5 then puffs_d = 5;
   else if puff_d >= 16.5 then puffs_d = 6;
label puffs_d = 'score on number of bronchodilator puffs, diary';

/* converting the average PEFR reading, diary, to a percent */
/* of predicted                                             */
p_perc_d = (pefr_d / p_pred);
label p_perc_d = 'PEFR percent of predicted, diary';

/* recoding the average percent of predicted PEFR reading, diary, */
/* to a 7 point scale                                             */
if p_perc_d > .95 then perc_p_d = 0;
   else if .95 >= p_perc_d >= .90 then perc_p_d = 1;
   else if .90 > p_perc_d >= .80 then perc_p_d = 2;
   else if .80 > p_perc_d >= .70 then perc_p_d = 3;
   else if .70 > p_perc_d >= .60 then perc_p_d = 4;
   else if .60 > p_perc_d >= .50 then perc_p_d = 5;
   else if p_perc_d < .50 then perc_p_d = 6;
label perc_p_d = 'score on PEFR percent of predicted, diary';

/* recoding the percent of predicted PEFR reading, questionnaire, */
/* to a 7 point scale                                             */
if p_perc_q > 95 then perc_p_q = 0;
   else if 95 >= p_perc_q >= 90 then perc_p_q = 1;
   else if 90 > p_perc_q >= 80 then perc_p_q = 2;
   else if 80 > p_perc_q >= 70 then perc_p_q = 3;
   else if 70 > p_perc_q >= 60 then perc_p_q = 4;
   else if 60 > p_perc_q >= 50 then perc_p_q = 5;
   else if p_perc_q < 50 then perc_p_q = 6;
label perc_p_q = 'score on PEFR percent of predicted, questionnaire';

/* recoding the percent of predicted FEV1 reading, */
/* questionnaire, to a 7 point scale               */
if f_perc_q > 95 then perc_f_q = 0;
   else if 95 >= f_perc_q >= 90 then perc_f_q = 1;
   else if 90 > f_perc_q >= 80 then perc_f_q = 2;
```

```
   else if 80 > f_perc_q >= 70 then perc_f_q = 3;
   else if 70 > f_perc_q >= 60 then perc_f_q = 4;
   else if 60 > f_perc_q >= 50 then perc_f_q = 5;
   else if f_perc_q < 50 then perc_f_q = 6;
label perc_f_q = 'score on FEV1 percent of predicted, questionnaire';

/* creating the different instruments */
d_s = mean(of woken_d morn_d limit_d short_d wheeze_d);
d_a = mean(of woken_d morn_d limit_d short_d wheeze_d puffs_d perc_p_d);
q_s = mean(of woken_q morn_q limit_q short_q wheeze_q);
q_a1 = mean(of woken_q morn_q limit_q short_q wheeze_q puffs_q perc_p_q);
q_a2 = mean(of woken_q morn_q limit_q short_q wheeze_q puffs_q perc_p_q
              perc_f_q);
label d_s = 'diary-symptoms';
label d_a = 'diary-all';
label q_s = 'questionnaire-symptoms';
label q_a1 = 'questionnaire-all-#1';
label q_a2 = 'questionnaire-all-#2';

/* setting the diary all instrument to missing for the first visit */
   if visit = 1 then d_all = .;
```

Bibliography

[1] Centor, R.M. (1985) A Visicalc Program for Estimating the Area Under a Receiver Operating Characteristic (ROC) Curve, *Medical Decision Making* **5** 139-148.

[2] Clark, T., and Rees, J. (1985) *Practical Management of Asthma*, Martin Dunitz Ltd, London.

[3] Deyo, R.A., and Centor, R.M. (1986) Assessing the Responseiveness of Functional Scales to Clinical Change: An Analogy to Diagnostic Test Performance, *Journal of Chronic Disease* **39** 897-906.

[4] Eliasziw, M., Young, S.L., Woodbury, M.G., Fryday-Field, K. (1994) Statistical Methodology for the Concurrent Assessment of Interrater and Intrarater Reliability: Using Goniometric Measurements as an Example, *Physical Therapy* **74** 777-788.

[5] Graybill, F.A. (1961) *An Introduction to Linear Statistical Models, Volume 1*, McGraw-Hill Inc, New York.

[6] Guyatt, G.H., Deyo, R.A., Charlson, M., Levine, M.N., and Mitchell, A. (1989) Responsiveness and Validity in Health Status Measurement: A Clarification, *Journal of Clinical Epidemiology* **42** 403-408.

[7] Hanley, J.A., and McNeil, B.J. (1982) The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve, *Radiology* **143** 29-36.

[8] Newhouse, M.T., and Barnes, P.J. (1991) *Conquering Asthma: An Illustrative Guide to Understanding and Care for Adults and Children*, Decker Periodicals Inc, Hamilton, Ontario.

[9] Pearson, R. (1990) *Asthma: Management in primary Care*, Radcliffe Medical Press Ltd, Oxford.

[10] Rosner, B. (1995) *Fundamentals of Biostatistics*, Wadsworth Publishing Company, Belmont, California.

[11] Sackett, D.L., Haynes, R.B., Guyatt, G.H., and Tugwell, P. (1991) *Clinical Epidemiology: A Basic Science for Clinical Medicine*, Little, Brown and Company, Boston.

[12] Scoggin, C.H., and Petty, T.L. (1982) *Clinical Strategies in Adult Asthma*, Lea & Febiger, Philadelphia.

[13] Shrout, P.E., and Fleiss, J.L. (1979) Intraclass Correlation: Uses in Assessing Rater Reliability, *Psychological Bulletin* **86** 420-428.

[14] Stratford, P.W., Binkley, J.M., and Riddle, D.L. (1996) Health Status Measures: Strategies and Analytic Methods for Assessing Change Scores, *Physical Therapy* **76** 1109-1123.

[15] Streiner, D.L., and Norman, G.R. (1989) *Health Measurement Scales: A Practical Guide to their Development and Use*, Oxford Univeristy Press, New York.

[16] Streiner, D.L., and Norman, G.R. (1996) *PDQ Epidemiology*, Decker, Philadelphia.

[17] Tames, P., with Toogood, J.H., and Malo, J-L. (1989) *Every Breath You Take: Coping with Asthma and Other Respiratory Diseases*, Grosvenor House Press Inc, Toronto.

[18] Winer, B.J. (1971) *Statistical Principles in Experimental Design*, McGraw-Hill Inc, New York.