New Zealand Mental Health Consumers and their Outcomes

Tom Trauer
Kathy Eagar
Phillipa Gaines
Alison Bower

Cover painting by Serena Young, Te Ata, Auckland

The Mental Health Research and Development Strategy is a partnership between the
Ministry of Health, Health Research Council of New Zealand,
and Mental Health Commission.

Published in June 2004 by the Health Research Council of New Zealand
PO Box 5541, Wellesley Street, Auckland, New Zealand
Telephone 09 379 8227, Fax 09 377 9988, Email info@hrc.govt.nz

This document is available
on the Health Research Council of New Zealand website: http://www.hrc.govt.nz
and the Mental Health Research and Development Strategy website
http://www.mhrds.govt.nz

ISB Number: 0-908-700-27-X

About this report

This report is the principal report on the analysis of consumer outcomes for mental health consumers in New Zealand. The data used in the compilation of this report were collected during the New Zealand Mental Health Classification and Outcomes Study (Gaines P, Bower A, Buckingham W, Eagar K, Burgess P. & Green J. (2003) *New Zealand Mental Health Classification and Outcomes Study: Final Report.* Health Research Council of New Zealand: Auckland) and includes information from the eight District Health Boards that participated in the Study.

This report presents the analysis of the outcomes data and is a compendium to the final report for New Zealand Mental Health Classification and Outcomes Study (NZ-CAOS). It provides only a brief overview of the key features of the NZ-CAOS study that are pertinent to outcome measurement. Readers seeking further detail about all aspects of the original study, including the selection of outcome measures, staff training and the data collection protocol are referred to the final report for NZ-CAOS at http://www.mhrds.govt.nz/dynz.aspx?ID1=3&ID2=63&ID3=104.

The purpose of this report is to provide examples of the different ways the clinical outcome data can be analysed. It does not aim to measure the performance of the services that participated in the study. Rather, the aim is to provide an opportunity for New Zealand mental health service stakeholders to consider what outcome measures may be useful for various purposes and in different situations. It does not interpret the findings in any detail as additional reports will be commissioned by the Mental Health Research and Development Strategy Steering Committee and that group will provide further discussion around the interpretation of the data from the perspective of the different stakeholder groups.

Acknowledgements

The Māori Monitoring and Review Group have provided us with expert guidance throughout our analysis and our interpretation of the findings, as has Professor Graham Mellsop, Chair of the National CAOS Reference Group. We also acknowledge the Victorian Department of Human Services in Australia for allowing us to use unpublished data from Victoria for comparative purposes. Finally, this study would not have been possible without the many consumers and clinicians who participated in the NZ-CAOS study.

Suggested citation

Trauer T, Eagar K, Gaines P and Bower A (2004) *New Zealand Mental Health Consumers and their Outcomes*. Health Research Council of New Zealand: Auckland

Table of Contents

1. Introduction 1

1.1 Project Overview 1

1.2 Project objective 2

1.3 What is a consumer outcome? 2

1.3.1 Implications for the analysis 3

2. Method 5

2.1 Overview of the study 5

2.1.1 Project Scope 5

2.1.2 Participating sites 6

2.1.3 Study sample 6

2.1.4 Service profile 7

2.1.5 The study episode model 7

2.1.6 Data collection triggered by 'episode of care events' 8

2.1.7 Overview of core consumer data set 10

2.2 Data management and preparation 11

2.2.1 Security of data 11

2.2.2 Preparation of the data for the outcomes analysis 12

2.2.3 Summary of compliance rates for each of the instruments 15

3. Results: completion rates of the individual items 19

3.1 Adult measures 19

3.1.1 HoNOS 19

3.1.2 LSP-16 21

3.2 Child and Youth measures 22

3.2.1 HoNOSCA 22

3.2.2 FIHS 24

3.3 Summary of results in relation to item completion 25

4. Results: technical performance of the instruments 27

4.1 HoNOS 27

4.2 LSP-16 29

4.3 HoNOSCA 30

4.4 FIHS 30

4.5 Summary of the technical performance of the instruments 30

5. Results: relationship between the outcome measures 33

5.1 Adults 33

5.1.1 Focus of Care and HoNOS 33

5.1.2 Focus of Care and LSP-16 35

5.1.3 HoNOS and LSP-16 35

5.2 Children and Youth 36

5.2.1 HoNOSCA and CGAS 36

5.2.2 HoNOSCA and FIHS 37

5.2.3 CGAS and FIHS 38

5.3 Summary of the relationship between the outcome measures 39

6. Results: relationship between the outcome measures and other clinical and socio-demographic variables 41
6.1 Adult measures 41
6.1.1 HoNOS 41
Age 41
Sex 42
Diagnosis 44
Legal status 45
6.1.2 LSP-16 47
Age 47
Sex 48
Diagnosis 50
Legal status 50
6.2 Child and Youth 51
6.2.1 HoNOSCA 51
Age 51
Sex 52
Diagnosis 53
6.2.2 CGAS 54
Age 54
Sex 54
Diagnosis 54
6.2.3 FIHS 55
Age 55
Sex 56
Diagnosis 56
6.3 Summary of the relationship between the outcome measures and other clinical and socio-demographic variables 57
7. Results: do consumers change over the course of an episode of mental health care? 59
7.1 Adults 59
7.1.1 The HONOS 59
Inpatient episodes 59
Community episodes 61
7.1.2 LSP-16 63
7.1.3 Adult change scores by Focus of Care 64
7.2 Child and Youth 66
7.2.1 HoNOSCA 66
7.2.2 CGAS 66
7.2.3 FIHS 66
7.3 Summary of consumer change over the course of an episode of mental health care 66
8. Results: comparison between outcomes for Māori, Pacific and Other Consumers 69
8.1 Adults 69
8.1.1 HoNOS profile 69
8.1.2 Change as measured by the HoNOS 73
Inpatient episodes 73
Community episodes 74
8.1.3 LSP-16 profile 75
8.1.4 Change as measured by the LSP-16 76

8.2 Child and Youth 77
8.2.1 HoNOSCA profile 77
8.2.2 Change as measured by the HoNOSCA 78
8.2.3 CGAS profile 79
8.2.4 Change as measured by the CGAS 79
8.2.5 FIHS profile 80
8.3 Summary of comparison between outcomes for Māori, Pacific and Other Consumers 80
Adult consumers rated on the HoNOS 80
Adult consumers rated on the LSP-16 80
Child/Youth consumers rated on the HoNOSCA 81
Child/Youth consumers rated on the CGAS 81
Child/Youth consumers rated on FIHS 81
9. Results: the relationship between ethnicity, deprivation and outcome profiles 83
9.1 Summary of the relationship between ethnicity, deprivation and outcome profiles 85
10. Results: outcomes for each casemix class 87
10.1 Change as measured by the HoNOS 87
10.1.1 Casemix-adjusted HoNOS outcomes by DHB 92
10.2 Change as measured by the LSP-16 93
10.2.1 Casemix-adjusted LSP-16 outcomes by DHB 95
10.3 Change as measured by the HoNOSCA 96
10.3.1 Casemix-adjusted HoNOSCA outcomes by DHB 99
10.4 Change as measured by the CGAS 99
10.4.1 Casemix-adjusted CGAS outcomes by DHB 102
10.5 Summary of the relationship between outcomes and casemix 102
11. Discussion of key findings and the implications for policy and practice 107
11.1 Compliance, completion and technical performance of the outcome measures 107
11.1.1 Implications for policy and practice 108
11.2 Consumer outcomes 108
11.2.1 Implications for policy and practice 109
11.3 Outcomes, ethnicity and social deprivation 109
11.3.1 Implications for policy and practice 110
12. References 111
13. Appendix Description of the consumer outcome measures 113
13.1 Health of the Nation Outcome Scales (HoNOS) 113
13.2 Abbreviated Life Skills Profile (LSP-16) 114
13.3 Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) 115
13.4 Children's Global Assessment Scale (CGAS) 116
13.5 Factors Influencing Health Status (FIHS) 116

List of Tables

Table 1. Broad indicators of the study for the period 2001/02. 6
Table 2. Services included in the study by team type. 7
Table 3. Overview of core consumer data set. 10
Table 4. Clinical rating instruments for measuring severity and level of functioning. 11
Table 5. Episode of Care by Treatment Setting and Sector. 12
Table 6. Number of ratings removed because of invalid dates. 13
Table 7. Number of ratings removed because they were duplicates. 13
Table 8. Number of ratings removed because collection date did not correspond with episode dates. 14
Table 9. Number of useable ratings after removal of ratings with all item values missing. 14
Table 10. Number of paired ratings. 15
Table 11. Summary of compliance rates on the various measures. 16
Table 12. Overall compliance rates. 17
Table 13. Percentage of missing HoNOS ratings, by setting and collection occasion. 19
Table 14. Percentage of HoNOS assessments with different numbers of missing items, by setting and collection occasion. 20
Table 15. Percentage of HoNOS assessments with no missing items, by setting, collection occasion and DHB. 21
Table 16. Percentage of LSP-16 assessments with different numbers of missing items, by setting and collection occasion. 21
Table 17. Percentage of LSP-16 assessments with no missing items, by setting, collection occasion and DHB. 22
Table 18. Percentage of missing HoNOSCA ratings in the community, by collection occasion. 22
Table 19. Percentage of HoNOSCA assessments conducted in the community with different numbers of missing items, by collection occasion. 23
Table 20. Percentage of HoNOSCA assessments conducted in the community with no missing items, by collection occasion and DHB. 23
Table 21. Percentage of missing FIHS ratings in the community, by collection occasion. 24
Table 22. Percentage of FIHS assessments conducted in the community with different numbers of missing items, by collection occasion. 24
Table 23. Percentage of FIHS assessments conducted in the community with no missing items, by collection occasion and DHB. 25
Table 24. Percentage of assessments conducted with no missing items, by instrument. 25
Table 25. Fit indices following CFA of HoNOS four subscale structure, by data source. 27
Table 26. Fit indices following CFA of HoNOS five subscale structure, by data source. 28
Table 27. Cronbach's alpha values of subscales in four- and five-subscale models. 29
Table 28. Fit indices following CFA of LSP-16 four subscale structure. 29
Table 29. Fit indices following CFA of HoNOSCA four subscale structure. 30
Table 30. Inter-correlations of HoNOS and LSP-16 subscale and total scores. 36
Table 31. Correlations between CGAS and HoNOSCA items, subscales and total score. 37
Table 32. Mean CGAS scores associated with FIHS items, and correlations with age. 38
Table 33. Correlations between age and HoNOS items, subscales and total score. 42
Table 34. Mean HoNOS item scores by sex. 43
Table 35. Mean HoNOS total score by setting and diagnosis. 44
Table 36. Correlations between age and LSP-16 items, subscales and total score. 48

Table 37. Mean LSP-16 items scores by sex. 49
Table 38. Mean LSP-16 total score by setting and diagnosis. 50
Table 39. Correlations between age and HoNOSCA items, subscales and total score..... 52
Table 40. Mean HoNOSCA items, subscales and total scores by sex. 53
Table 41. Mean HoNOSCA total score by setting and diagnosis. 54
Table 42. Mean CGAS total score by setting and diagnosis..... 55
Table 43. Mean ages associated with FIHS items, and correlations with age. 55
Table 44. Percentages of present ratings of FIHS items by sex..... 56
Table 45. Mean FIHS total score by setting and diagnosis..... 57
Table 46. HoNOS total scores at start and end of inpatient episodes by ethnicity grouping. 74
Table 47. HoNOS total scores at start and end of community episodes by ethnicity grouping. 74
Table 48. LSP-16 total scores at start and end of community episodes by ethnicity grouping. 76
Table 49. HoNOSCA total scores at start and end of community episodes by ethnicity groupings. 78
Table 50. CGAS total scores at start and end of community episodes by two ethnicity groupings. 80
Table 51. Number of observations in the HoNOS analysis. 83
Table 52. Number of observations in the LSP-16 analysis. 84
Table 53. Average improvement in HoNOS scores by casemix class. 89
Table 54. Average improvement in LSP-16 scores by casemix class. 94
Table 55. Average improvement in HoNOSCA scores by casemix class. 97
Table 56. Average improvement in CGAS scores by casemix class. 100
Table 57. Summary of percentage change from episode start to episode end with each measure. 108

List of Figures

Figure 1. Episode of Care model. 8
Figure 2. Examples of the consumer outcome data collection points. 9
Figure 3. Five subscale model of HoNOS. 28
Figure 4. HoNOS scale scores at episode start by Focus of Care – inpatient settings. 34
Figure 5. HoNOS scale scores at episode start by Focus of Care – community settings. 34
Figure 6. LSP-16 scale scores at episode start by Focus of Care – community settings. 35
Figure 7. HoNOSCA subscale and total scores by FIHS total score. 38
Figure 8. Mean HoNOS subscales scores by sex. 43
Figure 9. Mean HoNOS alternative subscales scores by sex. 44
Figure 10. HoNOS conventional summary scores by legal status – inpatient settings. 45
Figure 11. HoNOS alternative summary scores by legal status – inpatient settings. 46
Figure 12. HoNOS conventional summary scores by legal status – community settings. 46
Figure 13. HoNOS alternative summary scores by legal status – community settings. 47
Figure 14 . LSP-16 subscale and total scores by sex. 49
Figure 15. LSP-16 summary scores by legal status – community settings. 51
Figure 16. Mean admission and discharge HoNOS scores for episodes of schizophrenia – inpatient settings. 60
Figure 17. Mean admission and discharge HoNOS scores for episodes of mood disorder – inpatient settings. 61
Figure 18. Mean admission and discharge HoNOS scores for episodes of schizophrenia – community settings. 62
Figure 19. Mean admission and discharge HoNOS scores for episodes of mood disorder – community settings. 63
Figure 20. Mean admission and discharge LSP-16 subscale scores for episodes of schizophrenia and mood disorder – community settings. 64
Figure 21. Mean HoNOS total score, by Focus of Care and episode type. 64
Figure 22. Mean LSP-16 total score, by Focus of Care category – community episodes only. 65
Figure 23. Mean HoNOS item scores in acute inpatient settings, by ethnicity grouping. 70
Figure 24. Mean HoNOS subscales scores in acute inpatient settings, by ethnicity grouping. 70
Figure 25. Mean HoNOS alternative subscales scores in acute inpatient settings, by ethnicity grouping 71
Figure 26. Mean HoNOS item scores in community settings, by ethnicity grouping. 72
Figure 27. Mean HoNOS conventional subscale scores in community settings, by ethnicity grouping. 72
Figure 28. Mean HoNOS alternative subscales scores in community settings, by ethnicity grouping. 73
Figure 29. Changes in HoNOS total scores between start and end of inpatient and community episodes, by ethnicity grouping. 75
Figure 30. LSP-16 subscales and total score by ethnicity grouping. 76
Figure 31. Mean HoNOSCA item scores in community settings, by ethnicity grouping. 77
Figure 32. Mean HoNOSCA subscale scores in community settings, by ethnicity grouping. 78
Figure 33. Mean CGAS scores in community settings, by ethnicity grouping. 79
Figure 34. HoNOS total scores by deprivation quintile and ethnicity grouping. 84
Figure 35. LSP-16 total scores by deprivation quintile and ethnicity grouping. 85
Figure 36. Average HoNOS improvement by DHB. 87

Figure 37. Improvement on the HoNOS by inpatient class. 90
Figure 38. Improvement on the HoNOS by community class. 91
Figure 39. Summary of change profile on the HoNOS for each adult inpatient class. 91
Figure 40. Summary of change profile on the HoNOS for each adult community class. 92
Figure 41. Improvement on the HoNOS by DHB. 93
Figure 42. Improvement on the LSP-16 by community class. 94
Figure 43. Summary of change profile on the LSP-16 for each adult community class. 95
Figure 44. Improvement on the LSP-16 by DHB. 96
Figure 45. Improvement on the HoNOSCA by community class. 98
Figure 46. Summary of change profile on the HoNOSCA for each community class. 98
Figure 47. Improvement on the HoNOSCA by DHB. 99
Figure 48. Improvement on the CGAS by casemix class. 101
Figure 49. Summary of change profile on the CGAS for each casemix class. 101
Figure 50. Improvement on the CGAS by DHB. 102
Figure 51. Change profile of the adult inpatient classes. 103
Figure 52. Change profile of the adult community classes. 104
Figure 53. Change profile of the child and youth classes. 105

1. Introduction

This chapter provides an overview of mental health outcomes within the context of the Mental Health Research and Development Strategy (MHR&DS). It defines the objective of the study and introduces the concept of measuring consumer outcomes. The final section of this chapter identifies the implications of this discussion for the analysis presented in the remainder of the report.

1.1 Project Overview

This report on the outcomes component of the New Zealand Mental Health Classification and Outcomes Study (NZ-CAOS) encompasses one of the four priority areas defined under the Mental Health Research and Development Strategy (MHR&DS). The Strategy therefore provides the broad context to understand the origins and aspirations of the project.

The four priority areas of the Strategy are:

1. Epidemiology – which aims to measure the incidence and prevalence of different mental health problems in New Zealand populations.
2. Outcomes – which aims to develop and assess measures of mental health outcome (the subject of this report).
3. Casemix – which aims to develop and assess a casemix classification system to inform planning, purchasing and delivery of mental health services in New Zealand.
4. Quality and best practice.

The aim of the Research and Development Strategy is to foster research and development that will assist in the planning and improved delivery of services for those most in need (i.e. the 3% of people with serious mental illness). The Strategy has progressively developed in light of the Governments national mental health strategy, which started in 1994 with *Looking Forward: Strategic Directions for Mental Health Services*. It was further developed in the Ministry of Health's *Moving Forward: The National Plan for More and Better Mental Health Services* published in 1997. This was refined in the 1997 Mental Health Commission *Blueprint for Mental Health Services in New Zealand*, which spelt out the essential components and resource guidelines for mental health services. Funding to support the Mental Health Research and Development Strategy amounts to approximately $1.6 million per year and is administered by the Health Research Council on behalf of the Ministry of Health.

The overall objective of the MHR&DS is to use research and development to identify ways that will improve the planning, purchasing and delivery of mental health services in New Zealand, and which are consistent with the Treaty of Waitangi and the needs of consumers, family, whanau and other stakeholders. Specifically it aims to:

- Create a research and development culture within the mental health sector;
- Facilitate networking between researchers, providers and purchasers;
- Better utilise current research and development capacity;
- Build research and development capacity;
- Encourage evidence-based practice, and
- Collaborate with other initiatives in the mental health area.

This project was principally concerned with the strategy's casemix development objective. However, it was also designed to provide significant experience in the routine use of instruments to assess consumer outcomes. The casemix work has already been reported[1]. This compendium report focuses on consumer outcomes.

1.2 Project objective

The Mental Health Classification and Outcomes Study was a study with two objectives. One of those objectives is the subject of this report.

Objective

To trial the introduction of outcome measurement into routine clinical practice.

The National Mental Health Strategy advocates regular outcome measurement within clinical services. The commitment to incorporate outcome measurement into routine clinical practice in New Zealand is reflected in the Crown Funding Agreements for 2001/02 that specify the District Health Boards will develop a process for measuring outcomes. The Mental Health Research and Development Strategy also funds several projects that may assist in the measurement of outcomes in mental health in the medium- and long-term. Because the measures collected in this study included several standard outcome instruments, it provided an opportunity to introduce a significant number of clinical staff to routine outcome scales and to collect valuable data on consumer outcomes that could be used to inform future national policy directions.

The main aim of the analysis reported here is to contribute to service effectiveness (consumer gain) and efficiency (provider gain) by measuring the outcomes that various groups of consumers achieve. This analysis is designed to contribute towards the development of routine collection of outcome measurement in New Zealand (the MH-SMART initiative).

1.3 What is a consumer outcome?

The Australian Health Ministers Advisory Council (AHMAC), of which New Zealand is a member, defined a health outcome "a change in the health of an individual or group of individuals which is attributable to an intervention or a series of interventions". At first glance, this definition looks simple. But it contains two terms that are critical.

The first is the term "attributable". Its meaning is obvious. The change must be due (able to be attributed) to the intervention of the clinician or the health service. Thus when considering health outcomes, it is important to distinguish between health status (which is affected by social, demographic, and economic factors) and health care outcomes (which are affected specifically by health care interventions).

'Health status' is not the same as 'health outcomes' and 'health status' cannot be used as a proxy for 'health outcomes'. The health status of both an individual and a population can be attributed to a wide and complex interaction of factors, only one of which is the availability and usage of health services. In fact, many people would argue that the role of the health system is minor compared with factors such as demographic and socioeconomic status, the state of the economy, employment opportunities and so on.

---

[1] Gaines P, Bower A, Buckingham W, Eagar K, Burgess P. & Green J. (2003) *New Zealand Mental Health Classification and Outcomes Study: Final Report.* Health Research Council of New Zealand: Auckland.

The second key word is ‘change’. At the individual consumer level, "health outcome" has two distinct meanings[2]. In one, the health outcome is the difference in health status ‘before and after’ the intervention. This is the way that clinicians (and consumers) would typically judge the success of most interventions – after treatment, the consumer felt better (or worse) than they did before.

The problem is that this approach does not take into account what may have happened if the consumer had received either no treatment or another type of treatment. This alternative approach has been described as the ‘with and without’ model.

The ‘with and without’ approach is of particular relevance to chronic diseases where the goal of intervention may be to maintain current health status. Under the ‘before and after’ approach maintaining current health status would not necessarily rate as a good health outcome. Under the ‘with and without’ approach, maintaining current health status would be rated as a good health outcome if the alternative were a possible decline.

For example, a person with prolonged schizophrenia receives services from a specialist mental health service over a period of six months. At the end of the six months, she has the same functional status as she had at the beginning of the episode. Is this a good health outcome? Both she, and her clinician, rate this as a good outcome.

This example raises another important issue. The definition of a health outcome is dependent on the goal of intervention. For example, the consumer and her clinicians remain optimistic about her long-term outcomes and she is receiving a rehabilitation-focussed service directed towards her long-term recovery. But her current treatment goal is not ‘improvement’ at this particular stage. That is planned for the next stage. Rather her current goal is maintenance of her current quality of life, both for herself and her family.

This particular consumer’s outcomes can be best understood by taking into account the focus or goal of her care during the period in which her outcomes are rated. For this particular consumer at this point in time, no change is regarded as a good outcome.

1.3.1 Implications for the analysis

The analysis presented in this report is based mainly on the ‘before and after’ approach. While some analysis is included that takes account of the focus of the care provided, the data collected in the study do not include information to measure outcomes using a ‘with and without’ approach.

A further issue of significance to the outcomes reported in this study is that they are based solely on clinician ratings. In this regard, the study is no different to much of routine practice in which the assessment of outcomes is undertaken by health professionals as the ‘experts’. It is important to recognise, however, that this approach has certain inherent limitations.

First, it is only in relatively recent times that the importance of the consumer’s view and the family/carers view has been appreciated and, in consequence, there are few consumer-rated and carer-rated measures (as yet) in routine practice. While clinicians may rely upon consumers’ reports (e.g. about their mood) in order to form their clinical judgements, there is no provision

[2] See, for example, Eagar K (2000) *An overview of health outcomes measurement in Australia*. Mental Health Outcomes Research in Aotearoa: Proceedings of the September 2000 Mental Health Research and Development Strategy Outcomes Conference and Other Developments in Mental Health Outcomes Research. p 142-148.

for obtaining routine, systematic and comparable information from consumers or their families/carers. At worst, when the consumer or the carer has a different view from the clinician about the consumer's state of health, the consumer or carer may simply be assumed to be 'wrong' and the clinician 'right'.

The second critical issue is the cultural sensitivity and competency of the clinical workforce. Both Māori and Pacific representatives raised this issue during the planning and analysis of the NZ-CAOS study, and expressed significant concern that the interpretation of both clinical ratings and diagnosis may be affected by misdiagnosis and the lack of cultural competency by mainstream clinical staff. Their concerns, and the degree to which they were addressed, are discussed in the compendium casemix report.

Thus, this study is limited. It is an analysis of consumer outcomes as rated by clinicians who inevitably vary in their skills, experience and cultural competency. It is not an analysis that considers the consumer or family perspective on the direction or amount of change, on whether any change can be attributed to the mental health service or whether the outcomes would have been the same 'with or without' the intervention. These are important caveats in interpreting the findings presented in this report.

2. Method

This chapter is in two parts. The first part provides a short overview of the NZ-CAOS study. The NZ-CAOS study is the source of the data reported in this analysis. The second provides an overview of the analysis. This includes a section on how the data were prepared for the analysis and a consideration of data completeness.

2.1 Overview of the study

This section provides a short summary of key features of the NZ-CAOS study. It is designed simply to put the subsequent analysis into context rather than provide the reader with a detailed understanding of the full study. Full details of the study can be found in the study manual[3] developed for the study and in the compendium casemix report[4]. The NZ-CAOS study was itself based on a large Australian study called the MH-CASC study, and drew on many of the methods first developed in that study. Full details of the MH-CASC study can be found in the study report[5].

2.1.1 Project Scope

The scope included specialist mental health services provided directly by District Health Boards (DHBs), which are funded by the Mental Health Directorate of the Ministry of Health. This covers:

- Child and youth mental health services;
- Adult mental health inpatient and community care services, and
- Forensic mental health services.

Mental health services for the elderly did not easily fit within scope due to the variable funding arrangements that are shared between the Mental Health and the Disability Support Service Directorates. As a compromise, the study included those services managed by the participating DHB sites and which were funded by the Mental Health Directorate only.

The following services were excluded from the study:

1. *Services provided by non-government organisations (NGOs)* – a key consideration was whether or not NGO services were largely substitutable for services that would otherwise be provided by a DHB. Given that NGO services are predominantly orientated towards meeting the support needs of consumers rather than their clinical needs a decision was made to exclude them from the study. An additional consideration was that the cost and complexity of the data collection would have significantly increased the overall costs of the study and would more than likely have proven prohibitive for most NGOs.

---

[3] Gaines P, Bower A & Buckingham W (2001) *Mental Health Classification and Outcomes Study: Study Resource Manual*. Health Research Council of New Zealand. Auckland, New Zealand.

[4] Gaines P, Bower A, Buckingham W, Eagar K, Burgess P. & Green J. (2003*) New Zealand Mental Health Classification and Outcomes Study: Final Report.* Health Research Council of New Zealand: Auckland. Available at http://www.mhrds.govt.nz/dynz.aspx?ID1=3&ID2=63&ID3=104

[5] Buckingham B, Burgess P, Solomon S, Pirkis J and Eagar K (1998) *Developing a Casemix Classification for Mental Health Services.* Commonwealth Department of Health and Family Services, Canberra. Available at http://www.mentalhealth.gov.au/mhinfo/standards/casemix and also at http://www.mhri.edu.au/mh-casc/

2. *Alcohol and Drug services* – as A&D services are not included within the management of Australian mental health services they were not included in the design of the equivalent Australian casemix study, Mental Health Classification and Service Costs Project (MH-CASC). Consequently there was no foundation upon which to build a casemix classification in New Zealand that also covered primary alcohol and drug disorders. However, consumers with a diagnosis of substance abuse were included in the study, but only if they were treated within the specialist mental health sector.
3. *Residential services provided by DHBs* – a few of the sites provide residential accommodation services. A decision was made to treat these services the same as those residential services provided by NGOs and they were therefore excluded from the study.

2.1.2 Participating sites

Eight DHBs participated in the study. In some cases DHBs participated on a partial basis only because they considered that clinical staff in some parts of their organisation were already dealing with the pressure of significant workloads and could not sustain the collection of additional CAOS related data over the six-month period of the study. The eight DHB sites that participated in the study are as follows:

- Northland DHB;
- Waitemata DHB (regional forensic services only);
- Auckland DHB;
- Counties-Manukau DHB (excluding adult community health services);
- Waikato DHB (excluding adult community services);
- Lakes DHB;
- Capital and Coast DHB, and
- Otago DHB.

Criteria used in selecting the study sites are set out in the compendium casemix report.

2.1.3 Study sample

The study sample included services that are representative of New Zealand. Overall the eight DHBs covered approximately 22% of New Zealand mental health services. The broad indicators of the study are summarised in Table 1.

Table 1. Broad indicators of the study for the period 2001/02.

INDICATOR	The 8 DHBs	National total	% of total
CATCHMENT POPULATION	2.1 million	3.9 million	55.2%
DHB RESOURCES (in millions)			
Expenditure – Total (incl A&D)	$163.1	$496.5	32.9%
Expenditure – In Scope services	$103.2	$460.1	22.4%
ADULT DHB SERVICES*			
Expenditure – total in-scope	$91.9	$396.7	23.2%
Inpatient beds	599	1158	54.6%
Community FTEs	631	2690	34.2%

INDICATOR	The 8 DHBs	National total	% of total
CHILD & YOUTH DHB SERVICES*			
Expenditure – total in-scope	$11.4	$63.7	17.9%
Inpatient beds	27	60	45.0%
Community FTEs	169	582	29.0%

* Excludes A&D and all NGO services.

Source: Gaines et al. (2003)

In summary, the study sample included 22% of in-scope mental health resources. About half of all in-scope inpatient beds were included in the sample and about one third of all community mental health positions. The size and representativeness of the study suggests that the results can be generalised to New Zealand mental health services as a whole.

2.1.4 Service profile

In total, 103 teams and units participated in the study. A breakdown by type is shown in Table 2. While it may appear at first glance that forensic services are over-represented in the study, the number of consumer episodes collected by forensic teams was small compared to the overall sample.

Table 2. Services included in the study by team type.

Team Type	Number of teams/wards
Adult	58
Child and youth (excluding youth specialty and early intervention which came under 'adult')	11
Kaupapa Māori	7
Pacific Island	2
Services for the elderly	8
Forensics	17
Total	**103**

2.1.5 The study episode model

The 'episode of care' model that underpinned the design of the study, and subsequently the results reported here, is shown in Figure 1 below. This episode concept was fundamental to the study as it defined the unit of counting used to measure both costs and outcomes. The concept of *'episode of care'* is used widely throughout the health system, where 'episode' refers to a period of care with a discrete start and end point. Significant problems arise when translating this concept to mental health services because no concept of episode has been agreed to quantify community services and many consumers undergo care over extended periods. Additionally, multiple agencies may be involved in providing care during a particular period, with each agency regarding their intervention as a discrete episode.

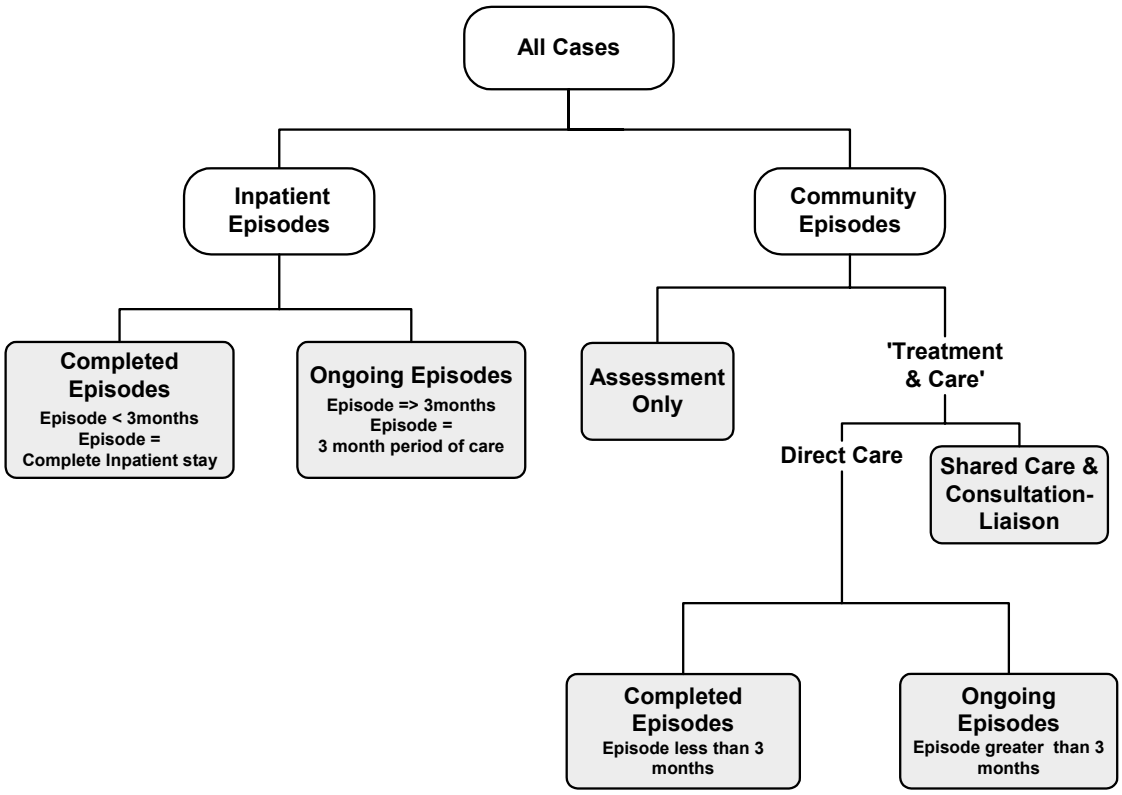
For the purposes of the study, an episode of care was defined as the whole or part of an episode of illness during which there is contact between a consumer and a health care provider or team of providers, is delivered in one setting and where there is no major change in the goal of intervention. Within this, episodes were defined as being either 'complete' or 'ongoing'.

Ongoing episodes were reviewed each three months and so, in these cases, an episode was defined as the care package delivered over a three month period.

Two principles governed the collection of episode data during the study:

- While an individual may have had multiple episodes of care over the course of the study, the person could be registered as having one episode at any one time.
- Each episode type has an associated data set. While the episode types share a number of data elements, all data were linked to a specific episode.

Figure 1. Episode of Care model.

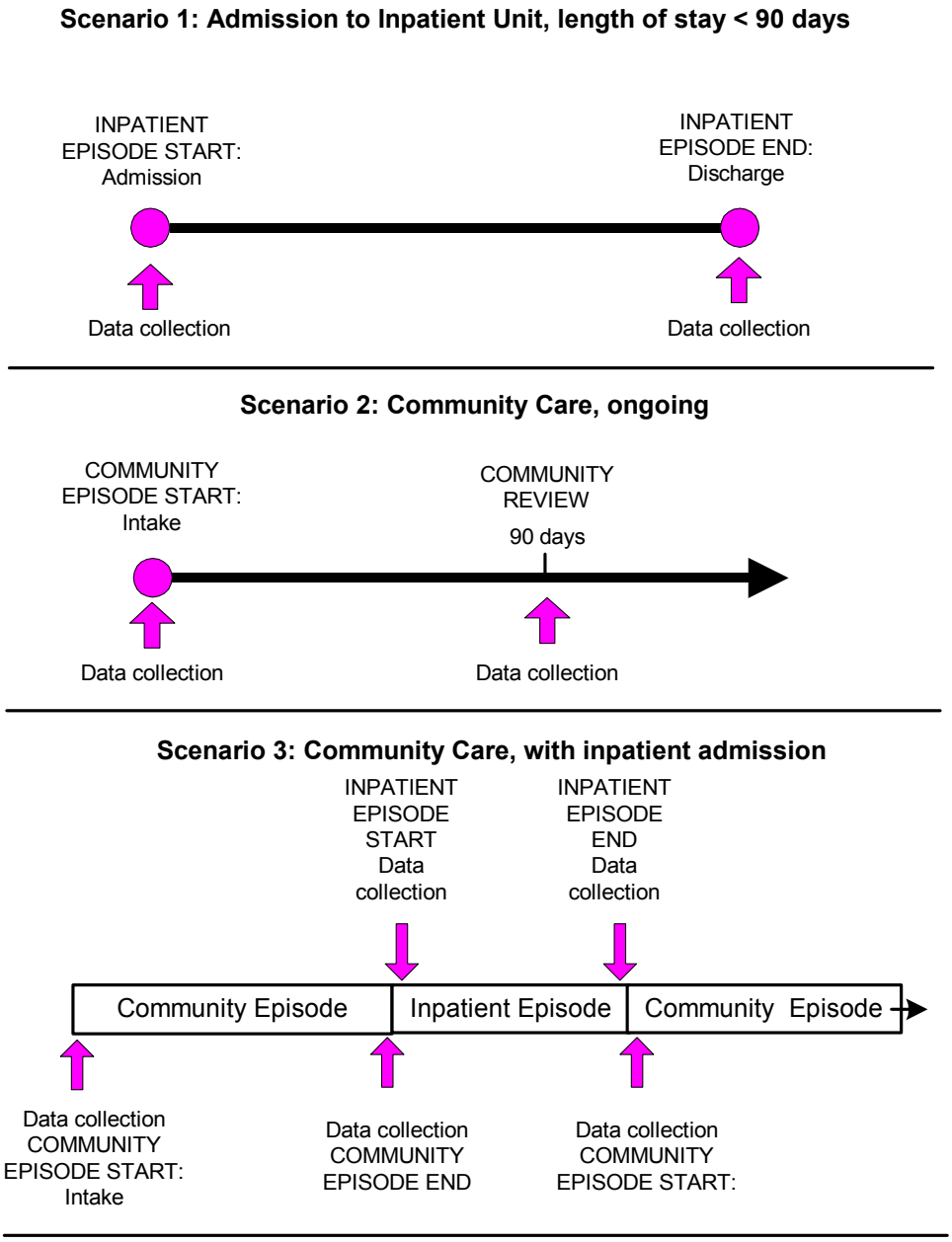


As will be seen in Figure 1, the study defined 6 types of episodes (shown in the shaded boxes). These are the episode types that are reported in this analysis. Clinical ratings (outcome measures) were not required for either Assessment Only or Shared Care episodes.

2.1.6 Data collection triggered by 'episode of care events'

The broad rule used in the collection of clinical data was that ratings were required at both *episode start* and *episode end.* Figure 2 provides some scenarios to illustrate the protocol which dictated those points at which the consumer outcome data were collected during the study.

Figure 2. Examples of the consumer outcome data collection points.



In many cases, the beginning and end of episodes was marked by some objective event such as discharge from hospital or completion of community treatment. However, because episodes were also deemed to have ended when a consumer has received 90 days of continuous care within the same treatment setting, the study protocol defined the following 3 critical events as the occasions for data collection to occur:

- *Episode Start*

This refers to the beginning of an inpatient or community episode of care. For the purposes of the study, episodes may start for a number of reasons. Regardless of the reason, the commencement of a new episode acts as a 'trigger' for a specific set of data to be collected.

- *Episode End*

This refers to the closure of an inpatient or community episode of care. As per Episode Start, episodes may end for a variety of reasons. Regardless of the reason, the end of an episode acts as a 'trigger' for a specific set of clinical data to be collected.

- *3-month review*

For the purposes of data collection, a 90-day review event was incorporated in the study design as a marker to identify ongoing episodes. In reality, it signalled two events – the end of one episode and the beginning of another. That is, it is both a Reason for Episode End and a Reason for Episode Start.

2.1.7 Overview of core consumer data set

Episode Start, Episode End and 3-month review events each had an associated data set. The relevant data sets for this analysis are:

- *Episode details* – these include items that identify episode type, start and end dates, and a number of data items that describe the consumer's situation during the episode and type of care provided. These represent attributes of the concept *Episode*. The specific data set to be collected is dependent on episode type.
- *Clinical ratings* – these comprise structured, multi-item rating scales that are administered by clinicians at the three episode events. The specific scales collected are dependent on the two factors: episode type and episode event (start, end, 3-month review).

Table 3 summarises the core consumer measures collected in this study. Details regarding each of the measures used in the study are provided in the National Study Manual (2001).

Table 3. Overview of core consumer data set.

Episode information	Socio-demographic details	Clinical details
Consumer NHI (de-identified) DHB HCAID Episode type Episode start date Episode end date Reason for episode start Reason for episode end	Date of birth Sex Ethnicity Area of usual residence	Principal psychiatric diagnosis Additional psychiatric diagnosis Level of clinical severity Level of functioning
		Plus these items for adults only Focus of care Legal status Criminal Justice Indicator (forensic services only)
	Plus for child and adolescent services only Living with Guardianship Family court involvement Juvenile Justice Indicator	Plus for child and adolescent services only Factors influencing health status (FIHS)

The consumer attribute data set comprised core attributes that were collected on all consumers in the study. In addition, a number of special items were collected for consumers over the age of 65 years and for consumers treated by child and adolescent mental health services. The specific instruments used to measure consumer outcomes are listed in Table 4. Further details and descriptions of the instruments are given in the appendix.

Table 4. Clinical rating instruments for measuring severity and level of functioning.

Consumer Population	Clinical ratings of severity	Levels of functioning
Adults	Health of the Nation Outcome Scales (HoNOS) The HoNOS is rated on a scale of 0 to 4. The higher the score, the greater the severity of the problem.	Life Skills Profile (LSP-16) The LSP is rated on a scale of 0 to 3. The higher the score, the greater the level of functional impairment.
Consumers over the age of 65 years	Health of the Nation Outcome Scales for consumers over the age of 65 (HoNOS 65+) The HoNOS 65+ is rated on a scale of 0 to 4. The higher the score, the greater the severity of the problem.	Resource Utilisation Group Activities of Daily Living (RUG-ADL) The RUG-ADL is rated on a scale from 1 to 4. The higher the score, the greater the level of functional impairment.
Children and adolescents	Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) The HoNOSCA is rated on a scale of 0 to 4. The higher the score, the greater the severity of the problem.	Children's Global Assessment Scale (CGAS) The CGAS is a single rating on a scale of 0 to 100. The lower the score, the lower the level of function

As shown in Table 3, other measures were also collected. They include diagnoses, legal/juvenile justice status, the Focus of Care (adults), Factors Influencing Health Status (child and youth) and family circumstances (child and youth). These measures are not outcome measures per se. Rather, they provide a context for interpreting the outcome measures collected in the study.

2.2 Data management and preparation

2.2.1 Security of data

The outcomes data that had been collected as part of the NZ-CAOS project were made available to the research team for detailed analysis purposes under the ethical clearance that has been obtained for the NZ-CAOS project.

The NZ-CAOS database had been developed with the primary objective of developing a casemix classification for specialist mental health services in New Zealand. With this objective in mind, a number of agreements were made with participating DHB sites, which formed part of the ethics approval granted for the study. The outcomes analysis is an extension of the NZ-CAOS project and arrangements made with the casemix analysis team to safeguard the privacy of individual consumers and individual sites extended to the analysis of the outcomes data. The database used was the final consolidated NZ-CAOS data set containing de-linked patient identifiers.

The security provisions were as follows:

- The researchers agreed to safeguard the privacy of personal information about consumers, as outlined in the Privacy Protocol for the Use of Personal Information in the NZ-CAOS Project (April 2001) and as set down in the New Zealand Privacy Act (1993), the New Zealand Health Information Privacy Code (1994) and in accordance with equivalent Privacy legislation in Australia (Commonwealth and State, as applicable).
- The researchers limited access to the de-identified data on a 'need-to-know' basis with the data protected against loss and unauthorised access, use, modification or disclosure.
- No data have been included in the report that would potentially allow for an individual consumer, staff member, or District Health Board site to be identified.

2.2.2 Preparation of the data for the outcomes analysis

For the NZ-CAOS study, several different 'episode of care' types were defined. One of these was 'Assessment Only' episodes. Outcome measures were not collected on 'Assessment Only episodes' and they have been excluded from the count of episodes shown in Table 5. Most of the clinical data were also not required for 'Shared Care' episodes[6]. In total, there were close to 18,000 episodes in the NZ-CAOS study. Under the study protocol, clinical ratings were required at the start and the end of each of these episodes.

Table 5. Episode of Care by Treatment Setting and Sector.

Episode of Care Type	Speciality					
	Adult		Child and Youth		Total	
	number	%	number	%	number	%
Inpatient	2755	19.0%	93	2.8%	2848	16.0%
Shared Community Care	897	6.2%	92	2.7%	989	5.6%
Direct Community Care	10815	74.8%	3167	94.5%	13982	78.5%
Total	**14467**	**100.0%**	**3352**	**100.0%**	**17819**	**100.0%**

The overview of clinical ratings reported in the NZ-CAOS report was based on the first clinical rating collected on each episode. It did not include any analysis of ratings at episode end or any analysis of change from start to end. Accordingly, for the outcomes analysis reported here, it was necessary to prepare a new outcomes data set that brought together the clinical ratings collected on each episode. For this reason, the outcomes data set is different to that used for the casemix analysis.

The data were prepared for the outcomes analysis using the following steps.

Step 1. Remove ratings with insufficient numbers for analysis

There were a total of only 247 RUG-ADL ratings in the outcomes data set. These RUG-ADL ratings had been collected on a total of 234 consumers. Only 13 consumers had two RUG-ADL ratings. Given this small number, the sample size was not sufficient to allow the RUG-ADL ratings to be analysed. The RUG-ADL was removed from the analysis.

[6] See the compendium casemix report for definitions of these types of episodes.

Step 2 Remove assessments that were not accompanied by a valid date

Ratings without a valid date were removed as it was not possible to verify that they were linked to a particular episode. As shown in Table 6, there were only a small number of ratings without valid dates.

Table 6. Number of ratings removed because of invalid dates.

Instrument	Number of ratings at beginning of this step	Number removed	Number remaining
HoNOS	24,452	14	24,438
LSP-16	20,135	14	20,121
HoNOSCA	5,697	11	5,686
CGAS	5,997	11	5,983
FIHS	3,090	11	3,079

Step 3. Remove duplicates

Different DHBs took different approaches in how they handled the '90 day review' rating (see Section 2.1.6). The 90-day review signalled both the end of one episode and the beginning of another. Some sites entered the 90-day review rating once but others entered it twice. In order to achieve consistency, all duplicate ratings were removed. An assessment was deemed to be a duplicate if, for the same individual, the assessment had the same episode start, episode end, and collection dates as another assessment.

Table 7. Number of ratings removed because they were duplicates.

Instrument	Number of ratings at beginning of this step	Number removed	Number remaining
HoNOS	24,438	2,078	22,360
LSP-16	20,121	1,792	18,329
HoNOSCA	5,686	626	5,060
CGAS	5,983	628	5,355
FIHS	3,079	17	3,062

Step 4. Remove assessments where the collection date did not correspond with either the start or the end date of the episode

Some ratings were recorded on dates that did not correspond to either the start or the end of an episode. As shown in Table 8, they represented only a small percentage and were removed.

Table 8. Number of ratings removed because collection date did not correspond with episode dates.

Instrument	Number of ratings at beginning of this step	Number removed	Number remaining
HoNOS	22,360	221	22,139
LSP-16	18,329	132	18,197
HoNOSCA	5,060	81	4,979
CGAS	5,355	78	5,277
FIHS	3,062	51	3,011

Step 5. Remove assessments where all constituent items were missing or invalid.

Some 'ratings' in the data set consisted completely of a row of numbers that indicated the item had not been completed. These were clearly not ratings and were removed. As shown in Table 9, the largest number of these were blank HoNOS ratings. They represented 6.3% of all HoNOS ratings at this point in the data preparation. 7.5% of ratings on the FIHS were also blank.

Table 9. Number of useable ratings after removal of ratings with all item values missing.

Instrument	Number of ratings at beginning of this step	Number removed	Number remaining
HoNOS	22,139	1,391	20,748
LSP-16	18,197	476	17,721
HoNOSCA	4,979	99	4,880
CGAS	5,177	117	5,160
FIHS	3,011	225	2,786

The ratings in the data set at the end of this step (shown in Table 9 above) constituted the data set used in the analysis to profile each measure. The remaining steps were then undertaken for the purpose of creating a subset of the episodes that had valid ratings at both the start and the end of the episode. These 'matched pairs' were used to measure change from the beginning to the end of each episode. This involved a considerable further loss of assessments (see Table 10). Further filtering later occurred during the analysis of each instrument based on the number of missing items in each assessment (see Chapter 3). Ratings with missing items were not removed from the analysis data set during the data preparation stage because they provide information about the items that are most prone to be omitted.

Step 6. Identify matching pairs of start and end assessments

This step involved creating a pair of ratings that matched the start and the end of an episode. As shown in Table 10, there were a significant number of ratings that were not part of a pair.

Table 10. Number of paired ratings.

	Start	End	All
HoNOS			
Both start and end ratings	7,805	7,805	7,805
Start or end, but not both	4,815	323	5,138
All	12,620	8,128	20,748
LSP			
Both start and end ratings	7,140	7,140	7,140
Start or end, but not both	3,071	370	3,441
All	10,211	7,510	17,721
HoNOSCA			
Both start and end ratings	1,716	1,716	1,716
Start or end, but not both	1,339	109	1,448
All	3,055	1,825	4,880
CGAS			
Both start and end ratings	1,994	1,994	1,994
Start or end, but not both	981	191	1,172
All	2,975	2,185	5,160
FIHS			
Both start and end ratings	268	268	268
Start or end, but not both	347	1,903	2,250
All	615	2,171	2,786

2.2.3 Summary of compliance rates for each of the instruments

Table 11 summaries the compliance rates on the various measures. The study protocol did not require ratings on assessment only or shared care episodes and these have been removed from this table.

Table 11. Summary of compliance rates on the various measures.

	Start	End	% total episodes
HoNOS			
Both start and end ratings	7,805	7,805	57.5%
Start or end, but not both	4,815	323	
Total episodes for which a rating was required	13,570	13,570	
% missing	7.0%	40.1%	
% episodes with at least one rating			95.4%
LSP			
Both start and end ratings	7,140	7,140	52.6%
Start or end, but not both	3,071	370	
Total episodes for which a rating was required	11,129	13,570	
% missing	8.2%	44.7%	
% episodes with at least one rating			78.0%
HoNOSCA			
Both start and end ratings	1,716	1,716	52.6%
Start or end, but not both	1,339	109	
Total episodes for which a rating was required	3,260	3,260	
% missing	6.3%	44.0%	
% episodes with at least one rating			97.1%
CGAS			
Both start and end ratings	1,994	1,994	61.2%
Start or end, but not both	981	191	
Total episodes for which a rating was required	3,260	3,260	
% missing	8.7%	33.0%	
% episodes with at least one rating			97.1%

In total, 95% of adult episodes had at least one HoNOS completed but only 58% had a HoNOS completed at both episode start and episode end. This is a little lower than the completion rates for the two Child and Youth measures. In total, 97% of episodes had at least one rating on the HoNOSCA and 53% had two. The rates for the CGAS were 97% and 61% respectively.

In the case of the LSP-16, the study protocol called for this measure to be collected at start, review and end for community episodes but not at episode start for inpatient episodes. The total number of episodes in this table has been adjusted to remove inpatient episode starts. The low proportion of episodes with at least one rating (78%) on the LSP-16 is due to low compliance at the end of episodes, not at the beginning.

Table 12 summaries the results across the four outcome measures collected in the study. If there had been full compliance with the study protocol, a total of almost 65,000 ratings would have been collected. In the final count, 48,509 ratings were collected. Only 57% of episodes in the study had ratings completed at both the start and the end of the episode, while a further 17%

had either a start rating or an end rating. No doubt some of these are attributable to the snapshot nature of the study rather than a reflection of what would happen in routine practice.

Table 12. Overall compliance rates.

	Start	End	Total	% of total
Both start and end ratings	18,655	18,655	37,310	57.5%
Start or end, but not both	10,206	993	11,199	17.3%
No rating	2,358	14,012	16,370	25.2%
Total	**31,219**	**33,660**	**64,879**	**100.0%**

Nevertheless, with 48,000 ratings, the NZ-CAOS study has produced a significant outcome data set for analysis. However, these figures relate to the total number of ratings in the data set. They include ratings that contain invalid or missing scores. Completion rates on the individual items are provided in the next chapter.

3. Results: completion rates of the individual items

The last chapter outlined the preparation of the data and gave compliance rates with the study protocol. In summary, over 48,000 ratings were collected, with 57% of episodes having ratings completed at both the start and the end of the episode and a further 17% having either a start rating or an end rating. This chapter takes those 48,000 ratings and examines the completion rates of the individual items for each instrument. The completion rates of the individual items give some indication of data quality, as well as the ease or difficulty that staff had in using the instruments. The chapter begins by looking at the two adult measures and moves on to consider the child and youth measures.

3.1 Adult measures

3.1.1 HoNOS

The distribution of missing or invalid ratings of the valid HoNOS assessments (see Table 9) was examined. Table 13 shows this, divided into inpatient and community settings. There were a little over 4,500 assessments in the former and 16,000 in the latter. In this and subsequent tables, HoNOS data and HoNOS65 data are grouped together.

Table 13. Percentage of missing HoNOS ratings, by setting and collection occasion.

Item	Inpatient		Community		All Inpatient	All Comm	All Start	All End
	Start	End	Start	End				
1	0.8	0.2	0.2	0.1	0.5	0.2	0.4	0.2
2	2.2	0.5	0.3	0.3	1.4	0.3	0.7	0.4
3	5.4	1.5	0.7	0.6	3.6	0.7	1.6	0.9
4	1.7	0.2	0.3	0.2	1.0	0.3	0.6	0.2
5	2.1	0.4	0.3	0.2	1.3	0.3	0.7	0.2
6	1.7	0.5	0.6	0.4	1.2	0.6	0.8	0.5
7	2.3	0.7	0.3	0.3	1.6	0.3	0.7	0.4
8	5.8	4.2	2.4	1.9	5.1	2.3	3.1	2.6
9	7.3	2.0	0.5	0.2	4.9	0.4	1.9	0.6
10	5.5	1.9	0.3	0.3	3.9	0.3	1.4	0.7
11	17.2	8.6	0.9	0.7	13.3	0.8	4.3	2.8
12	17.6	8.5	1.5	1.3	13.4	1.5	4.8	3.2

There is a consistent pattern for more items to be missing at the start of episodes than at the end, and in inpatient settings compared to community settings. Items 11 (accommodation problems) and 12 (occupation and leisure problems) are the most prone to attract missing ratings across most settings and occasions, followed by item 8 (other psychological problems), item 9 (relationship problems), item 10 (activities of daily living), and item 3 (drug and alcohol misuse). All other items are missing in less than 2% of assessments. The most commonly missing item in assessments conducted in the community is item 8, followed by item 12.

This pattern of missing ratings is very consistent with findings from Australia. The Victorian HoNOS field trial (Trauer et al., 1999)[7], the Victorian "Round One" analysis (Trauer, 2003)[8], and the Australian casemix study (MH-CASC, Buckingham et al., 1998)[9] all found higher rates of missing ratings in acute inpatient settings than in the community. In the Victorian "Round One" analysis and the MH-CASC study, items 8, 11 and 12 were the most frequently missed, while in the HoNOS field trial, the most frequently missed item was 11, followed by items 12, 3 and 8.

Table 14 shows the percentages of HoNOS assessments with different numbers of missing items.

Table 14. Percentage of HoNOS assessments with different numbers of missing items, by setting and collection occasion.

No. of missing items	Inpatient		Community		All Inpat.	All Comm	All Start	All End
	Start	End	Start	End				
0	73.4	85.1	94.4	95.4	78.8	94.8	90.1	92.7
1	7.3	6.0	4.2	3.5	6.7	4.0	4.8	4.2
2	9.8	5.8	0.8	0.6	8.0	0.7	2.6	2.0
3	3.2	1.6	0.2	0.2	2.4	0.2	0.8	0.5
4	2.8	0.8	0.2	0.2	1.9	0.2	0.7	0.3
5	1.7	0.5	0.0	0.0	1.2	0.0	0.4	0.2
6-11	1.8	0.2	0.2	0.1	1.0	0.1	0.5	0.1

As shown in Table 11 on page 16, there were more HoNOS assessments completed at the start of episodes than at the end. However, as Table 14 above shows, when the HoNOS was actually completed, 12% more HoNOS assessments were fully completed at the end of inpatient episodes than at the start. Very few items were omitted from assessments conducted in the community, either at the start or at the end. Apart from the start of inpatient episodes, assessments rarely had more than two items missing, suggesting that few assessments would be lost if two missing items was the permitted maximum.

Table 15 shows the percentage of assessments with no missing item by DHB. In this and subsequent tables, each DHB has been assigned a number. In this case, the DHBs are numbered from 1 to 8 based on the percentage of assessments at episode end with no missing items (the last column). The number assigned to each DHB varies according to the table. DHB 1 in this table is not the same DHB as DHB 1 in another table.

With one notable exception, all 12 items of the HoNOS were completed between 80% and 100% of the time. In one DHB (DHB 8) the rate was much lower in their inpatient service. In that service, nearly half of all HoNOS assessments had two items missing, and in the vast majority of cases, these were HoNOS items 11 and 12. In one DHB (DHB 1) 97% or more of HoNOS assessments were completed in full regardless of setting or occasion.

7 Trauer T, Callaly T, Little J, Shields R, Smith J. Health of the Nation Outcome Scales (HoNOS): results of the Victorian field trial. British Journal of Psychiatry 1999; 174: 380-388.

8 Trauer T, Analysis of outcome measurement data from the four Victorian "Round One" agencies. Unpublished report, 2003.

9 Buckingham W, Burgess P, Solomon S, Pirkis J, Eagar K, Casemix classification of mental health services. Canberra: Commonwealth Department of Health and Family Services, 1998.

By way of comparison, in the three studies referred to earlier in this section, the proportion of HoNOS assessments with no missing ratings was 96% or better.

Table 15. Percentage of HoNOS assessments with no missing items, by setting, collection occasion and DHB.

DHB	Inpatient		Community		All Inpat	All Comm	All Start	All End
	Start	End	Start	End				
1	96.8	98.0	99.8	100.0	97.3	99.9	98.9	99.4
2	88.4	92.2	98.2	99.8	90.2	98.8	97.0	98.5
3	80.6	93.4	96.4	96.6	86.2	96.5	93.4	95.9
4	86.6	96.1	91.2	95.3	91.9	93.4	88.6	95.8
5	90.3	88.8	97.0	97.8	89.7	97.4	94.7	95.1
6	77.3	83.1	94.6	94.9	79.9	94.7	92.2	92.7
7	79.1	91.6	91.2	89.0	85.0	90.5	88.5	89.9
8	17.9	37.5	88.6	91.4	25.6	89.5	72.7	74.7
All	73.4	85.1	94.4	95.4	78.8	94.8	90.1	92.7

Note: DHBs are ranked in this and subsequent tables based on the percentage of assessments with no missing items. Thus, the number assigned to each DHB varies by table.

3.1.2 LSP-16

As shown in Table 11 on page 16, the compliance rate with the LSP-16 was the lowest of all of the measures, with only about 500 LSP-16 assessments conducted in inpatient settings and 17,000 conducted in the community. However, when the LSP-16 was actually collected, the number of items missing was very small. For each of the sixteen items of the LSP-16, the missing rate was less than or equal to 0.5%, irrespective of setting (inpatient or community) or collection occasion (start or end of episode).

Table 16. Percentage of LSP-16 assessments with different numbers of missing items, by setting and collection occasion.

No of missing items	Inpatient		Community		All Inpat	All Comm	All Start	All End
	Start	End	Start	End				
0	98.1	97.4	98.5	98.8	99.8	98.6	98.5	98.7
1	1.9	2.2	0.9	0.8	2.0	0.9	0.9	0.9
2			0.4	0.3		0.4	0.4	0.3
3		0.4	0.1		0.2		0.1	
4-15			0.1	0.1		0.1	0.1	0.1

Across all settings and occasions, 97% or more of LSP-16 assessments were completed with no missing ratings. This is somewhat higher than the 93% found in the Victorian "Round One" analysis (see footnote 8). Very small numbers indeed have more than one item missing.

Table 17. Percentage of LSP-16 assessments with no missing items, by setting, collection occasion and DHB.

DHB	Inpatient		Community		All Inpat	All Comm	All Start	All End
	Start	End	Start	End				
1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
2	100.0	100.0	99.9	99.8	100.0	99.9	99.9	99.8
3	89.3	89.3	100.0	100.0	89.3	100.0	99.4	99.2
4	100.0	100.0	98.9	98.3	100.0	98.7	99.0	98.4
5	-	-	98.6	99.5	-	99.0	98.6	99.5
6	100.0	94.4	98.1	99.3	96.8	98.7	98.2	98.7
7	100.0	100.0	97.5	96.7	100.0	97.2	97.5	96.7
8	97.1	97.1	96.3	95.1	97.1	95.9	96.3	95.3
All	98.1	97.4	98.5	98.8	99.8	98.6	98.5	98.7

Note: DHBs are ranked in this table based on the percentage of assessments with no missing items. Thus, the number assigned to each DHB varies by table.

As shown in Table 17, high rates of completion were achieved at all eight DHBs. One DHB did not have any recorded LSP-16 administrations in the inpatient setting. Every LSP-16 undertaken by DHB 1 had every item completed.

3.2 Child and Youth measures

This section presents the same analysis, but this time for the Child and Youth measures. The CGAS is not included as it consists of a single rating.

3.2.1 HoNOSCA

The distribution of missing or invalid ratings of the 4,880 valid HoNOSCA assessments (see Table 9) was examined. Of the 4,880, all but 104 were conducted in the community. This analysis only includes the 4,776 community assessments. Of these, 2,963 (62%) were collected at the start of an episode, and 1,813 (38%) at the end.

Over 99% of items 1 to 12 on the HoNOSCA were complete. Items 13 (Poor school attendance), 14 (Problems with knowledge or understanding about the nature of the child or adolescent's difficulties), and 15 (Problems with lack of information about services or management of the child or adolescent's difficulties) were missed in between 1.5% and 2% of assessments.

Table 18. Percentage of missing HoNOSCA ratings in the community, by collection occasion.

Item	Start	End	All
1	0.1	0.1	0.1
2	0.3	0.2	0.2
3	0.3	0.0	0.2
4	1.0	0.5	0.8
5	0.8	0.3	0.6
6	0.4	0.4	0.4

Item	Start	End	All
7	0.8	0.3	0.6
8	0.7	0.2	0.5
9	0.2	0.3	0.2
10	0.2	0.2	0.2
11	0.3	0.2	0.3
12	0.2	0.1	0.2
13	1.9	1.4	1.7
14	1.4	1.4	1.4
15	1.5	1.5	1.5

Table 19 shows the same information grouped by the number of missing items. About 95% or more of HoNOSCA assessments were completed with no missing ratings at both start and end. Very small numbers indeed have more than two items missing.

Table 19. Percentage of HoNOSCA assessments conducted in the community with different numbers of missing items, by collection occasion.

No of missing items	Start	End	All
0	94.5	95.8	95.0
1	3.0	2.3	2.8
2	1.5	1.3	1.4
3	0.5	0.4	0.5
4-14	0.4	0.2	0.3

Table 20 provides completion results by DHB. High rates of completion were achieved in the seven DHBs that collected the HoNOSCA.

Table 20. Percentage of HoNOSCA assessments conducted in the community with no missing items, by collection occasion and DHB.

DHB	Start	End	All
1	99.9	99.8	99.5
2	98.0	98.1	98.1
3	93.0	95.3	94.0
4	93.2	90.9	92.2
5	91.9	91.4	91.7
6	92.6	84.6	90.6
7	89.5	100.0	89.6
All	**94.5**	**95.8**	**95.0**

Note: DHBs are ranked in this and subsequent tables based on the percentage of assessments with no missing items. Thus, the number assigned to each DHB varies by table.

3.2.2 FIHS

The distribution of missing or invalid ratings of the 2,786 valid FIHS assessments (see Table 9) was examined. Of the 2,786, all but 93 were conducted in the community. This analysis only includes the 2,693 community assessments. Of these, 615 (23%) were collected at the start of an episode, and 2,078 (77%) at the end. Table 21 shows that ratings are missing in between 1% and 5% of assessments. There is a slight tendency for more ratings to be missed in assessments conducted at the end of episodes than at the start.

Table 21. Percentage of missing FIHS ratings in the community, by collection occasion.

Item	Start	End	All
1	3.6	4.7	4.4
2	2.1	2.8	2.6
3	2.8	3.0	3.0
4	1.3	1.7	1.6
5	2.4	3.7	3.4
6	4.1	3.7	3.8
7	4.9	5.2	5.1

Table 22 shows the percentages of FIHS assessments with different numbers of items missing. Overall, about 94% of assessments have no missing items. For the few assessments that have missing items, they are more likely to have 4, 5 or 6 items missing than 1, 2 or 3.

Table 22. Percentage of FIHS assessments conducted in the community with different numbers of missing items, by collection occasion.

No of missing items	Start	End	All
0	94.8	94.1	94.3
1	0.5	0.6	0.6
2	0.3	0.4	0.4
3	0.8	0.6	0.7
4	1.3	1.1	1.1
5	1.3	1.8	1.7
6	1.0	1.4	1.3

Table 23 shows the completion rates on the FIHS instrument by DHB. Most FIHS assessments were complete in 6 of the 7 DHBs that provided data. Of the 2,693 FIHS assessments, only 154 had any missing items, and 137 of these came from the one DHB.

Table 23. Percentage of FIHS assessments conducted in the community with no missing items, by collection occasion and DHB.

DHB	Start	End	All
1	100.0	100.0	100.0
2	100.0	100.0	100.0
3	100.0	100.0	100.0
4	100.0	100.0	100.0
5	99.5	98.0	98.4
6	95.1	98.0	97.5
7	75.0	68.4	70.1
All	**94.8**	**94.1**	**94.3**

Note: DHBs are ranked in this and subsequent tables based on the percentage of assessments with no missing items. Thus, the number assigned to each DHB varies by table.

3.3 Summary of results in relation to item completion

As shown in Table 24, the overall item completion rates were very high. The HoNOS had the poorest rate of item completion and the LSP-16 had the best. The completion rates for the two Child and Youth measures were similar. With all instruments having item completion rates of over 90%, it can be concluded that the data are of a sufficiently high quality for the analyses presented in the remainder of the report.

Table 24. Percentage of assessments conducted with no missing items, by instrument.

Instrument	Start	End
HoNOS	90.1%	92.7%
LSP-16	98.1%	98.7%
HoNOSCA	95.8%	95.0%
FIHS	94.1%	94.3%

4. Results: technical performance of the instruments

Multi-item instrument, such as the HoNOS, HoNOSCA, and LSP-16, may be scored at the item, subscale, and total score level. Different users of outcome measures may find different levels of scoring suitable for their purpose. The subscales are based on content similarity as well as on statistical grounds, i.e. the tendency for the items comprising the subscale to correlate with each other. This chapter reports on the technical performance of the instruments in the New Zealand context. It includes a comparison with relevant Australian results.

4.1 HoNOS

The subscale structure of the HoNOS was explored using confirmatory factor analysis (CFA), which is a form of structural equation modelling[10]. One main application of CFA is to assess the goodness of fit between a set of data and a presumed underlying structure. The original work on the current version of HoNOS was based on exploratory factor analysis, but no details were published[11]. For the present purposes, the only published study of the subscale structure of the HoNOS[12] was used. That study used 2,035 HoNOS assessments with no missing ratings from the Victorian HoNOS field trial (see footnote 7). The present data comprises 18,908 HoNOS records with no missing ratings.

In CFA, assessment of the goodness of fit between the model and the data is achieved by the use of a variety of fit indices. These include the Root Mean Square (RMS), the Root Mean Square Error of Approximation (RMSEA), the Normed Fit Index (NFI), and the Tucker-Lewis Index (TLI)[13]. Acceptable levels for these are <.05, .05 to .08, ≥.90, and ≥.90 respectively. Table 25 shows the fit indices for the conventional four-factor structure from the earlier study and the present data. In both these analyses, the same technical adjustments (allowing certain errors to be correlated) were made.

Table 25. Fit indices following CFA of HoNOS four subscale structure, by data source.

	Acceptable level	Trauer 1999	NZ-CAOS 2003
RMR	<.05	.10	.11
RMSEA	0.5-0.8	.07	.12
NFI	≥.90	.81	.82
TLI	≥.90	.75	.74

Table 25 shows that RMR, NFI and TLI for the conventional structure are well outside of the recommended levels of acceptability. RMSEA was acceptable in the 1999 study but not in the current data. On the basis of the poor fit on the conventional structure, Trauer (1999) proposed an alternative, five-factor, structure. This retains the conventional Impairment and Social subscales, but has a new Behaviour subscale comprising only items 1 and 3, a new Depression subscale comprising items 2, 7, 8 and 9, and a single item Hallucinations/delusions "subscale" comprising only item 6. This is displayed diagrammatically in Figure 3. The ovals show the

[10] See Tabachnick, B. G. & Fidell, L. S. Using Multivariate Statistics. 1996 New York: Harper Collins. Chapter 14.

[11] College Research Unit, Health of the Nation Outcome Scales: Report on Research. 1996 London: Royal College of Psychiatrists.

[12] Trauer T. The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health* 1999; 8: 499-509.

[13] Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990; **107**: 238-246.

subscales, the rectangles the items, and the arrows the relationships between the items and the subscales.

Figure 3. Five subscale model of HoNOS.

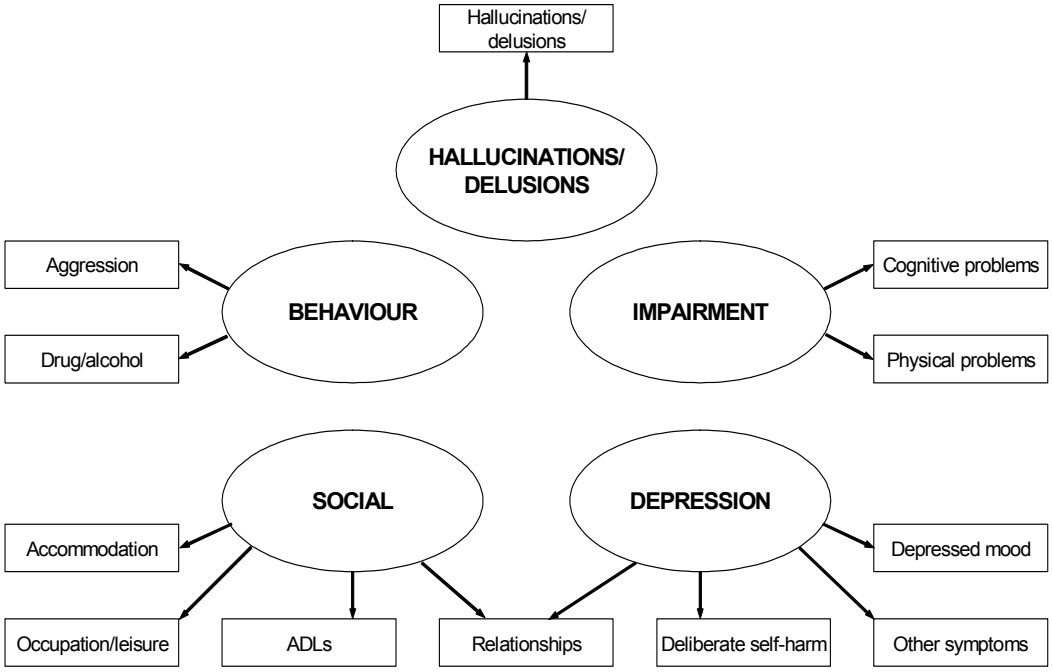


Table 26 shows the corresponding indices for the alternative five factor structure proposed by Trauer (1999).

Table 26. Fit indices following CFA of HoNOS five subscale structure, by data source.

	Acceptable level	Trauer 1999	NZ-CAOS 2003
RMR	<.05	.05	.06
RMSEA	0.5-0.8	.04	.08
NFI	≥.90	.93	.93
TLI	≥.90	.90	.90

It shows that the fit indices for the five factor structure are within or around the recommended levels, indicating that the five factor structure represents a far better fit to the data than the conventional four factor structure.

One desirable feature of a subscale is internal consistency, which suggests that the items are tapping different aspects of the same attribute, and not several different attributes. Internal consistency is usually measured by Cronbach's alpha. Table 27 shows the alpha values for the subscales of the four- and five-subscale models.

Table 27. Cronbach's alpha values of subscales in four- and five-subscale models.

Sub Scale	Four subscales	Five subscales
Behaviour	.57	.53
Impairment	.44	.44
Symptoms	.30	-
Social	.76	.76
Depression	-	.62
Hallucinations/delusions	-	-

The Impairment and Social subscales are identical in both models, hence the alphas are the same. Internal consistency is not applicable to the Hallucinations/delusions "subscale", since it comprises a single item. The Depression subscale, however, shows a much higher internal consistency than the Symptoms subscale. The Cronbach's alpha for the total score (i.e. all twelve items) is .80. This is considerably higher than the values of .72 found in the Victorian Round One study (see footnote 8), .73 found in the Stedman et al. study[14] and .75 obtained in the Victorian HoNOS field trial (see footnote 7).

One further matter of interest in the performance of the HoNOS is the performance of item 8. This item allows the clinician to select a problem from a provided list, and rate it in the usual fashion, i.e. with a rating of 1 to 4. The provided list of problems is: A Phobic; B Anxiety and panics; C Obsessive and compulsive problems; D Reactions to severely stressful events and traumas; E Dissociative problems; F Somatoform problems; G Problems with appetite, over and under eating; H Sleep; I Sexual problems; and J Other. Of the 20,748 HoNOS administrations, 10,526, or just over half, had item 8 rated as zero, which meant that none of the list of problems applied. Of the 10,222 assessments in which item 8 was rated, B Anxiety and panics was rated in 33.9% of assessments, H Sleep in 17.7%, and D Reactions to severely stressful events and traumas in 16.6%. Thus, of the nine substantive problems listed, three accounted for about 74% of the ratings. This is similar to findings elsewhere. The same three choices jointly accounted for 46% of all item 8 assessments in the Victorian HoNOS field trial (see footnote 7) and for 68% in the Victorian "Round One" data (see footnote 8).

4.2 LSP-16

A CFA similar to that performed with the HoNOS data was conducted on the 17,471 LSP-16 records with no missing ratings. The goodness of fit indices are reported in Table 28. None of the indices achieve the recommended levels, indicating that there is a poor fit between the data and the four subscale structure.

Table 28. Fit indices following CFA of LSP-16 four subscale structure.

	Acceptable level	NZ-CAOS 2003
RMR	<.05	.08
RMSEA	0.5-0.8	.12
NFI	≥.90	.80
TLI	≥.90	.77

[14] Stedman T, Yellowlees P, Mellsop G, Clarke R, Drake S, Measuring consumer outcomes in mental health. Canberra, A.C.T.: Department of Health and Family Services, 1997.

The Cronbach's alpha values for the Withdrawal, Self Care, Compliance and Antisocial subscales are .83, .77, .81 and .80 respectively, indicating reasonable levels of internal consistency. The Cronbach's alpha for the whole scale (i.e. all sixteen items) is .89. This is an identical value to that found in the Victorian Round One findings (see footnote 8).

4.3 HoNOSCA

There were 4,596 HoNOSCA assessments on consumers between the ages of 3 and 20, and which had no missing items. Of these, 75, or 1.6%, were conducted in inpatient settings; in the interests of consistency, these have been omitted, leaving 4,521. A CFA similar to that performed with the HoNOS and LSP-16 data was conducted on these assessments. Preliminary analyses indicated that the fit would be improved by allowing certain error terms to be correlated. Table 29 shows the fit indices after allowing the error terms of items 3 and 4, 1 and 12, and 4 and 13, to be correlated. Three of the indices (RMR, NFI and TLI) are outside acceptable levels, and the fourth, RMSEA, is marginal. Taken together, these results suggest that the present HoNOSCA data do not fit the four subscale structure well. Several other possible subscale structures were tried, but none achieved acceptable levels of fit with the data.

Table 29. Fit indices following CFA of HoNOSCA four subscale structure.

	Acceptable level	NZ-CAOS 2003
RMR	<.05	.07
RMSEA	0.5-0.8	.08
NFI	≥.90	.85
TLI	≥.90	.80

The Cronbach's alpha values for the Behaviour, Impairment, Symptoms, Social and Information subscales are .52, .31, .34, .62 and .81 respectively. Thus, the internal consistencies of the first three subscales are weak, the fourth is moderate, and the fifth is very strong. The Cronbach's alpha for the whole scale (based on the first thirteen items) is .76, which is moderately strong.

4.4 FIHS

The inter-correlations of all seven FIHS items are all positive, ranging from .13 to .54, with a mean of .31. The degree of internal consistency among the items is represented by a Cronbach's alpha of .76. The correlations of the items with the total score range from .54 to .72. These results suggest that the FIHS total score reflects to a reasonable degree the global effect of the several factors influencing health status.

4.5 Summary of the technical performance of the instruments

The factor structure of the HoNOS is consistent with work in Australia that indicated that the conventional four subscale structure does not fit the obtained data particularly well. Better fit, and arguably better clinical meaning, with no loss of internal consistency, can be achieved through a five subscale structure, although at the expense some loss of parsimony (i.e. five subscales are less parsimonious than four).

Neither does a four subscale structure fit the LSP-16 particularly well, although there is no alternative currently available. However, all four subscales show good internal consistency, as does the total scale.

A weak degree of fit of subscales with the data was also observed in the HoNOSCA, but here the situation was exacerbated by most of the subscales having poor internal consistency as well. Only the Information subscale, which is not used in the derivation of the total score, and the total score itself, showed acceptable level of internal consistency. These findings suggest that not too much reliance should be placed on HoNOSCA subscales.

The internal consistency of the FIHS is the same as for the HoNOSCA and indicates that the total score is a reasonably robust measure of the combined effects of the factors included in the scale.

5. Results: relationship between the outcome measures

This chapter reports on the relationship between the various outcome measures. The first section looks at the adult measures – the relationship between the Focus of Care and the HoNOS, between the Focus of Care and the LSP-16 and between the HoNOS and the LSP-16. The second section looks at the child and youth measures – the relationship between the HoNOSCA and the CGAS, between the HoNOSCA and FIHS and between the CGAS and FIHS. It concludes with a brief summary of findings.

Throughout this chapter, results are described as being either 'statistically significant' or not. It is important to emphasise at the beginning that, because of the large volume of ratings, almost all results are statistically significant. Statistical significance must be distinguished from the concept of clinical significance or importance. Some results presented in this chapter are statistically significant but clinically unimportant.

5.1 Adults

5.1.1 Focus of Care and HoNOS

The Focus of Care (FoC) is a retrospectively completed instrument, signifying the main goal of treatment over the preceding period of care. Thus, it was expected to be completed at review and end of episodes, but not at the start. It was included in the data collection as it was expected to help explain differences in the cost and outcomes between adult consumers. This section explores the relationship between the FoC and the outcome measures. The FoC was not collected for Child and Youth episodes.

There were 1,970 review or end of episode assessments in inpatient settings with a FoC rating. There were 8,645 similar assessments in the community.

Figure 4 shows the HoNOS subscale and total scores at episode start for the four types of FoC in inpatient settings. Episodes rated as 'Acute FoC' have higher Behaviour and Symptoms scores, while the 'Maintenance FoC' is associated with lower scores across all four subscales. The 'Functional Gain FoC' has the highest score on the Social subscale while the 'Intensive Extended FoC' sits in between on all subscales.

Figure 4. HoNOS scale scores at episode start by Focus of Care – inpatient settings.

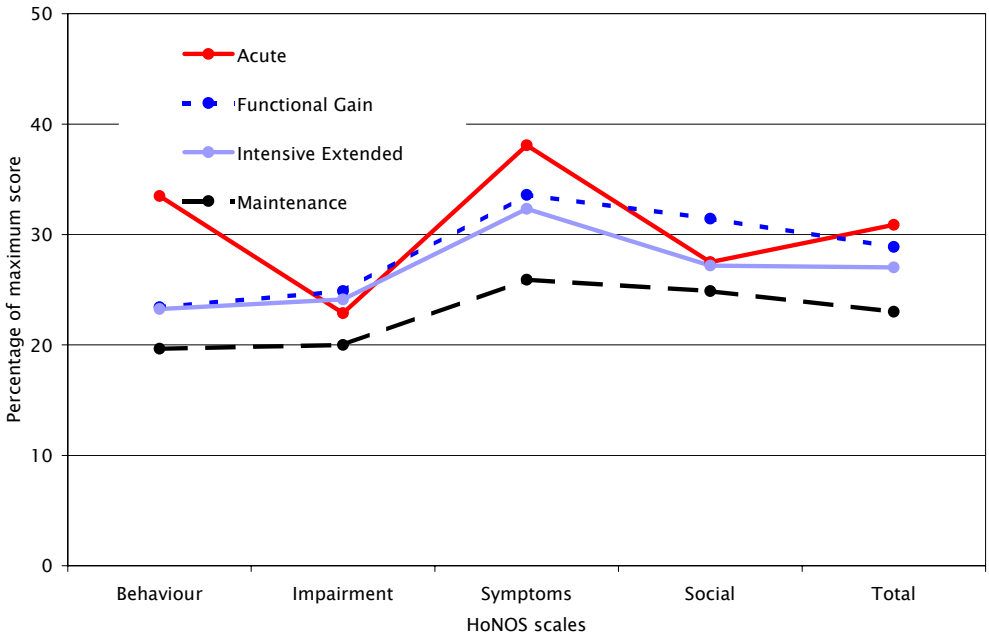
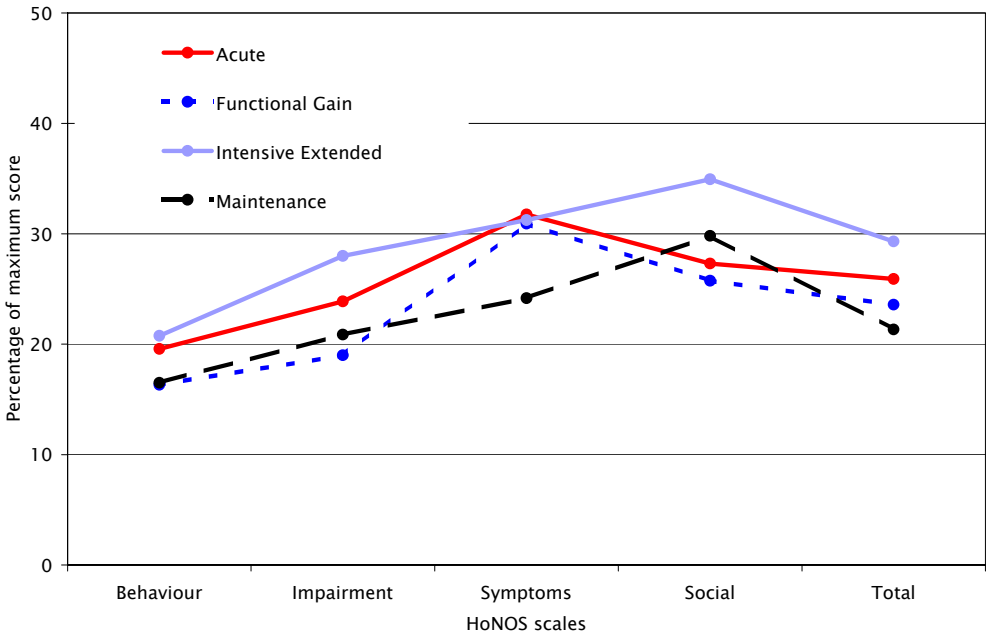


Figure 4 can be compared with the corresponding community profiles displayed in Figure 5. In the community, the Intensive Extended profile of HoNOS subscale scores is generally the highest, followed by the 'Acute FoC'. The 'Maintenance FoC' group have low Symptom and Behaviour scores while the 'Functional Gain FoC' is distinguished by low scores on the Behaviour, Impairment and Social subscales.

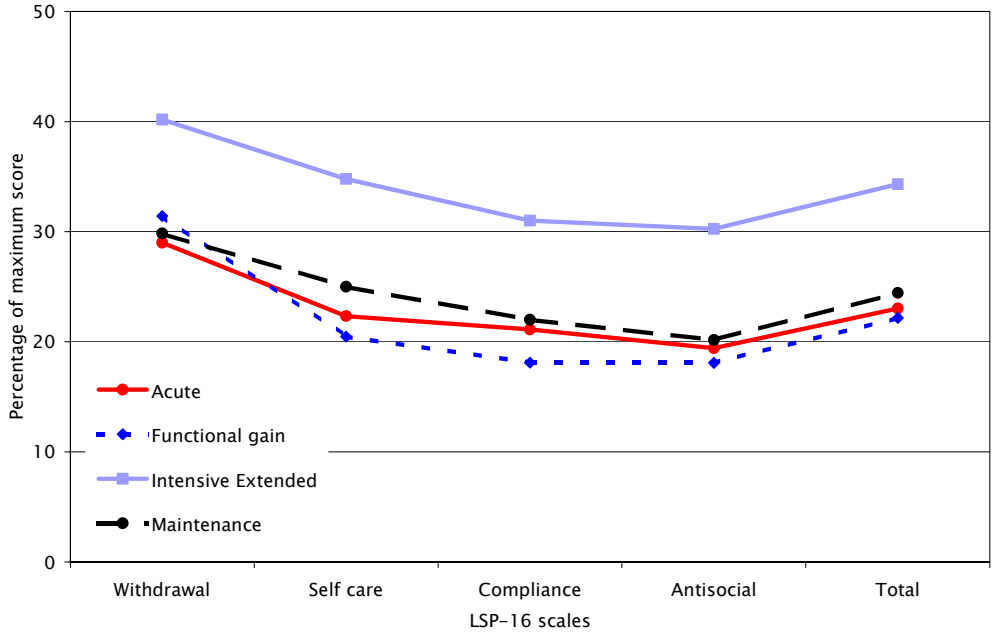
Figure 5. HoNOS scale scores at episode start by Focus of Care – community settings.



5.1.2 Focus of Care and LSP-16

Figure 6 shows the LSP-16 subscale and total scores for associated with the four type of FoC in community settings. The Intensive Extended category is associated with considerably higher scores than the other three FoC categories, which are not much different from each other.

Figure 6. LSP-16 scale scores at episode start by Focus of Care – community settings.



5.1.3 HoNOS and LSP-16

There were 15,187 occasions when a HoNOS and an LSP-16 were obtained on the same consumer, on the same date, at the same point (start or end) of an episode, and with no missing ratings on either instrument.

Table 30 shows the inter-correlations between the subscale and total scores. A recent study (see footnote 8) reported similar inter-correlations based on the initial implementation of adult outcome measures in Victoria, Australia. These values are included in parentheses for comparison purposes.

Table 30. Inter-correlations of HoNOS and LSP-16 subscale and total scores.

	Behaviour	Impairment	Symptoms	Social	HoNOS total
Withdrawal	.29 (.22)	.33 (.31)	.37 (.41)	.58 (.57)	.56 (.57)
Self care	.27 (.28)	.48 (.43)	.23 (.32)	.52 (.59)	.52 (.59)
Compliance	.39 (.32)	.27 (.21)	.22 (.22)	.41 (.41)	.44 (.42)
Antisocial	.59 (.54)	.30 (.25)	.31 (.23)	.52 (.48)	.60 (.54)
LSP-16 total	.48 (.43)	.45 (.40)	.36 (.39)	.66 (.67)	.68 (.69)

The highest correlations of subscales between the two instruments are the Behaviour subscale of the HoNOS with the Antisocial subscale of the LSP-16, and the Social subscale of the HoNOS with the Withdrawal subscale of the LSP-16. There are also very substantial correlations between Social (HoNOS) and both Self Care and Antisocial (LSP-16), and between Impairment (HoNOS) and Self care (LSP-16). The correlation of .68 between the two total scores is very high, and represents 46% (.68 squared) of common variance.

The present correlations are remarkably close to those obtained from the Victorian data, especially the correlation of the two total scores. It is also noteworthy that a correlation of -.65 between the total scores of the HoNOS and the full form of the LSP was found by Stedman et al. (see footnote 14) in their field trial of outcome measures (their correlation was negative because the full LSP is scored so that a low score means low function while the LSP-16 is scored so that a low score means high function. The scoring in the LSP-16 was reversed to make it consistent with the scoring of the HoNOS.).

5.2 Children and Youth

5.2.1 HoNOSCA and CGAS

There were 4,524 occasions when a HoNOSCA and a CGAS were obtained on the same consumer, on the same date, at the same point (start or end) of an episode, and with no missing ratings on the HoNOSCA.

Table 31 shows that the correlations between the CGAS and HoNOSCA scores range from moderate to very large. All the correlations are negative because higher scores on the HoNOSCA are indicative of greater severity, but on the CGAS higher scores indicate better functioning. The correlation of -.63 between the CGAS and the HoNOSCA total score is very close to the -.6 reported by Gowers et al.[15]

[15] Gowers S, Bailey-Rogers S, Shore A, Levine W. The Health of the Nation Outcome Scales for Child & Adolescent Mental Health (HoNOSCA). Child Psychology & Psychiatry Review 2000; 5: 50-56.

Table 31. Correlations between CGAS and HoNOSCA items, subscales and total score.

	Correlation	Qualitative description of the effect[16]
1 Disruptive, antisocial, or aggressive behaviour	-.41	Large
2 Over-activity, attention or concentration	-.38	Large
3 Non-accidental self-injury	-.24	Moderate
4 Alcohol, substance/solvent misuse	-.15	Moderate
5 Scholastic or language skills	-.37	Large
6 Physical illness or disability problems	-.18	Moderate
7 Hallucinations, delusions	-.24	Moderate
8 Non-organic somatic symptoms	-.16	Moderate
9 Emotional and related symptoms	-.38	Large
10 Peer relationships	-.46	Very large
11 Self-care and independence	-.39	Large
12 Family life and relationships	-.36	Large
13 Poor school attendance	-.29	Large
14 Lack of knowledge – nature of difficulties	-.29	Large
15 Lack of information – services/management	-.28	Large
Behaviour Subscale	-.48	Very large
Impairment Subscale	-.39	Large
Symptoms Subscale	-.41	Large
Social Subscale	-.55	Very large
Information Subscale	-.30	Large
Total	-.63	Very large

5.2.2 HoNOSCA and FIHS

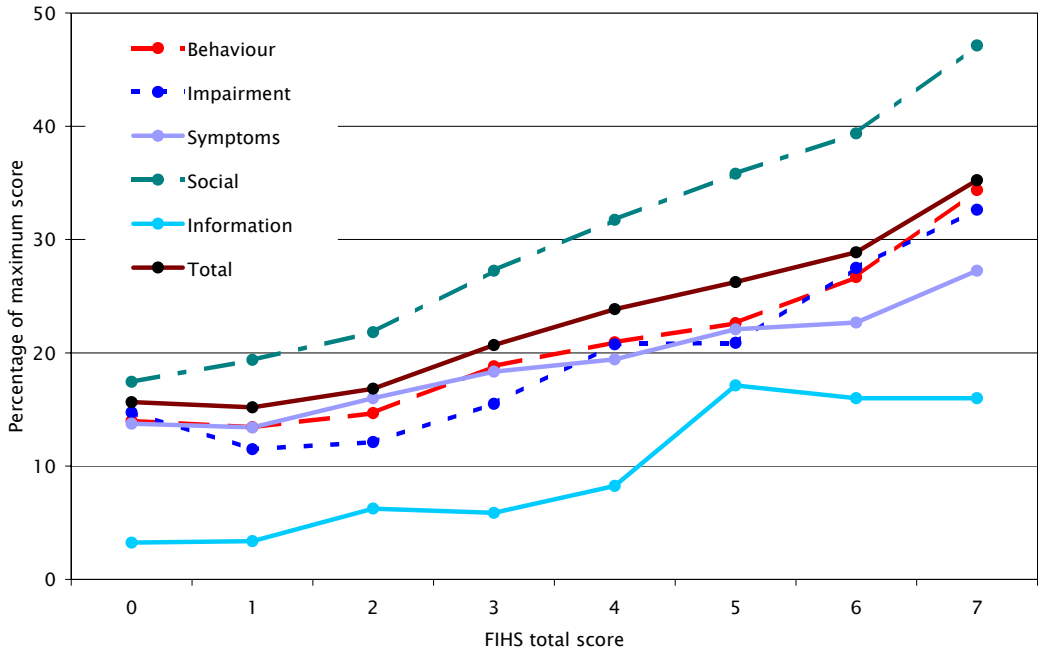
There were 2,159 occasions when a HoNOSCA and a FIHS were obtained on the same consumer, on the same date, at the same point (start or end) of an episode, and with no missing ratings on the HoNOSCA or the FIHS.

Figure 7 shows the relationships between the FIHS total score and the HoNOSCA subscales and total score. It shows that consumers with FIHS total scores of zero were rated on average between 14% and 17% of the maximum score on Behaviour, Impairment, Symptoms, Social and total score, and 3% of the maximum score of Information.

[16] See Sulloway FJ. *Born to rebel: birth order, family dynamics, and creative lives*, New York: Vintage. 1997, page 372. As Sulloway points out: "Assigning qualitative descriptions to effect sizes is dependent on the context." Also Rosenthal R, Rubin DB. A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* 1982; 74: 166-169.

There is a general tendency for HoNOSCA scores to increase with increasing FIHS total scores. The Behaviour, Impairment, and total scores of consumers with FIHS total scores of seven are about double those whose total score is zero, while the Social score are around two-and-a-half times as high. Information scores are the lowest of all, but it can be seen that the Information scores of those consumers with FIHS total scores of 5 and over are distinctly higher than those whose FIHS scores are 4 or lower.

Figure 7. HoNOSCA subscale and total scores by FIHS total score.



5.2.3 CGAS and FIHS

There were 2,605 occasions when a CGAS and a FIHS were obtained on the same consumer, on the same date, at the same point (start or end) of an episode, and with no missing ratings on the FIHS. Table 32 shows the mean CGAS scores associated with the presence or absence of each of the FIHS items.

Table 32. Mean CGAS scores associated with FIHS items, and correlations with age.

		FIHS rating	
		Yes	No
1	Maltreatment syndromes	55.6	61.4
2	Problems related to negative life events in childhood	58.2	62.5
3	Problems related to upbringing	57.4	63.7
4	Problems related to primary support group, including family circumstances	58.5	63.6
5	Problems related to social environment	57.2	62.8
6	Problems related to certain psychosocial circumstances	56.7	62.0
7	Problems related to other psychosocial circumstances	53.1	61.5

The differences in mean CGAS scores between the presence and absence of each of the FIHS items is significant, although the correlations are all in the .05 to .07 range, representing a qualitatively very small effect. In all cases the presence of an FIHS factor is associated with a lower CGAS score. The differences are fairly consistent; with the presence of a factor on the FIHS being associated with a decrement of 4 to 8 points in the CGAS. The correlation between the FIHS total score and CGAS is -.29.

5.3 Summary of the relationship between the outcome measures

In summary, only the Intensive Extended category of the Focus of Care measure in the community setting is reliably associated with different ratings on the two adult outcome measures. There were only small and inconsistent differences in HoNOS scores between the other three Focus of Care categories: Acute, Functional Gain, and Maintenance, especially between Acute and Maintenance, which should arguably be clinically quite different from each other. The relationship between the Focus of Care and the LSP-16 is very similar: Intensive Extended is associated with elevated LSP-16 scores, with the other three groups being virtually indistinguishable.

The relationship between the HoNOS and the LSP-16 is very strong, particularly in their total scores, and this is consistent with findings elsewhere. However, the correlation is not strong enough to suggest that one may substitute for the other – apart from the Social subscale of the HoNOS, the content and coverage of the two instruments is different.

The relationship between the HoNOSCA and CGAS is also strong, and also consistent with findings elsewhere. There is a clear association between FIHS and HoNOSCA, especially the Social subscale. Elevated scores on the Information subscale only appeared to be associated with FIHS total scores of five and above. The relationship between FIHS and CGAS, while clearly present, seemed to be smaller in overall effect. On average, the presence of each of the FIHS factors was associated with 4 to 8 points less on the CGAS.

6. Results: relationship between the outcome measures and other clinical and socio-demographic variables

This chapter presents results on the relationship between the outcome measures and other information about consumers that was captured in the study. It begins with an analysis of the adult measures – the HoNOS and the LSP-16. The relationship between each of these two measures and the age, sex, diagnosis and legal status of the consumer is examined. The final section looks at the child and youth measures – the HoNOSCA, the CGAS and FIHS. The relationship between each of these measures and the age, sex and diagnosis of the consumer is examined. Two other key consumer attributes were captured in the study – ethnicity and a measure of social deprivation. These are reported in Chapter 9 and Chapter 11.

As with other chapters, results are described in this chapter as being either 'statistically significant' or not. However, it is important to again emphasise that, because of the large volume of ratings, almost all results are statistically significant. Statistical significance must be distinguished from the concept of clinical significance or importance. Some results presented in this chapter are statistically significant but lack clinical meaning.

6.1 Adult measures

6.1.1 HoNOS

Age

Correlations were calculated between consumer age and HoNOS item, subscale and total scores. All valid assessments were used for this purpose, so some consumers contributed more than one observation. The numbers of observations employed varied between 19,981 and 20,745.

Because of the large numbers of observations, even small correlations achieved statistical significance. All coefficients in Table 33 are significant apart from those for the Social subscale and the total score. The Behaviour items and subscale are negatively related to age (younger consumers are rated higher), the Impairment subscale is positively related to age (older consumers were rated higher on Impairment), and the correlations of the Symptom items and subscale are all quite close to zero, as are three of the Social items. Item 10 (activities of daily living) was rated somewhat higher in older consumers.

The correlation between age and the total HoNOS score was very close to zero despite some significant correlations with the subscales. It appears that the combined effect of positively and negatively correlated items is to produce a total score that is unrelated to age.

Table 33. Correlations between age and HoNOS items, subscales and total score.

Item	Correlation	Qualitative description of the association[17]
1 Aggressive/overactive/ agitated behaviour	-.07	Small
2 Non-accidental self-injury	-.16	Moderate
3 Problem drinking or drug-use	-.23	Moderate
4 Cognitive problems	.17	Moderate
5 Physical illness or disability problems	.34	Large
6 Hallucinations/delusions	-.06	Small
7 Depressed mood	-.02	Very small
8 Other behavioural or psychological problems	-.03	Very small
9 Relationship problems	-.04	Very small
10 Activities of daily living	.13	Small
11 Problems with accommodation	-.07	Small
12 Problems with occupation/leisure	-.06	Small
Behaviour Subscale	-.20	Moderate
Impairment Subscale	.32	Large
Symptoms Subscale	-.06	Small
Social Subscale	-.00	Very small
Total	-.01	Very small

Sex

Of the 20,748 valid HoNOS assessments, 10,603 (51%) were of males and 10,145 (49%) were of females. Table 34 presents the mean scores of males and females on the HoNOS items. For 8 of the 12 HoNOS items, males scored significantly higher (worse) than females. On three (items 5, 7 and 8) females scored significantly worse than males, and on one (item 2) there was very little difference. Despite being statistical significance on eleven items, the differences in the average scores are, in real terms, very small.

[17] See footnote 16.

Table 34. Mean HoNOS item scores by sex.

Item	Males	Females
1 Aggressive/overactive/ agitated behaviour	1.01	.78
2 Non-accidental self-injury	.56	.58
3 Problem drinking or drug-use	.82	.41
4 Cognitive problems	.96	.78
5 Physical illness or disability problems	.74	.88
6 Hallucinations/delusions	1.08	.68
7 Depressed mood	1.01	1.29
8 Other behavioural or psychological problems	1.00	1.31
9 Relationship problems	1.58	1.30
10 Activities of daily living	1.30	1.10
11 Problems with accommodation	1.00	.63
12 Problems with occupation/leisure	1.13	.74

Figure 8 shows the conventional subscale and total scores broken down by sex. It is apparent that males score much higher than females on the Behaviour and Social subscales, females score slightly higher than males on the Symptom subscale, and there is little difference on the Impairment subscale. The next analysis looks at the differences between the sexes on an alternative set of subscales (see Section 4.1). This shows a greater sex difference on the Behaviour subscale, while females are rated higher on the Depression subscale than males, and the reverse is true for the Hallucinations/delusions item/"subscale".

Figure 8. Mean HoNOS subscales scores by sex.

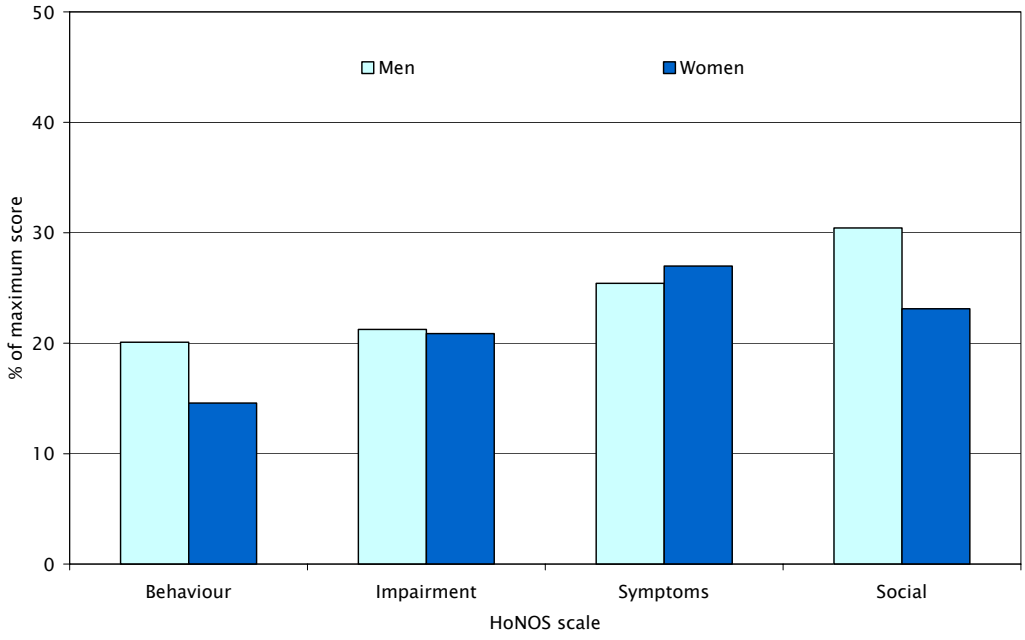
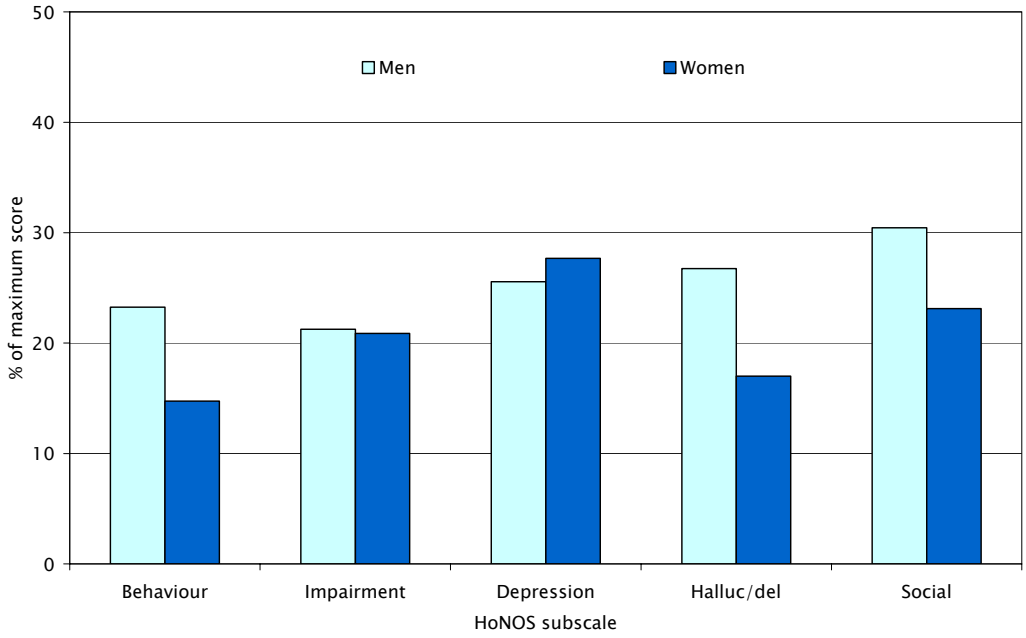


Figure 9. Mean HoNOS alternative subscales scores by sex.



The mean total score (12.01) for males was significantly higher than that of females (10.35) ($t_{(20746)}$ = 15.2, p < .001). However, in real terms, the average difference was 1.66 points on a scale with a maximum score of 48 points. It is doubtful that this difference is clinically significant.

Diagnosis

Principal psychiatric diagnoses were combined into 16 major diagnostic classes. The Somatoform, Physiological, Sexual and Retardation classes were omitted on account of small numbers, Developmental and Childhood disorders were omitted on account of inappropriateness (for the HoNOS), and the class of Other was omitted on account of lack of informational value. The remaining nine diagnostic classes are represented in Table 35, which shows the mean total HoNOS scores in inpatient and community settings.

Table 35. Mean HoNOS total score by setting and diagnosis.

	Inpatient	Community	All
Substance misuse	13.7	15.5	15.0
Personality disorder	11.8	15.7	14.7
Organic	13.5	13.4	13.4
Schizophrenia	11.8	11.8	11.8
Obsessive-compulsive	8.7	10.9	10.6
Stress	10.5	10.0	10.1
Eating	15.6	9.2	10.0
Mood disorders	11.0	9.5	9.8
Anxiety	11.4	9.4	9.5

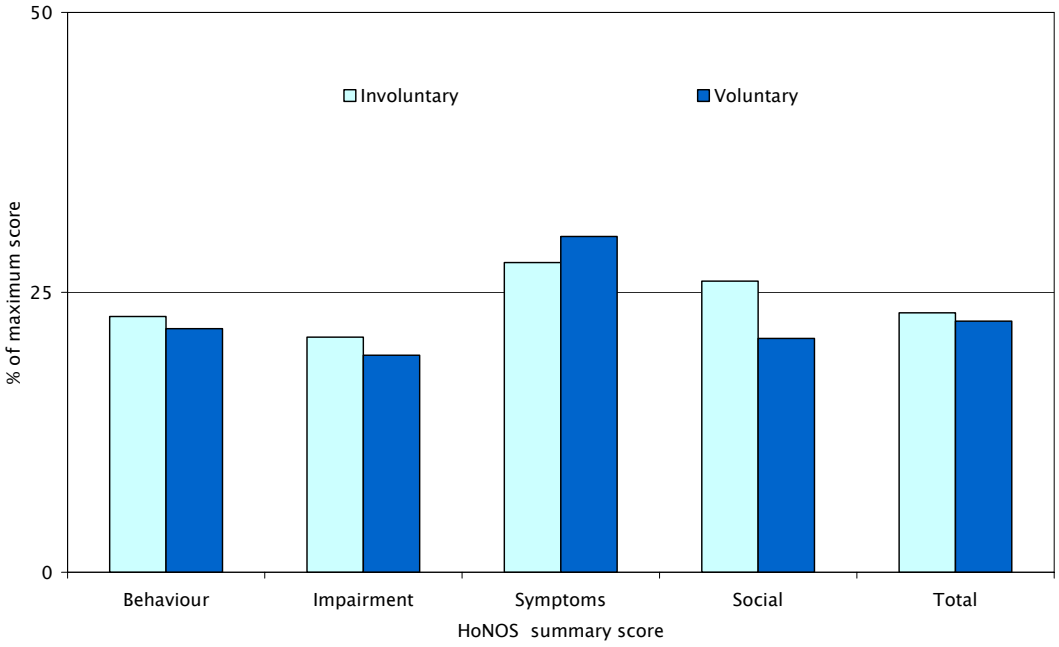
For certain of the diagnostic groups (Organic, Schizophrenia and Stress) there was little difference in the HoNOS total scores between inpatient and community settings. For others (Mood, Anxiety and Eating), inpatient scores were higher than community scores and, for Substance Misuse and Personality Disorders, the community mean was higher than the inpatient mean. The highest HoNOS totals were associated with the Substance Misuse, Personality, and Organic disorders, and the lowest with Anxiety, Mood, Eating, Stress, and Obsessive-Compulsive disorders.

Legal status

Of the 20,748 valid HoNOS assessments, 1,410 (6.8%) had no recorded legal status. Of the remainder, 4,587 were conducted in inpatient settings, and of these 3,278 (72%) were involuntary, and 14,751 were conducted in the community, and of these 2,763 (18%) were involuntary.

Figure 10 shows the mean HoNOS subscale and total scores for voluntary and involuntary episodes in inpatient settings. The difference in the total scores is small and not significant, as is the difference on the Behaviour subscale. Impairment and Social subscale scores are significantly higher in involuntary episodes, but the Symptom subscale scores are lower.

Figure 10. HoNOS conventional summary scores by legal status – inpatient settings.



These subscale results can be compared with the alternative set of subscales (see Section 4.1) which are shown in Figure 11. The Impairment, Social and Total scores are identical to those in Figure 10. Involuntary episodes are associated with significantly higher scores on the Behaviour and Hallucination/delusions subscale, and significantly lower scores on the Depression subscale.

Figure 11. HoNOS alternative summary scores by legal status – inpatient settings.

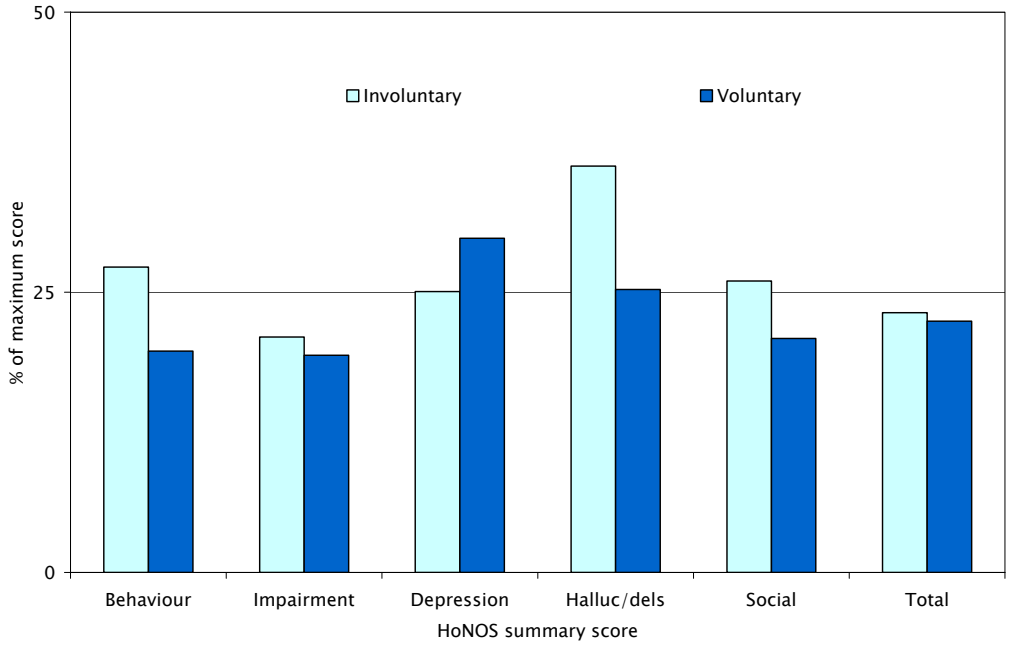
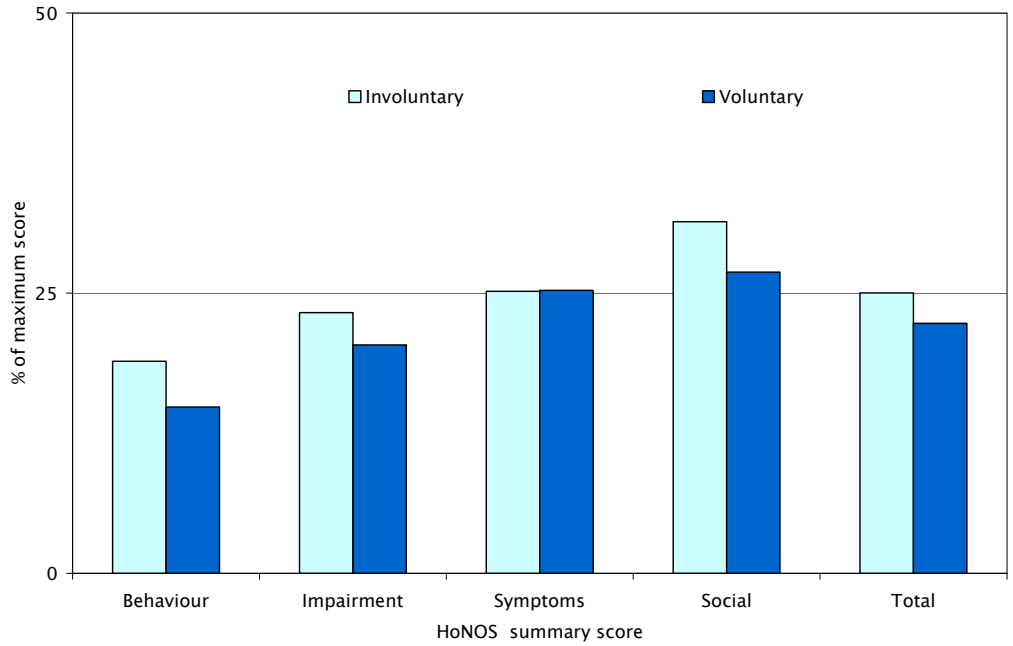


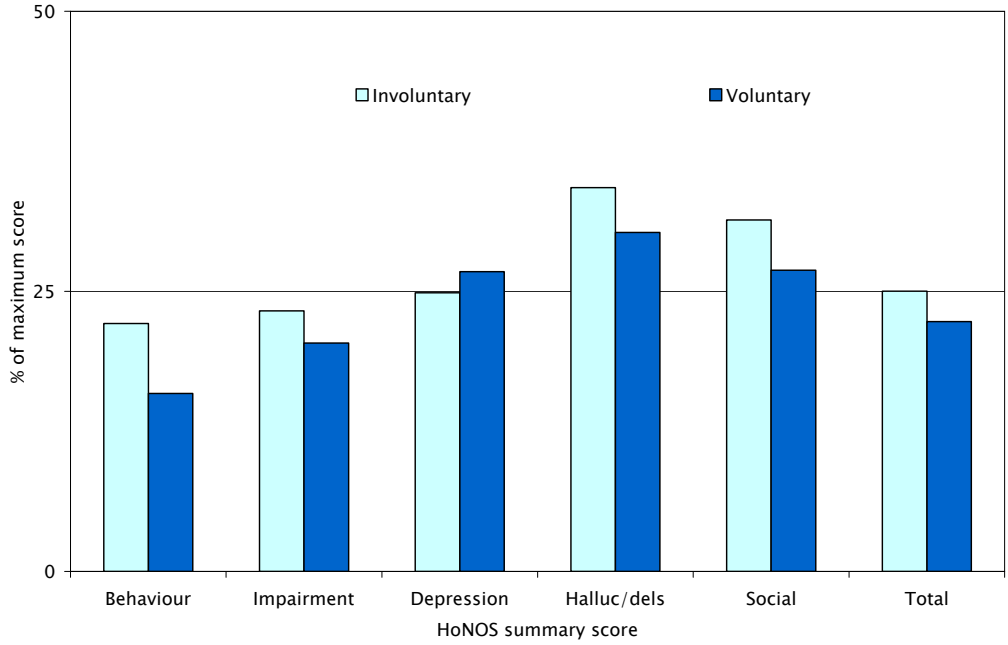
Figure 12 uses the conventional subscales to compare the differences between voluntary and involuntary HoNOS assessments in the community. Behaviour, Impairment and Social subscale scores are significantly higher in involuntary than voluntary episodes, but there is no difference on the Symptoms subscale.

Figure 12. HoNOS conventional summary scores by legal status – community settings.



These results can be compared to the alternative set of HoNOS subscales in Figure 13. Once again, the Impairment, Social and Total scores are the same as in Figure 12. The lack of difference on the Symptom subscale obscures significant differences on the Depression and Hallucinations/delusions subscales, where voluntary episodes are higher on the former and involuntary episodes higher on the latter.

Figure 13. HoNOS alternative summary scores by legal status – community settings.



6.1.2 LSP-16

Age

Correlations were calculated between consumer age and LSP-16 item, subscale and total scores. All valid assessments were used for this purpose, so some consumers contributed more than one observation. The numbers of observations varied between 17,468 and 17,717 depending on the item.

Table 36. Correlations between age and LSP-16 items, subscales and total score.

Item	Correlation	Qualitative description of the association[18]
1 Initiating and responding to conversation	.01	Very small
2 Withdrawal from social contact	.02	Very small
3 Warmth to others	-.05	Very small
4 Well groomed	.03	Very small
5 Does this person wear clean clothes	.04	Very small
6 Neglect of physical health	.03	Very small
7 Violence to others	-.11	Small
8 Make or keep friendships	.04	Very small
9 Maintenance of an adequate diet	-.01	Very small
10 Look after/take own prescribed medication	.01	Very small
11 Willing to take prescribed medication	-.07	Small
12 Cooperation with health services	-.08	Small
13 Problems living with others in the household	-.06	Small
14 Offensive behaviour	-.06	Small
15 Irresponsible behaviour	-.13	Small
16 What sort of work is this person capable of	.46	Very large
Withdrawal Subscale	-.00	Very small
Self care Subscale	.15	Moderate
Compliance Subscale	-.05	Small
Antisocial Subscale	-.08	Small
Total	.03	Very small

Correlations between age and LSP-16 items, subscales and total score are mostly small to very small, but there is one major exception. Item 16 (What sort of work is this person generally capable of?) correlated very strongly with age, indicating that older consumers were rated as having greater difficulties in this respect.

Sex

Of the 17,721 valid LSP-16 assessments, 8.901 (50%) were male and 8,821 (50%) were female. Table 37 presents the mean scores of males and females on the LSP-16 items.

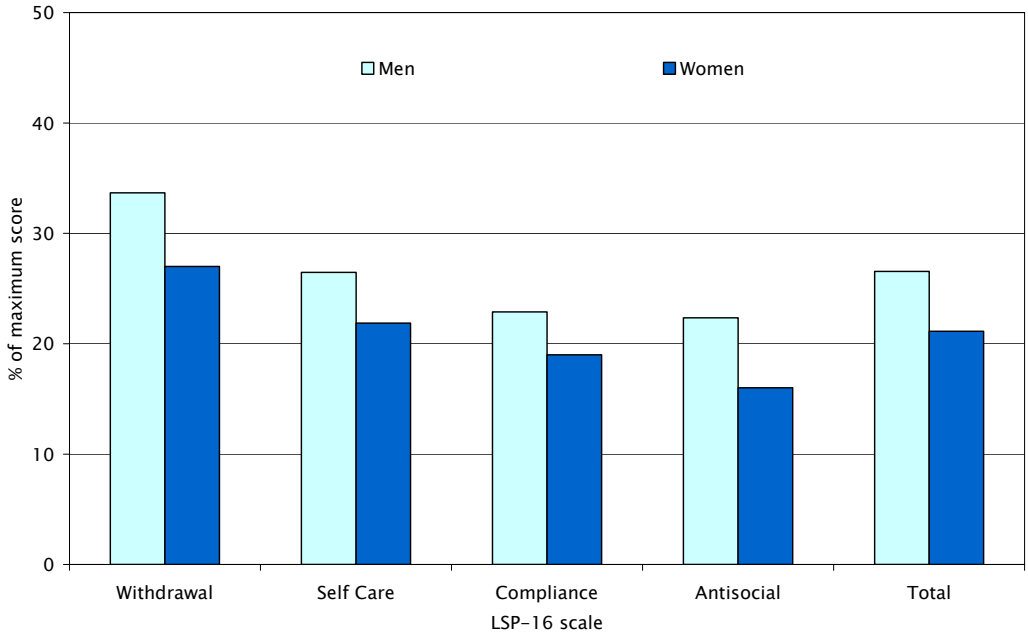
[18] See footnote 16.

Table 37. Mean LSP-16 items scores by sex.

Item	Males	Females
1 Initiating and responding to conversation	.78	.53
2 Withdrawal from social contact	1.06	.94
3 Warmth to others	1.08	.81
4 Well groomed	.79	.46
5 Does this person wear clean clothes	.65	.32
6 Neglect of physical health	.60	.48
7 Violence to others	.51	.29
8 Make or keep friendships	1.12	.95
9 Maintenance of an adequate diet	.58	.63
10 Look after/take own prescribed medication	.76	.59
11 Willing to take prescribed medication	.65	.57
12 Cooperation with health services	.64	.54
13 Problems living with others in the household	.83	.76
14 Offensive behaviour	.49	.28
15 Irresponsible behaviour	.84	.59
16 What sort of work is this person capable of	1.34	1.39

For 15 of the 16 LSP-16 items, males scored significantly higher (worse) than females. The exception was item 9 (Maintenance of an adequate diet), on which females scored significantly worse than males. Despite being statistical significant, the actual difference is very small for all 16 items. The next figure (Figure 14) shows the subscale and total scores. Consistent with the majority of the item scores, males scored higher on all the subscales scores and the total score. For both sexes, Withdrawal was the most highly rated subscale, and Antisocial the lowest.

Figure 14. LSP-16 subscale and total scores by sex.



Diagnosis

The same set of diagnoses that was used in relation to the HoNOS (see Section 6.1.1) is used here. Most (97%) LSP-16 assessments were conducted in the community. The only diagnoses for which reasonable numbers of LSP-16 assessments were obtained in inpatient settings were Schizophrenia and Mood Disorders. Table 38 shows the mean LSP-16 total scores broken down by setting and diagnosis.

Table 38. Mean LSP-16 total score by setting and diagnosis.

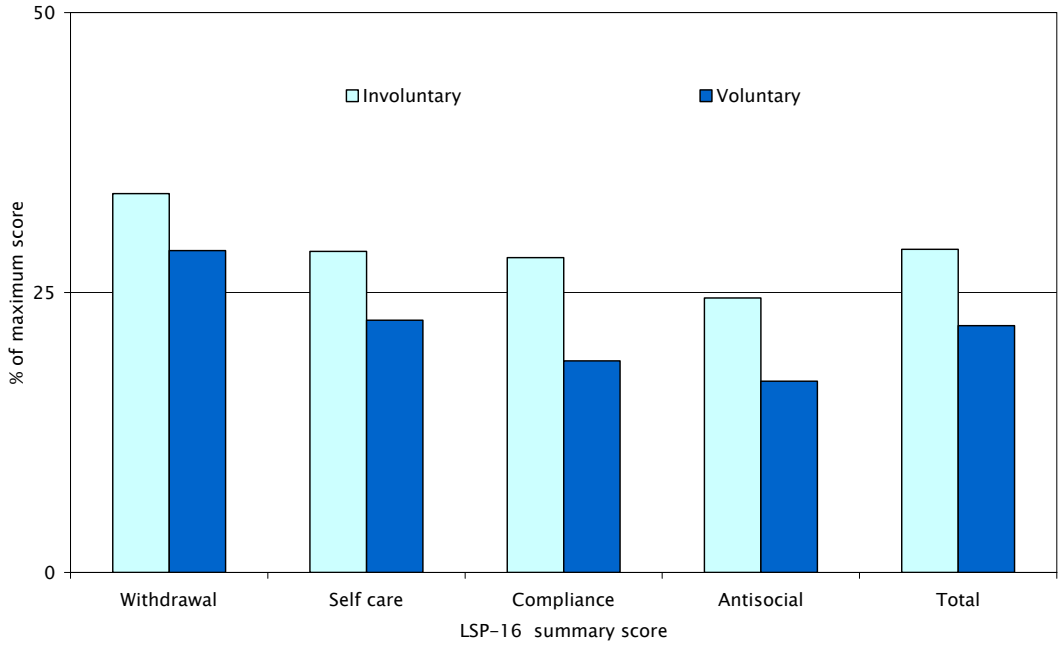
	Inpatient	Community	All
Organic		16.3	16.3
Personality disorder		13.9	14.0
Schizophrenia	17.3	13.4	13.6
Substance misuse		13.1	13.2
Obsessive-compulsive		9.7	9.7
Eating		9.1	9.5
Stress		9.0	9.0
Mood disorders	17.1	8.7	8.8
Anxiety		8.2	8.3

In terms of the LSP-16 total score, there are two distinct groups. Organic, Substance misuse, Schizophrenia and Personality Disorders are associated with relatively high scores, and Mood, Anxiety, Obsessive-Compulsive, Stress and Eating Disorders are associated with relatively low scores. However, there is almost no difference between the scores for Schizophrenia and Mood Disorders in the inpatient setting.

Legal status

Of the 17,721 valid LSP-16 assessments, 1,166 (6.6%) had no recorded legal status. Of the remainder 16,025, or 97%, were conducted in community settings, and of these 2,998 (19%) were involuntary, and 13,027 (81%) were voluntary. Given the comparatively small number of LSP-16 assessments conducted in inpatient settings, only the community episodes were analysed. Figure 15 shows the LSP-16 summary scores broken down by legal status. Involuntary status is clearly associated with significantly higher scores on all LSP-16 subscales and the total score.

Figure 15. LSP-16 summary scores by legal status – community settings.



6.2 Child and Youth

6.2.1 HoNOSCA

Age

Correlations were calculated between consumer age and HoNOSCA item, subscale and total scores. Ages recorded ranged from 1 to 27 years. For the purpose of examining the relationship with age, assessments where age was less than 3 and greater than 20 years were omitted. Otherwise, all valid assessments were used for this purpose, so some consumers contributed more than one observation. The numbers of observations varied between 4,648 and 4,857 depending on the item.

Table 39. Correlations between age and HoNOSCA items, subscales and total score.

Item	Correlation	Qualitative description of the association[19]
1 Disruptive, antisocial, or aggressive behaviour	-.18	Moderate
2 Over-activity, attention or concentration	-.20	Moderate
3 Non-accidental self-injury	.20	Moderate
4 Alcohol, substance/solvent misuse	.33	Large
5 Scholastic or language skills	-.07	Small
6 Physical illness or disability problems	.02	Very small
7 Hallucinations, delusions	.21	Moderate
8 Non-organic somatic symptoms	.03	Very small
9 Emotional and related symptoms	.10	Small
10 Peer relationships	-.01	Very small
11 Self-care and independence	.01	Very small
12 Family life and relationships	-.01	Very small
13 Poor school attendance	.21	Moderate
14 Lack of knowledge – nature of difficulties	-.02	Very small
15 Lack of information – services/management	-.01	Very small
Behaviour Subscale	-.00	Very small
Impairment Subscale	.14	Small
Symptoms Subscale	.16	Moderate
Social Subscale	.07	Small
Information Subscale	-.02	Very small
Total	.06	Small

In Table 39, the HoNOSCA total score was based on the first 13 items. As with the HoNOS, the low correlation of age with total score obscured the fact that certain items have opposite relationships with age: only the first two items (Disruptive, antisocial, or aggressive behaviour and Over-activity, attention or concentration) are associated with younger age, while items 4 (Alcohol, substance/solvent misuse), 7 (Hallucinations, delusions), and 13 (Poor school attendance) are associated with older age.

Sex

Of the 4,880 valid HoNOSCA assessments, 4,776 (98%) were conducted in the community, leaving too few inpatient assessments to analyse. Of the 4,776 assessment in the community, age was known for 4,774. 2,939 (62%) were male and 1,835 (38%) were female. The next table presents the mean scores of males and females on the HoNOSCA items.

[19] See footnote 16.

Table 40. Mean HoNOSCA items, subscales and total scores by sex.

Item	Males	Females
1 Disruptive, antisocial, or aggressive behaviour	1.68	1.05
2 Over-activity, attention or concentration	1.61	0.87
3 Non-accidental self-injury	0.16	0.48
4 Alcohol, substance/solvent misuse	0.33	0.49
5 Scholastic or language skills	1.37	0.79
6 Physical illness or disability problems	0.30	0.35
7 Hallucinations, delusions	0.16	0.20
8 Non-organic somatic symptoms	0.30	0.47
9 Emotional and related symptoms	1.75	2.06
10 Peer relationships	1.61	1.42
11 Self-care and independence	0.75	0.63
12 Family life and relationships	2.03	2.01
13 Poor school attendance	0.53	0.60
14 Lack of knowledge – nature of difficulties	0.58	0.59
15 Lack of information – services/management	0.35	0.33
Behaviour Subscale	3.81	2.94
Impairment Subscale	1.70	1.27
Symptoms Subscale	2.23	2.80
Social Subscale	4.93	4.71
Information Subscale	0.94	0.96
Total	12.59	11.56

While most of the sex differences in Table 40 are statistically significant, they are all quite small in terms of the actual scores. The clearest differences between the sexes are on items 1 (Disruptive, antisocial, or aggressive behaviour), 2 (Over-activity, attention or concentration), 5 (Scholastic or language skills) and the Behaviour subscale, on which the males scored much higher than the females, and items 3 (Non-accidental self-injury) and the Symptoms subscale, on which the females scored much higher than the males. Sex differences on the two Information items (14 and 15) and the Information subscale were very small and non-significant. On average, males scored about one point higher on the total score (calculated on the basis of items 1 to 13) than females. This is not clinically significant.

Diagnosis

Most (98%) of HoNOSCA assessments were conducted in the community; the only diagnoses for which reasonable numbers of HoNOSCA assessments were obtained in inpatient settings were Schizophrenia and Mood disorders. There were too few instances of diagnoses of Organic, Somatoform, Physiological and Personality Disorder to allow for meaningful analysis. Table 41 shows the mean HoNOSCA total scores (based on items 1 to 13) broken down by setting and diagnosis.

Table 41. Mean HoNOSCA total score by setting and diagnosis.

	Inpatient	Community	All
Retardation		17.4	17.4
Schizophrenia	20.8	14.6	15.7
Developmental		15.3	15.3
Substance misuse		14.0	14.0
Childhood		13.1	13.2
Stress		11.8	11.9
Mood disorders	15.4	10.7	10.9
Obsessive-compulsive		10.0	10.7
Anxiety		10.2	10.2
Eating		6.3	7.1

In terms of the HoNOSCA total score, the highest scores are associated with Retardation, Schizophrenia, Developmental and Substance Misuse, and the lowest with Eating Disorders. The largest diagnostic class, Childhood Disorders, accounts for about one third of the assessments in Table 41. On average, HoNOSCA total scores of consumers with Schizophrenia and Mood disorders diagnoses in inpatient settings are 43% higher than in community settings.

6.2.2 CGAS

Age

Omitting ratings where the recorded age was less than 3 and more than 20 years, the mean CGAS score of the remaining 5,140 assessments was 58.0 (SD = 14.4). Scores ranged from 0 to 100, with a lower score indicating lower functioning. The correlation with age was -.053, which represents a very small effect.

Sex

There are 3,155 CGAS assessments of males and 2,003 of females. Most of these (95%) were made in the community. In inpatient settings, the average male score (49.0) was not significantly higher than the average female score (46.0). This difference is reversed in the community, where females (59.1) scored significantly higher than males (57.7).

Diagnosis

Most (96%) of CGAS assessments were conducted in the community; the only diagnoses for which reasonable numbers of HoNOSCA assessments were obtained in inpatient settings were Schizophrenia and Mood Disorders. There were too few instances of diagnoses of Organic, Somatoform, Physiological and Personality Disorder to allow for meaningful analysis. Table 42 shows the mean CGAS total broken down by setting and diagnosis.

Table 42. Mean CGAS total score by setting and diagnosis.

	Inpatient	Community	All
Retardation		43.5	43.5
Schizophrenia	45.0	50.4	48.7
Developmental		52.1	52.0
Childhood		57.3	57.2
Mood disorders	48.4	58.8	58.0
Stress		58.1	58.1
Obsessive-compulsive		59.2	59.2
Substance misuse		59.8	59.9
Anxiety		60.4	60.4
Eating		64.9	61.8

The mean CGAS scores form three distinct clusters. Those associated with the diagnoses Substance Misuse, Mood, Anxiety, Obsessive-Compulsive, Stress, Eating and Childhood are all in a narrow range between 57 and 62. Schizophrenia and Developmental are around 50, and Retardation is much lower (worse), at 43.5.

6.2.3 FIHS

Age

The total score of the FIHS is simply the number of items that were answered in the positive direction; since there are seven items, FIHS total scores range from 0 to 7. Considering only consumers between the ages of 3 and 20 years, the correlation between the FIHS total score and age is .04, indicating a very small effect of age. Table 43 shows that the effect of age on each of the FIHS items is small or very small.

Table 43. Mean ages associated with FIHS items, and correlations with age.

FIHS item	FIHS rating		Correlation	Qualitative description of the association[20]
	Yes	No		
1 Maltreatment syndromes	12.9	12.6	.02	Very small
2 Problems related to negative life events in childhood	12.7	12.7	.01	Very small
3 Problems related to upbringing	12.5	12.9	-.05	Small
4 Problems related to primary support group, including family circumstances	12.6	12.8	-.02	Very small
5 Problems related to social environment	13.0	12.5	.06	Small
6 Problems related to certain psychosocial circumstances	13.3	12.4	.12	Small
7 Problems related to other psychosocial circumstances	13.4	12.6	.09	Small

[20] See footnote 16.

Sex

Of the 2,786 valid FIHS assessments, 2,693 (97%) were conducted in the community, leaving too few inpatient assessments to analyse. Of the 2,693 assessment in the community, 62% were male and 38% were female, and in one case, sex was not recorded. Table 44 presents the percentages of males and females for whom the FIHS items were rated as present.

Table 44. Percentages of present ratings of FIHS items by sex.

	Males	Females	Correlation	Qualitative description of the association[21]
1 Maltreatment syndromes	20.5	25.2	-.05	Very small
2 Problems related to negative life events in childhood	53.2	63.4	-.10	Small
3 Problems related to upbringing	54.6	61.0	-.06	Very small
4 Problems related to primary support group, including family circumstances	66.8	73.2	-.07	Very small
5 Problems related to social environment	47.3	49.4	-.02	Very small
6 Problems related to certain psychosocial circumstances	33.0	39.2	-.06	Very small
7 Problems related to other psychosocial circumstances	16.3	15.6	.01	Very small

On six of the seven FIHS items, more females were rated positive than males, but differences in most cases were very small. A somewhat larger effect was present in item 2 (Problems related to negative life events in childhood) on which 10% more females than males were rated positive. Item 4 (Problems related to primary support group, including family circumstances) was the most often rated problem, and item 7 (Problems related to other psychosocial circumstances) the least. The total FIHS score of the females (3.2) was significantly higher than that of the males (2.8) ($t_{(2690)} = 4.4$, $p < .001$). However, with the real difference being just 0.4 on a 7 point scale, this difference is not meaningful.

Diagnosis

Table 45 shows the mean FIHS total scores broken down by setting and diagnosis. As with the HoNOSCA and the CGAS, there were only sufficient numbers of assessments associated with Schizophrenia and Mood Disorders to allow for meaningful analysis. The highest FIHS scores are associated with Substance Misuse and Stress and the lowest with Anxiety, Obsessive-Compulsive, Eating and Developmental diagnoses. Schizophrenia, Mood, Retardation and Childhood diagnoses are all associated with mean FIHS scores of about 3.

[21] See footnote 16.

Table 45. Mean FIHS total score by setting and diagnosis.

	Inpatient	Community	All
Substance misuse		4.1	4.1
Stress		3.6	3.6
Childhood		3.0	3.1
Mood disorders	4.2	2.9	3.0
Retardation		3.0	3.0
Schizophrenia	3.3	2.5	2.8
Anxiety		2.2	2.2
Developmental		2.0	2.1
Obsessive-compulsive		2.0	2.0
Eating		1.8	1.9

6.3 Summary of the relationship between the outcome measures and other clinical and socio-demographic variables

Some HoNOS items were associated with younger age, while others were associated with older age. However, the total score on the HoNOS was unrelated to age.

There were differences between males and females on all but one HoNOS item and also at the subscale level. But none of the differences between the sexes seem to be clinically important.

The three diagnostic categories associated with the highest HoNOS total scores were Substance misuse, Personality disorder, and Organic disorders. There was little difference in HoNOS total scores obtained in the community between Obsessive-compulsive, Stress, Eating, Mood and Anxiety disorders. It was not possible to test the effect of co-morbidities.

In both inpatient and community settings, involuntary status was associated with higher scores on the Behaviour, Impairment, Social subscales and the total score. High scores on the Hallucinations/delusions item were associated with involuntary status, while high scores on the depression related items were more associated with voluntary status.

Correlations between age and LSP-16 items, subscales and total score were small to very small, with one exception (Item 16: What sort of work is this person generally capable of?). Males had higher (worse) ratings on 15 of the 16 items, all four subscales, and on the total score. However, the actual difference in scores was very small for all 16 items.

Consumers with Organic, Substance misuse, Schizophrenia and Personality Disorders had relatively high scores on the LSP-16, while consumers with Mood, Anxiety, Obsessive-Compulsive, Stress and Eating Disorders on average had lower scores. Involuntary legal status was consistently associated with higher scores on all four LSP-16 subscales and the total score.

On the HoNOSCA, certain items were associated with younger age, while others were associated with older age, but the total score was unrelated to age. On average, males obtained total scores about one point higher than females. The highest HoNOSCA total scores were associated with diagnoses of Retardation, Schizophrenia, Developmental and Substance Misuse, and the lowest with Eating Disorders.

Scores on the CGAS were largely unrelated with age. Males tended to obtain higher (better) ratings in inpatient settings, and females obtained higher ratings in community settings. The lowest ratings were associated with diagnoses of Retardation, followed by diagnoses of Schizophrenia and Developmental disorder.

As with the other child and youth measures, there was a negligible correlation between the FIHS total score and age. Item 4 (Problems related to primary support group, including family circumstances) was the most frequently identified problem. Six of the seven FIHS items were more commonly identified in females than males. The highest scores were associated with Substance misuse and Stress disorders.

7. Results: do consumers change over the course of an episode of mental health care?

For the purposes of assessing change, the assessment pairs reported in Table 10 were used. Using only these paired assessments ensures that the comparison is 'like with like'. Thus, this chapter looks at those consumers who had a rating at both the beginning and the end of an episode of care and measures the differences between the two ratings.

As with other chapters, results are described in this chapter as being either 'statistically significant' or not. However, it is important to emphasise again that, because of the large volume of ratings, almost all results are statistically significant. Statistical significance must be distinguished from the concept of clinical meaning or importance. As with other chapters, some results presented in this chapter are statistically significant but lack clinical meaning.

7.1 Adults

7.1.1 The HONOS

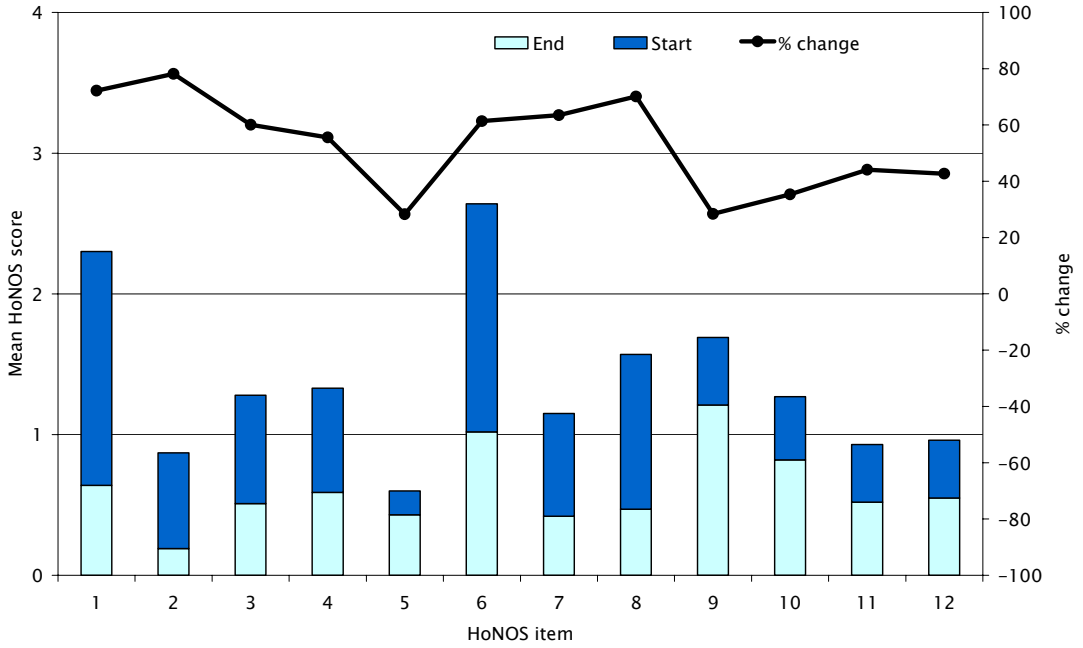
Inpatient episodes

In inpatient settings, the intervals between the start and end HoNOS assessments ranged from 1 to 157 days. Episodes were only included in the following analysis where the episode lasted more than three days and there was a maximum time of 31 days between assessments. In total, 611 episodes met these criteria. Of these, the majority (443) began with an actual admission and ended with an actual discharge, 58 began with an admission and ended with a review, 97 began with a review and ended with a discharge and, for just 13 episodes both assessments were reviews. The mean interval between assessments was 15.2 days (range 4 to 31).

The mean HoNOS total score at the start of these episodes was 14.5, much higher than the mean of 7.2 at the end. This almost 50% difference is statistically significant. Examination of changes in the subscales scores revealed that the greatest changes were in the Behaviour (66%) and Symptoms subscales (59%), with lesser changes in the Impairment (37%) and Social subscales (33%).

The two most numerous diagnostic categories for which relevant data were available are schizophrenia and mood disorders. The next analysis looks at changes at the HoNOS item level between the starts and ends of episodes. Figure 16 shows the admission and discharge HoNOS item scores for episodes where the principal diagnosis was schizophrenia. The numbers of data points vary somewhat from item to item, between a low of 302 for item 11 (Occupation problems) and a high of 372 (Aggressive/overactive/agitated behaviour). The top of each dark vertical bar represents the mean admission score, and the top of the light part of the bar is the mean discharge score. The line indicates the percentage change, the values of which are on the right axis. For example, the admission score for item 1 was 2.2 and the discharge score was 0.7, a change of 67%.

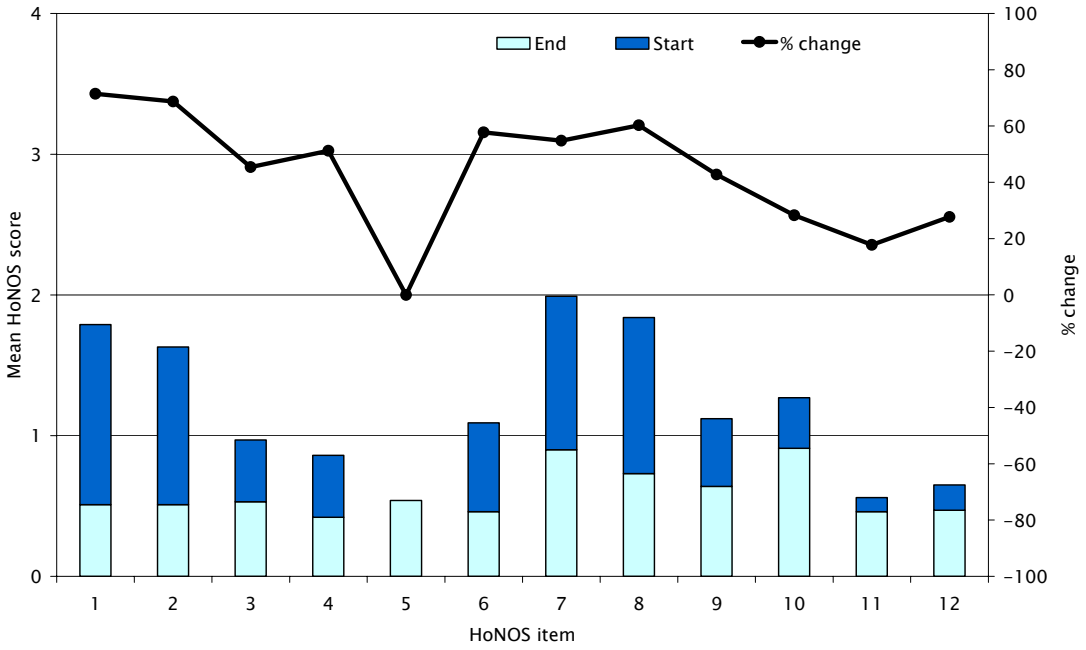
Figure 16. Mean admission and discharge HoNOS scores for episodes of schizophrenia – inpatient settings.



For inpatient episodes of schizophrenia, the greatest problems at initial assessment were aggressive/overactive/agitated behaviour and hallucinations and delusions. Substantial changes are observed on all items, but the greatest (over 50%) were on all of the items comprising the Behaviour subscale (items 1, 2 and 3) and all of the items comprising the Symptoms subscale (items 6, 7 and 8). The smallest changes, at 22%, is seen on item 9 (Relationship problems), which is the third highest at episode start.

Figure 17 shows the same results for those with mood disorders. The highest initial ratings are on items 7 (Depressed mood) and 8 (Other Behavioural or psychological problems), closely followed by items 1 (Aggressive/overactive/agitated behaviour) and 2 (Non-accidental self-injury). Five items (1, 2, 6, 7 and 8) show changes of 50% or more. Very little change (about 5%) was detected on item 5 (Physical illness or disability problems).

Figure 17. Mean admission and discharge HoNOS scores for episodes of mood disorder – inpatient settings.



Community episodes

In community settings, the intervals between the start and end HoNOS assessments ranged from 1 to 180 days. Episodes were only included in the following analysis if they had an episode length greater than three days and a maximum of 91 days between ratings. This corresponds to the cycle of assessment adopted in the study. In total, 3,681 episodes met these criteria. Of these, the majority (2,489) began and ended with a review assessment, 655 began as a new case and ended with a review, 254 began with a review and ended with a case closure. For 283 episodes, the initial assessment was at case opening and the final assessment at case closure. The mean interval between assessments was 69.3 days (range 4 to 91).

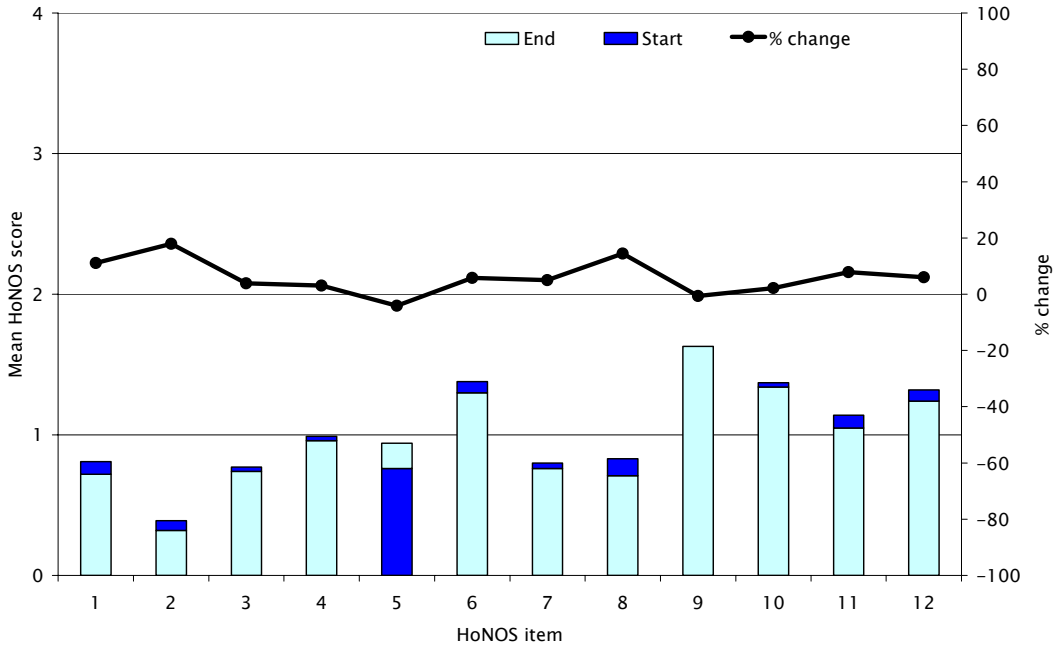
The mean HoNOS total score at the start of these episodes was 11.5, somewhat higher than the mean of 10.2 at the end. This difference of about 11% is highly statistically significant because of the large numbers involved. Examination of changes in the subscales scores revealed that, as with the inpatient episodes, the greatest changes were in the Behaviour (16%) and Symptoms subscales (15%), with lesser changes in the Impairment (8%) and Social subscales (9%).

One of the reasons for community episodes ended was because the consumer was admitted to a psychiatric unit. There were 163 instances of these where the initial assessment had been a review. For these episodes, which averaged 42.6 days, the start and end HoNOS total scores were 13.8 and 18.7 respectively, a significant worsening of about 36%.

The next analysis looks at changes at the HoNOS item level for the two most numerous diagnostic groups. Figure 18 shows the start and end HoNOS item scores for episodes where the principal diagnosis was schizophrenia.

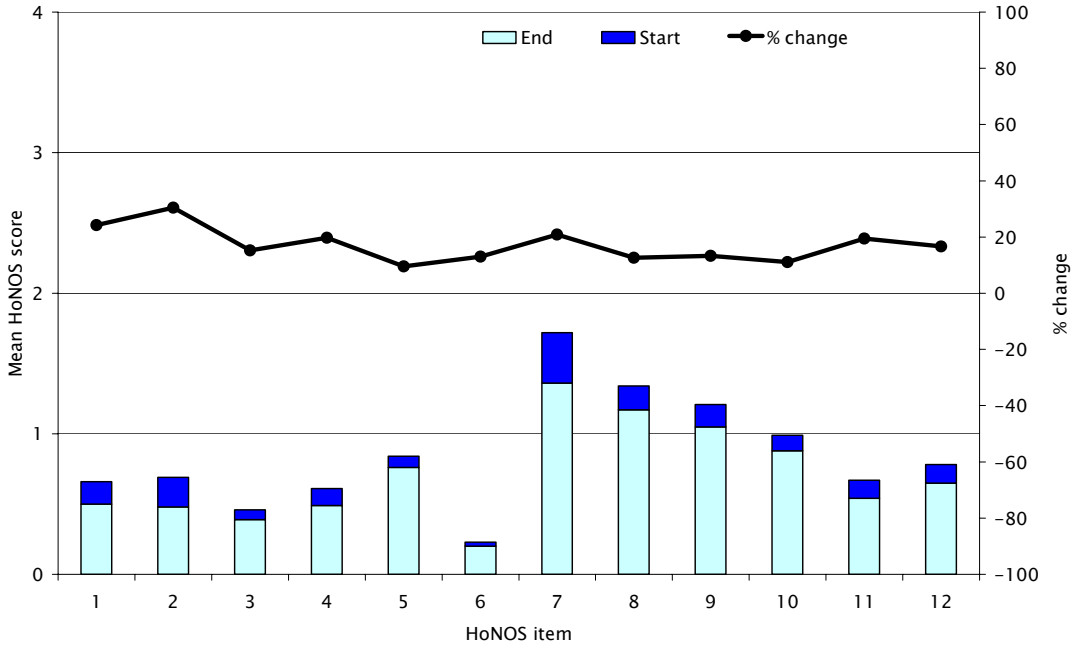
As before, the top of each dark vertical bar in Figure 18 represents the mean score at episode start and the top of the pale part of the bar is the mean discharge score. The line indicates the percentage change, the values of which are on the right axis. For eight of the twelve items the changes are small – 8% or less. In the case of item 5 (Physical illness and disability), the mean end of episode rating is actually higher (worse) than the initial ratings. Item 9 (Relationship problems) is the worst at the start of episodes and remains unchanged at the ends of episodes. Clearly, the magnitude of change is less than in the inpatient setting.

Figure 18. Mean admission and discharge HoNOS scores for episodes of schizophrenia – community settings.



The next analysis looks at the start and end HoNOS scores of episodes of mood disorder in the community. This is shown in Figure 19. Once again, only episodes that lasted between three and 91 days are included. The number of matched start and end assessments available for analysis ranged between 1,400 and 1,433, depending on the HoNOS item. The highest rating at the start of these episodes was Item 7 (Depressed mood). The "% change" profile shows that the greatest proportional change was seen on item 2 (Non-accidental self-injury). While all changes are for the better, the magnitude of change is moderate, being of the order of 10% to 20%.

Figure 19. Mean admission and discharge HoNOS scores for episodes of mood disorder – community settings.



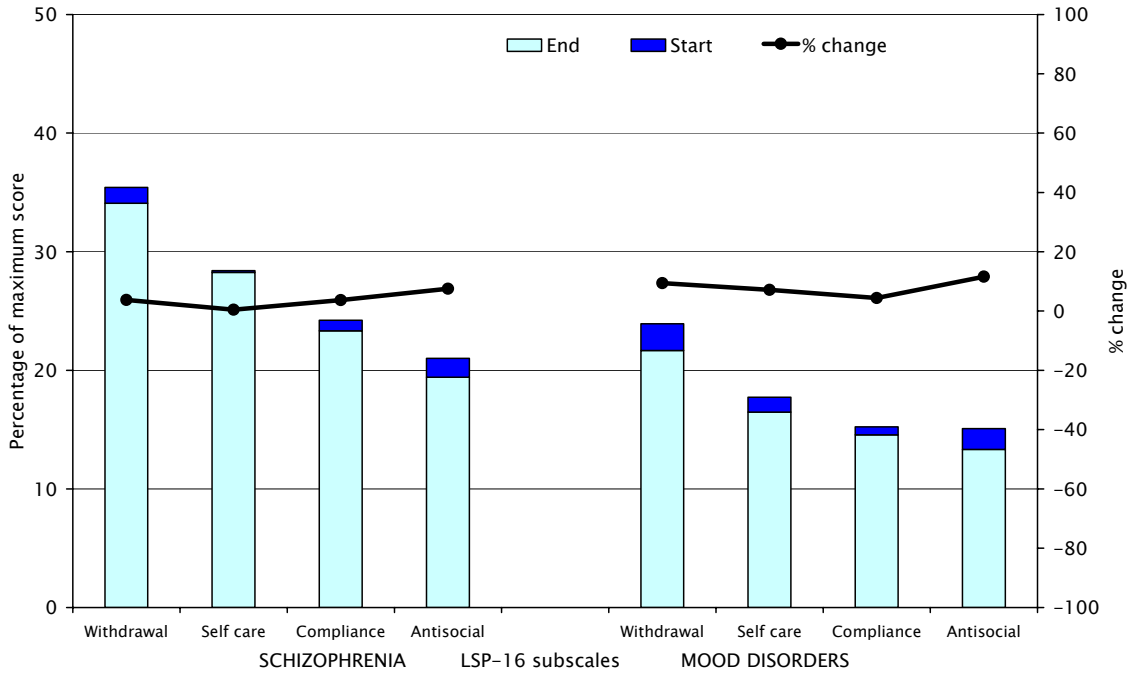
7.1.2 LSP-16

There were too few instances of matched start and end LSP-16 assessments in inpatient settings for analysis. In community settings, the same rules used in the HoNOS analysis were used to create a data set for analysis (see Section 7.1.1). In total, 4,438 episodes met the criteria and were included in the analysis. Of these, the majority (3,125) began and ended with a review assessment, 732 began with a case opening and ended with a review, 272 began with a review and ended with a case closure. For 309 episodes the initial assessment was a case opening and the last was at case closure. The mean interval between assessments was 71.1 days (range 4 to 91).

The mean LSP-16 total score at the start of these episodes was 11.2, somewhat higher than the mean of 10.5 at the end. The difference of about 6% is statistically significant because of the large numbers involved.

As with the HoNOS, the two most numerous diagnostic categories for which relevant data were available are schizophrenia and mood disorders. The next analysis looks at changes in the LSP-16 subscale scores between the initial and final assessments. Figure 20 shows the start and end LSP-16 subscale scores for episodes where the principal diagnosis was schizophrenia (on the left) and mood disorders (on the right). The number of paired assessments of the former was 1,777 and of the latter 1,620. As with Figure 19, the top of each dark vertical bar represents the mean score at episode start and the top of the pale part of the bar is the mean score at episode end. The line indicates the percentage change, the values of which are on the right axis. Despite starting with higher levels of disability on all the LSP-16 subscales, consumers with schizophrenia made small improvements, of the order of 0% to 7%, while those with mood disorders began with lower level of disability but made slightly larger improvements, of the order of 4% to 12%.

Figure 20. Mean admission and discharge LSP-16 subscale scores for episodes of schizophrenia and mood disorder – community settings.



7.1.3 Adult change scores by Focus of Care

Section 5.1 examined the differences in HoNOS and LSP-16 scores at episode start between the Focus of Care (FoC) categories. In this section we examine whether there are differences between the FoC categories in relation to the amount of change achieved over the course of an episode.

Figure 21. Mean HoNOS total score, by Focus of Care and episode type.

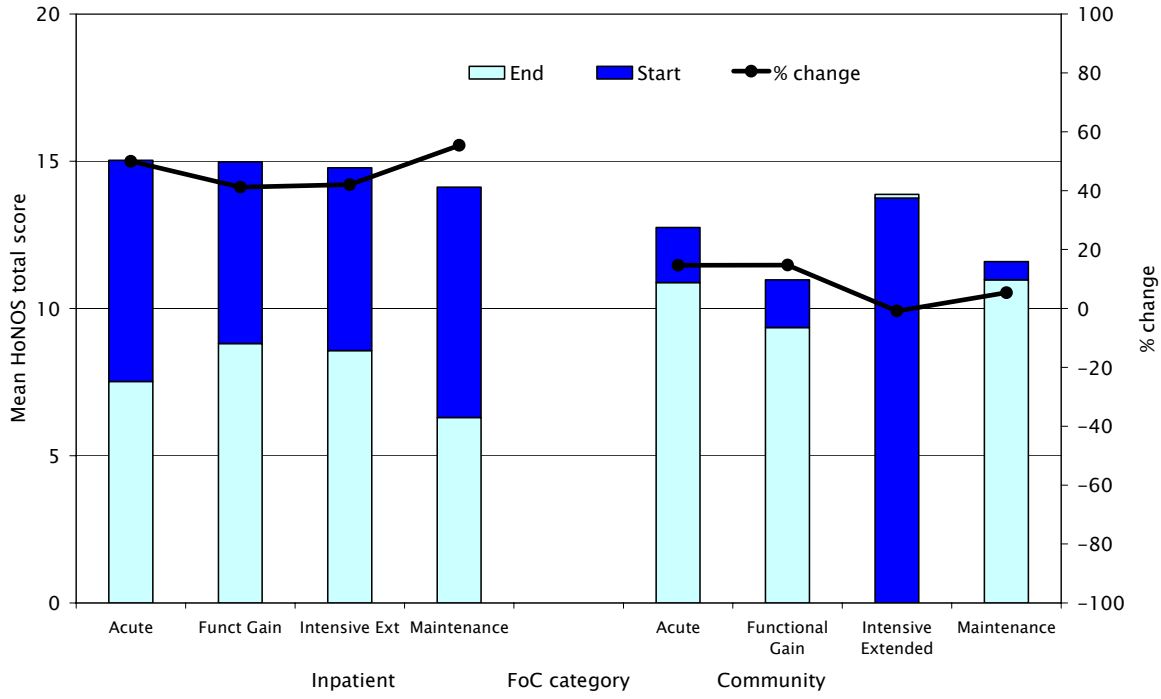
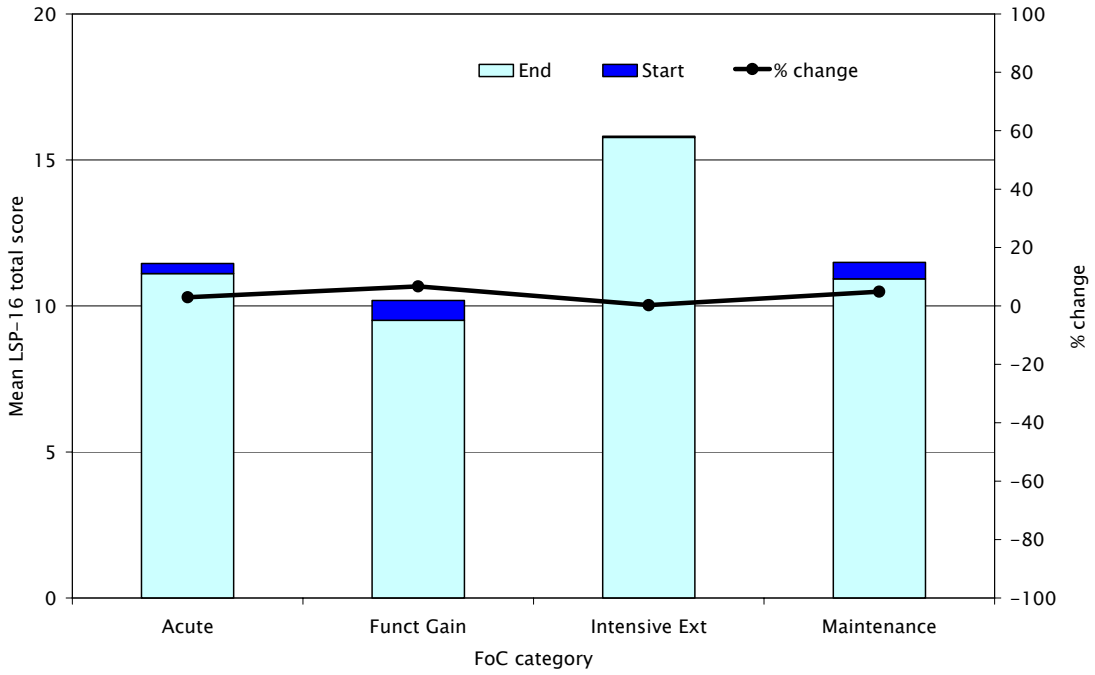


Figure 21 shows the mean changes in HoNOS total score, broken down by FoC category, separately for inpatient and community episodes. For inpatient episodes, substantial improvements on HoNOS total scores, of the order of 50%, were observed between start and end of inpatient episodes in all four FoC categories. Raw differences were between 6.2 and 7.8 points and not statistically significant. Within each FoC category, the greatest improvements were seen on the Behaviour and Symptoms subscales and the least on the Impairment and Social subscales.

Somewhat more separation in HoNOS total scores between the FoC categories was found in the community episodes. On average, there were improvements of around 15% in the Acute and Functional Gain categories, about 5% for Maintenance, and no real change in the Intensive Extended category. While these changes are not statistically significant, they were nearly so ($F_{(3,613)} = 2.29$, $p = .077$). For the Acute FoC, changes were between 11% and 17% across the four HoNOS subscales, while for the Functional Gain FoC, changes on Behaviour and Symptoms were around 20% and on Impairment and Social, around 10%.

This same analysis was undertaken using the LSP-16. There were too few LSP-16 assessments conducted in inpatient settings, so we examine only those performed in community settings. Figure 22 shows the mean changes in LSP total score, broken down by FoC category, in community episodes

Figure 22. Mean LSP-16 total score, by Focus of Care category – community episodes only.



In the community, small changes, of the order of 3% to 7%, are seen in episodes of the Acute, Functional Gain and Maintenance type, and there was virtually no mean change in Intensive Extended episodes. The differences in LSP-16 total scores between the FoC categories is just significant ($F_{(3,5023)} = 2.63$, $p = .048$). For episodes of the Acute and Functional Gain types, the greatest changes were seen on the Withdrawal and Antisocial subscales. As with the HoNOS, the Intensive Extended FoC showed an inconsistent pattern of very small changes close to zero. The Maintenance FoC showed changes of between 3% and 9% on the four LSP-16 subscales.

7.2 Child and Youth

7.2.1 HoNOSCA

There were too few instances of matched start and end HoNOSCA assessments in inpatient settings for analysis. In community settings, the intervals between the start and end HoNOSCA assessments ranged from 1 to 180 days and the same criteria for including episodes in the analysis were applied (only episodes that lasted more than three days and that had a maximum time of 91 days between assessments). The number of pairs of assessments meeting these criteria was 896; of these, 315 began and ended with a review assessment, 373 began with a case opening and ended with a review and 110 began with a review and ended with a case closure. For 98 episodes, the initial assessment was a case opening and the last was a case closure. The mean interval between assessments was 68.0 days (range 4 to 91).

The mean HoNOSCA total score at the start of these episodes was 13.2, somewhat higher than the mean of 10.9 at the end. This 17% difference is statistically significant. Examination of changes in the subscales scores revealed that the greatest changes were in the Behaviour (21%) with somewhat smaller changes Impairment and Symptoms subscales (both 17%) and the Social subscale (15%).

7.2.2 CGAS

There were too few instances of matched start and end CGAS assessments in inpatient settings for analysis. The number of pairs of community assessments that met the criteria described above was 1,073. Of these, 405 began and ended with a review, 430 began with a case opening and ended with a review, 120 began with a review and ended with a case closure. For 118 episodes, the initial assessment was a case opening and the last was a case closure. The mean interval between assessments was 68.5 days (range 4 to 91).

The mean CGAS rating at the start of these episodes was 55.7, which is five points lower (worse) than the mean of 60.7 at the end. The difference of about 9% is statistically significant.

7.2.3 FIHS

As may be seen from Table 10, there were only 268 instances in the data of FIHS assessments at both the start and end of episodes. Under the study data collection protocol, the FIHS was not required to be collected at the start of episodes as it is not, strictly speaking, an outcome measure.

7.3 Summary of consumer change over the course of an episode of mental health care

In inpatient settings, the HoNOS reflected large changes, of the order of 50% of the initial total score, between the start and the end of episodes. Smaller changes, around 11%, were found in the community. While changes were observed in most items, the greatest improvements occurred on the Behaviour and Symptoms subscales. In inpatient settings, the greatest changes were generally in those items that were highest at the start of the episode, but no such pattern was clearly discernible in community episodes. This is probably because ratings in the community were generally much lower than those reported in inpatient settings.

Overall changes between start and end of episodes in the community on the LSP-16 were smaller than on the HoNOS – of the order of 6%. Comparing the two largest diagnostic

groupings, those with schizophrenia made smaller changes than those with mood disorders, despite starting with higher levels of disability.

For inpatient episodes, there were no significant differences between the four FoC categories in relation to improvement between the start and end of episodes. For all four FoC categories, the greatest improvements were seen on the Behaviour and Symptoms subscales and the least on the Impairment and Social subscales. The FoC categories were better differentiated in the community but, again, the differences between the four categories in the amount of change was not statistically significant. However, on average, there were improvements of around 15% in the Acute and Functional Gain categories, about 5% for Maintenance, and no real change in the Intensive Extended category.

The differences in LSP-16 total scores between the FoC categories just reached statistical significance. For Acute and Functional Gain episodes, the greatest changes were seen on the Withdrawal and Antisocial subscales. As with the HoNOS, the Intensive Extended FoC showed an inconsistent pattern of very small changes close to zero. The Maintenance FoC showed changes of between 3% and 9% on the four LSP-16 subscales.

HoNOSCA assessments in the community indicated an overall change of just over two points, or about 17%, with the greatest change observed on the Behaviour subscale (21%). On average, CGAS assessments in the community improved five points, or 9%, between starts and ends of episodes.

8. Results: comparison between outcomes for Māori, Pacific and Other Consumers

The analysis in this section on 'ethnicity' uses 'ethnicity' data as collected in the study. These were collected through a combination of records supplied by the participating sites and, in those instances where this field was left blank by the site, by NZHIS. The study asked for only one consumer self-identified ethnic group to be recorded on the study episode forms, unlike NZHIS where a maximum of three ethnic groups can be stored on the MHINC database. If a consumer identifies with more than three ethnic groups, then a prioritisation algorithm for ethnicity codes is used by NZHIS to reduce the ethnic groups to a maximum of three in ranked order. The decision by NZ-CAOS to combine data received from both DHB sites and the NZHIS in order to ensure a high completion rate means that this prioritisation algorithm was not applied to all the data. For the purposes of the following analysis, ethnicity was grouped up into three high level groupings – Māori, Pacific Island, with all other ethnicity classifications grouped together to form an 'All Other' grouping.

As with other chapters, results are described in this chapter as being either 'statistically significant' or not. However, it is important to emphasise once more that, because of the large volume of ratings, almost all results are statistically significant. Statistical significance must be distinguished from the concept of clinical meaning or importance. Some results presented in this chapter are statistically significant but lack clinical meaning.

8.1 Adults

8.1.1 HoNOS profile

Figure 23 shows the HoNOS item scores in inpatient settings for the three high level ethnicity groupings. This figure is based on the mean HoNOS item scores of approximately 1,250 assessments of Māori, 400 Pacific Island, and just under 3,000 for the 'All Other' grouping (the numbers vary slightly from item to item on account of missing ratings). The differences between the three groups are statistically significant for all items apart from Item 4 (cognitive problems) and Item 10 (activities of daily living).

Figure 23. Mean HoNOS item scores in acute inpatient settings, by ethnicity grouping.

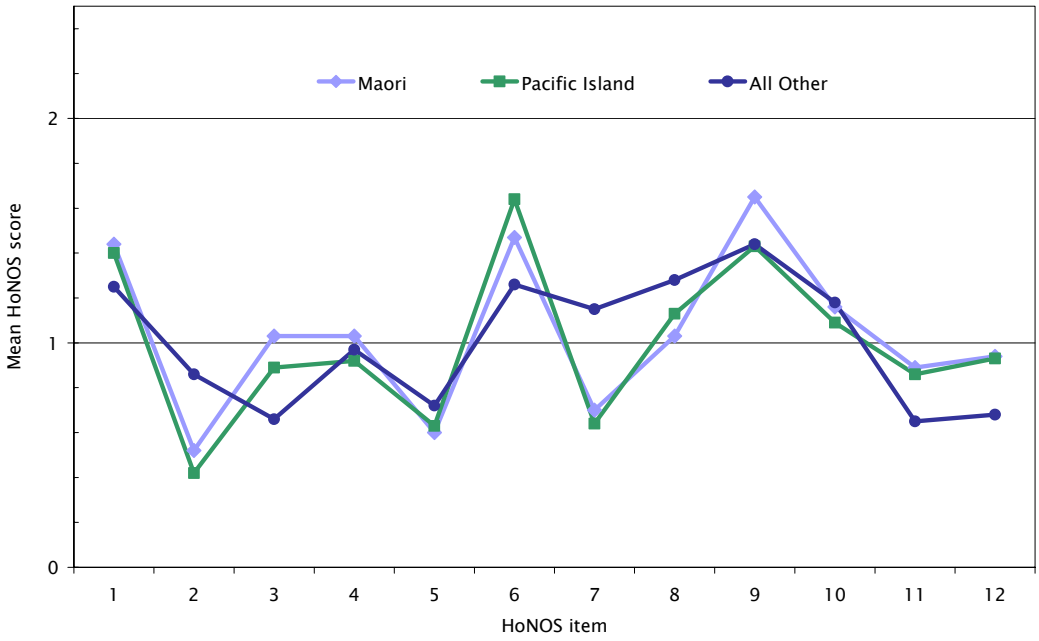
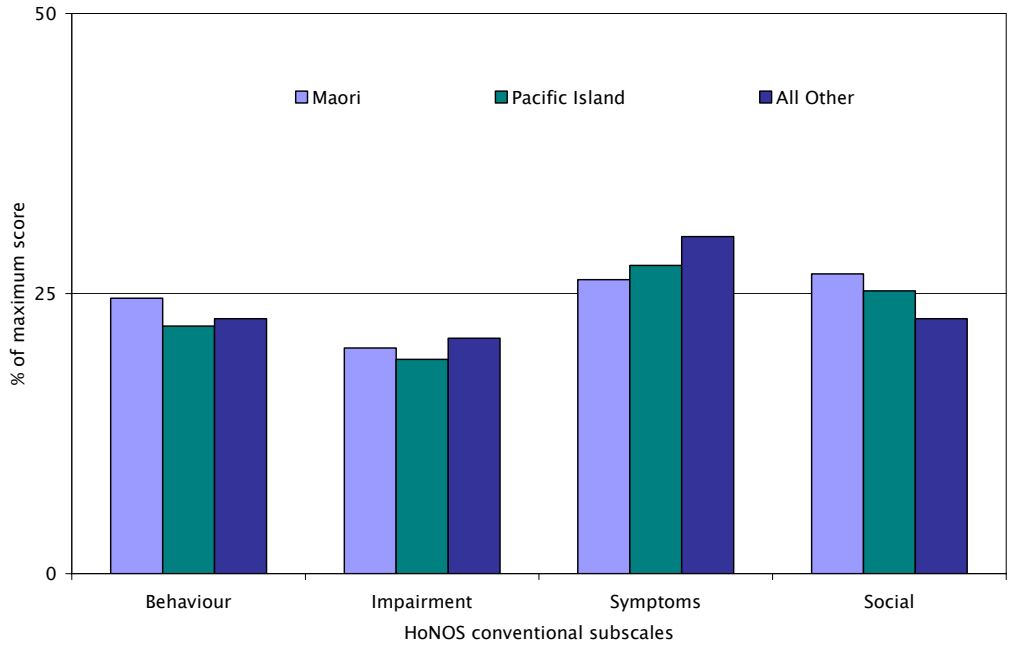


Figure 24. Mean HoNOS subscales scores in acute inpatient settings, by ethnicity grouping.

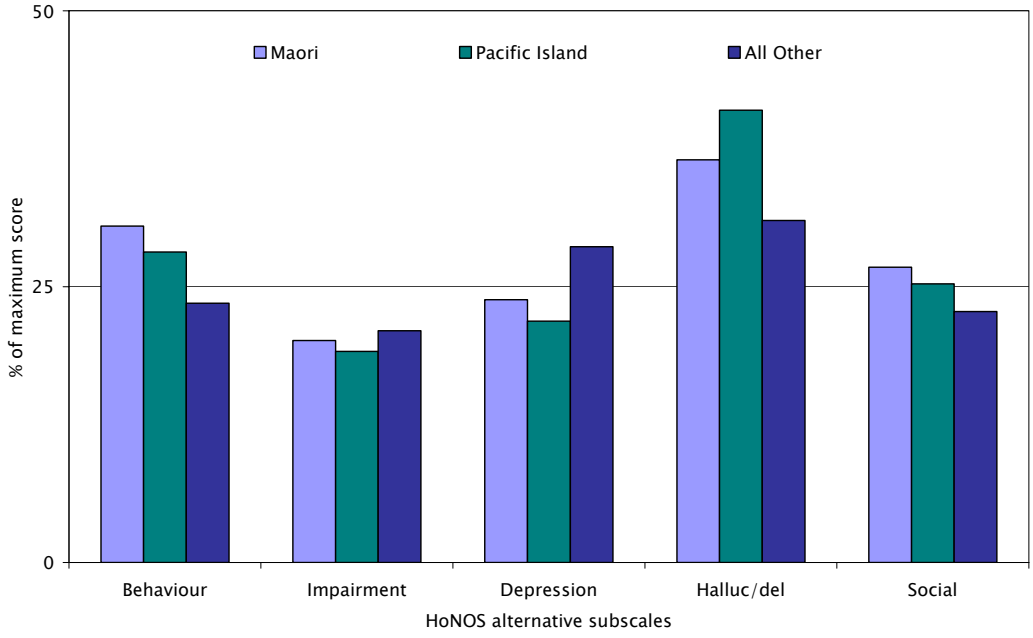


Differences between the groups become clearer by examining the subscales scores, which are presented in Figure 24. It shows that, on the conventional HoNOS subscales, Māori are rated higher than Pacific Island and All Others on the Behaviour subscale. The differences on the Impairment subscale are small and non-significant. The 'All Other' group score higher on the Symptom subscale but lower on the Social subscale than the Māori and Pacific Island groups.

An alternative set of subscales proposed by Trauer in 1999 was included in the technical analysis in Chapter 4. Figure 25 shows the differences between the three ethnicity groupings on these alternative subscales.

Looking only at the subscales that are different from the conventional set, Māori and Pacific Island assessments are significantly higher on the Behaviour subscale. Māori and Pacific Island assessments are significantly lower on Depression but higher on Hallucinations/delusions than the 'All Other' grouping.

Figure 25. Mean HoNOS alternative subscales scores in acute inpatient settings, by ethnicity grouping



Turning now to assessments in the community, Figure 26 shows the HoNOS items for the three high level ethnicity groupings. This figure shows the mean HoNOS item scores of approximately 2,800 assessments of Māori, 770 Pacific Island, and about 12,200 All Others. Differences are generally more marked than in inpatient settings and the differences between the three groups on all 12 items are statistically significant ($p < .0001$). The Māori and Pacific Island groups scored higher than the All Other grouping on items 1, 3, 6, 10, 11 and 12, while the reverse was true for items 7 and 8.

Figure 26. Mean HoNOS item scores in community settings, by ethnicity grouping.

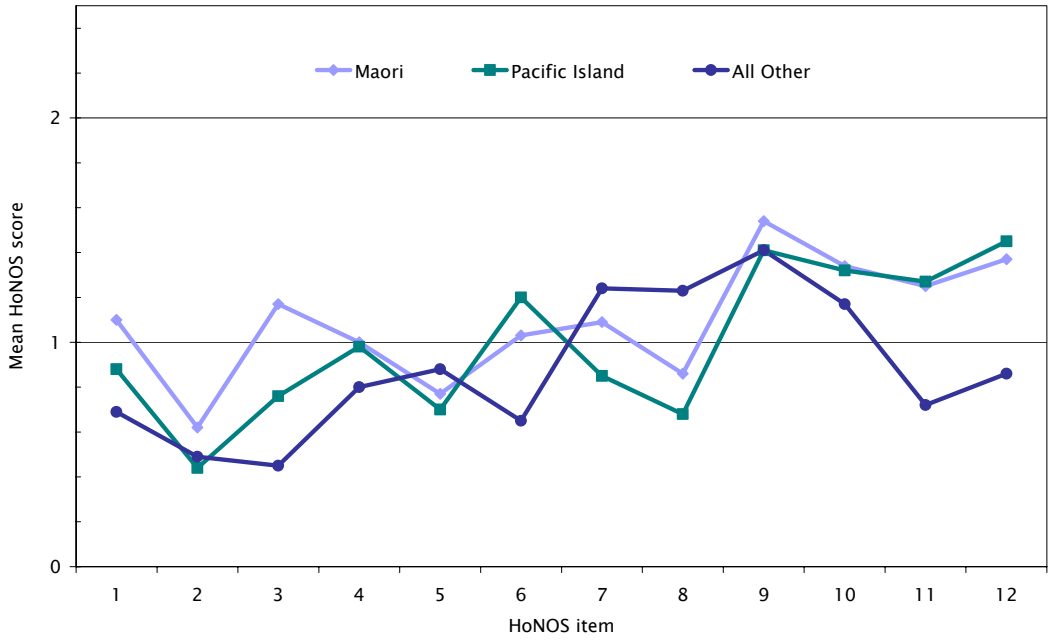


Figure 27. Mean HoNOS conventional subscale scores in community settings, by ethnicity grouping.

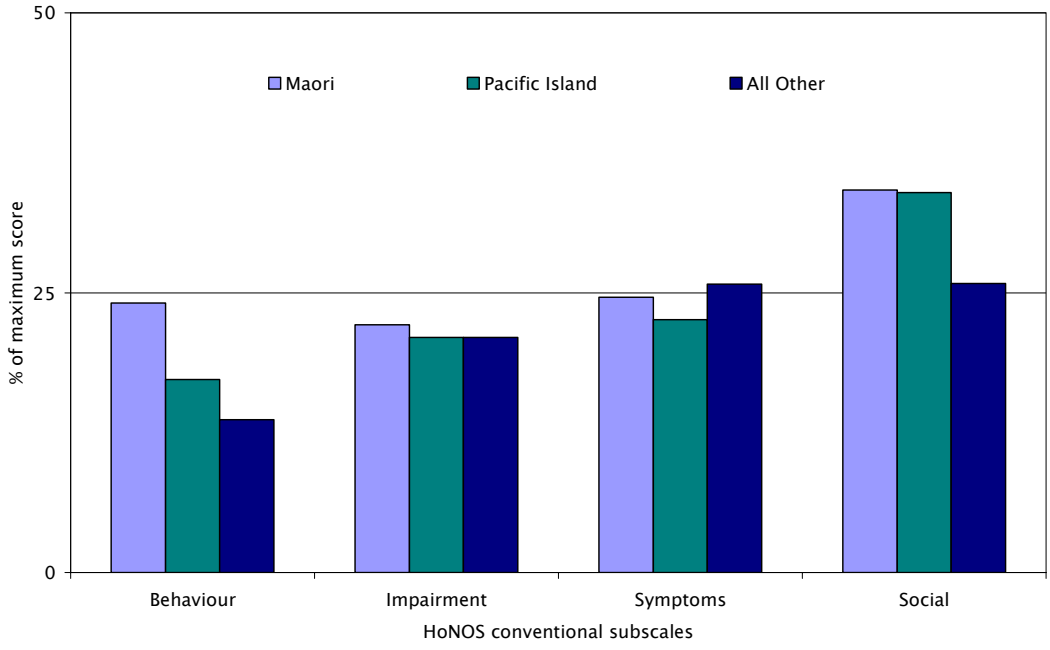
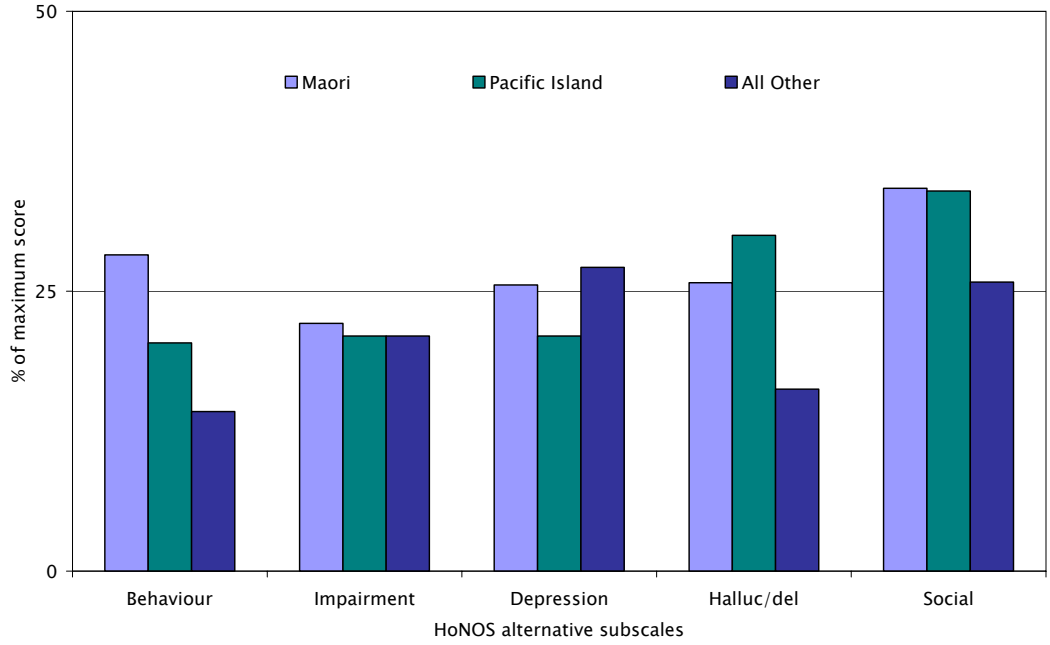


Figure 27 above shows the conventional subscale scores. Māori assessments in the community setting had higher Behaviour subscale scores than those of Pacific Island, which are, in turn, higher than those of the All Other grouping. Differences on the Impairment subscale were small and non-significant. There were inconsistent differences on the Symptoms subscale. Māori and Pacific Island scores on the Social subscale were higher than those of the All Other grouping.

Once again, the alternative set of subscales was included in the analysis. The results are shown in Figure 28. Looking only at the subscales that are different from the conventional, the differences between the ethnic groups are more clearly delineated. The Māori and All Other groups score significantly higher on the Depression subscale than the Pacific Islander grouping and the Māori and Pacific Island groups score significantly higher than the All Other grouping on Hallucinations/delusions.

Figure 28. Mean HoNOS alternative subscales scores in community settings, by ethnicity grouping.



8.1.2 Change as measured by the HoNOS

Of the 7,805 pairs of HoNOS assessments (see Table 10) 1,578 (20%) were of Māori, 432 (6%) of Pacific Island, 5,739 (74%) All Other. No ethnicity was recorded for 56 episodes. These numbers include both inpatient and community episodes, each of which are separately reported below.

Inpatient episodes

This section reports similar analyses to those reported in Chapter 7, but this time broken down by ethnicity grouping. Table 46 shows that the change results are fairly similar for all three ethnicity groupings (although the number of Pacific Island assessments is quite small). Improvements of the order of 50% were achieved in episodes of approximately 15 days.

These results were further analysed using multiple regression. The purpose of the multiple regression was to analyse the relationship between independent or predictor variables (in this case, variables such as ethnicity grouping and start scores) and a dependent or criterion variable (in this case the difference between start and end HoNOS total scores). This analysis showed that greater improvements are significantly associated with higher start scores. Ethnicity grouping effects were not significant. In other words, the amount of change is largely explained by the score at the start of the episode. Once differences in starting scores are taken into account, there were no differences between the three ethnicity groupings in relation to the amount of change from episode start to episode end.

Table 46. HoNOS total scores at start and end of inpatient episodes by ethnicity grouping.

Ethnicity grouping	N	Days	Start	End	% change
Māori	160	15.5	16.7	7.8	53.5%
Pacific Island	35	15.3	14.8	7.2	51.2%
All Other	410	15.2	13.6	6.9	49.1%

Community episodes

The next analyses repeat the preceding analyses, but this time for community episodes. The same rules for including episodes in the analysis were applied. The results are presented in Table 47. It shows that the change results are fairly similar for all three ethnicity groupings. Improvements of the order of 7% to 12% were achieved in episodes of approximately 69 days.

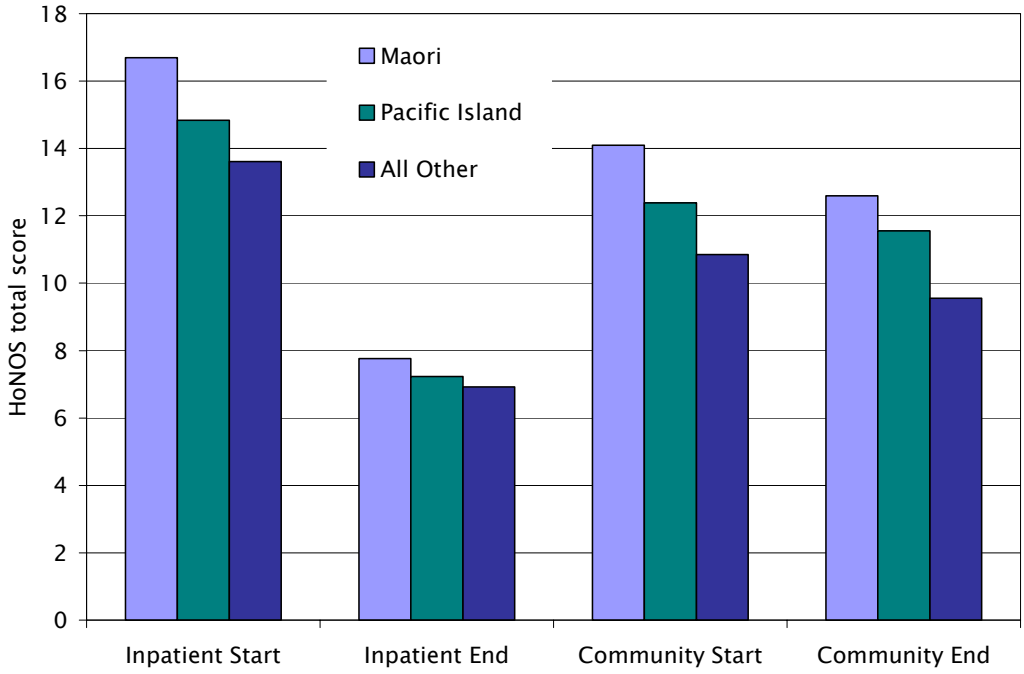
These results were further analysed using multiple regression (see previous section for an explanation of this technique). This analysis showed that greater improvements are again significantly associated with higher start scores. Ethnicity grouping effects just reach statistical significance once starting HoNOS scores are taken into account. The improvement (in raw HoNOS score) for Māori and Pacific Island is less than that for the 'All Other' grouping when their higher start scores are taken into account.

Table 47. HoNOS total scores at start and end of community episodes by ethnicity grouping.

Ethnicity grouping	N	Days	Start	End	% change
Māori	669	68.7	14.1	12.6	10.7%
Pacific Island	168	68.5	12.4	11.5	6.8%
All Other	2,822	69.5	10.8	9.5	12.0%

These results are summarised graphically in Figure 29. Māori consumers were rated higher (more severity) than Pacific Island at all points, and Pacific Island were similarly rated higher than All Others. The changes observed in inpatient settings are much larger than in the community and there were very similar rates of improvement.

Figure 29. Changes in HoNOS total scores between start and end of inpatient and community episodes, by ethnicity grouping.

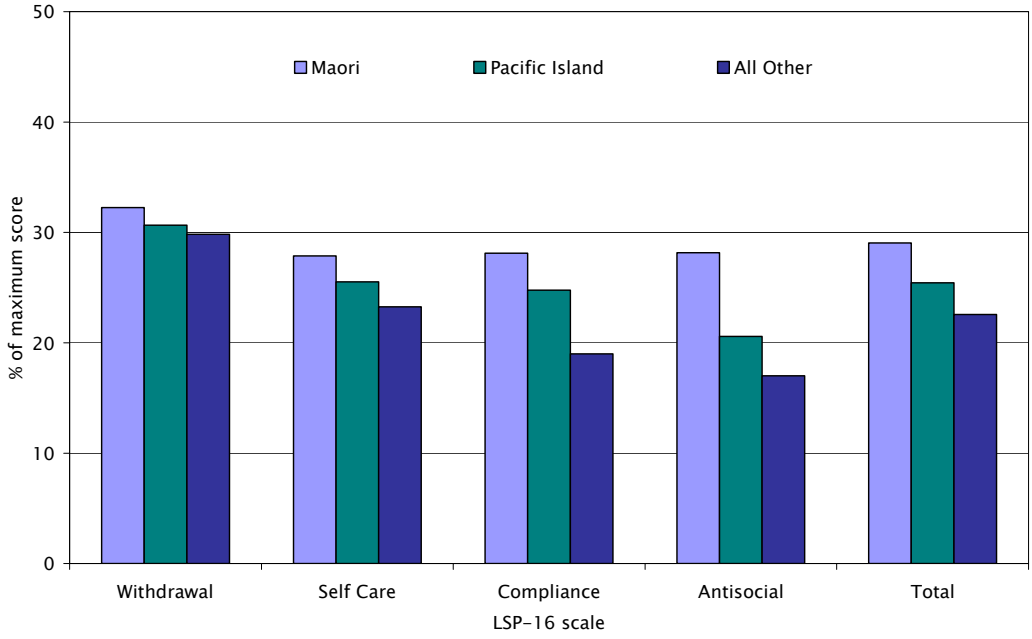


8.1.3 LSP-16 profile

LSP-16 total scores were available for 3,110 Māori, 934 Pacific Island, and 13,315 Other assessments. The mean scores were 13.95, 12.21 and 10.83 respectively. These differences are statistically significant ($F_{(2, 17356)} = 209.7$, $p < .001$). Pairwise comparisons found that the Māori mean was significantly higher than the Pacific Island mean which is, in turn, significantly higher than the All Other mean.

Figure 30 shows the mean LSP-16 subscales and total score by ethnicity grouping expressed as percentages of the maximum score for each subscale. The pattern (Māori higher than Pacific Island higher than All Other) observed in the total score is present in each of the four subscale scores. Pairwise comparisons showed that all three pairwise differences were significant for the Self Care, Compliance and Antisocial subscales. For the Withdrawal subscale, only the difference between Māori and All Other was significant. Effect sizes for the Withdrawal and Self Care subscales were very small and, for the Compliance and Antisocial subscales, were small.

Figure 30. LSP-16 subscales and total score by ethnicity grouping.



8.1.4 Change as measured by the LSP-16

Of the 7,140 pairs of LSP-16 assessments (see Table 10) all but 8 were obtained in community settings. Of these, 1,184 (17%) were of Māori, 390 (6%) of Pacific Island, 5,517 (77%) All Other, and for 41 episodes (0.6%) ethnicity grouping was not known.

The next analyses repeat the preceding analyses of the HoNOS data, but this time for the LSP-16 and community episodes. The same rules for including episodes in the analysis were applied. The results, presented in Table 48, are quite similar to the HoNOS (see Table 47). The pattern of scores is again Māori, followed by Pacific Island, followed by All Others. Changes are small, being of the order of one half to one point, and represent between 5% and 8% of initial scores.

Table 48. LSP-16 total scores at start and end of community episodes by ethnicity grouping.

Ethnicity grouping	N	Days	Start	End	% change
Māori	761	69.8	14.2	13.2	7.5%
Pacific Island	238	72.1	12.4	11.6	6.8%
All Other	3,413	71.3	10.4	9.8	5.5%

A multiple regression analysis similar to the ones in the previous section was performed. The difference between start and end LSP-16 total scores was predicted in terms of the start total score and ethnicity grouping. This analysis showed that greater improvements are significantly associated with higher start scores. There was no significant association between the amount of change and ethnicity. Any differences in change between the ethnicity groupings are largely attributable to differences in the start levels.

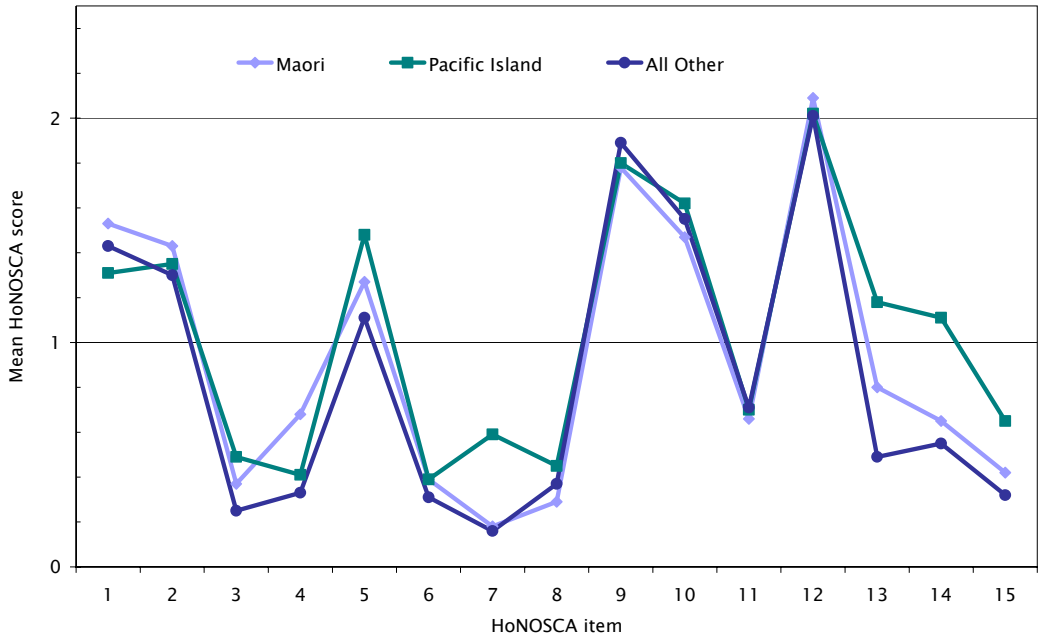
8.2 Child and Youth

This section repeats the same analyses, but this time for child and youth episodes.

8.2.1 HoNOSCA profile

Figure 31 shows the mean HoNOSCA item scores of 770 assessments of Māori, 113 Pacific Island, and about 3,850 Others. The biggest differences between the groups are on items 3 (Non-accidental self-injury), 4 (Alcohol, substance or solvent misuse), and 5 (Problems with scholastic or language skills). In both cases, Māori and Pacific Island consumers were rated higher than the All Other grouping. The Pacific Island grouping was rated highest on item 7 (Hallucinations, delusions or abnormal perceptions) and on items 13 (Poor school attendance), 14 (Lack of knowledge or understanding about the nature of the child or adolescent's difficulties), and 15 (Lack of information about services or management of the child or adolescent's difficulties), on which the Pacific Island group was rated highest, followed by the Māori, followed by All Others.

Figure 31. Mean HoNOSCA item scores in community settings, by ethnicity grouping.

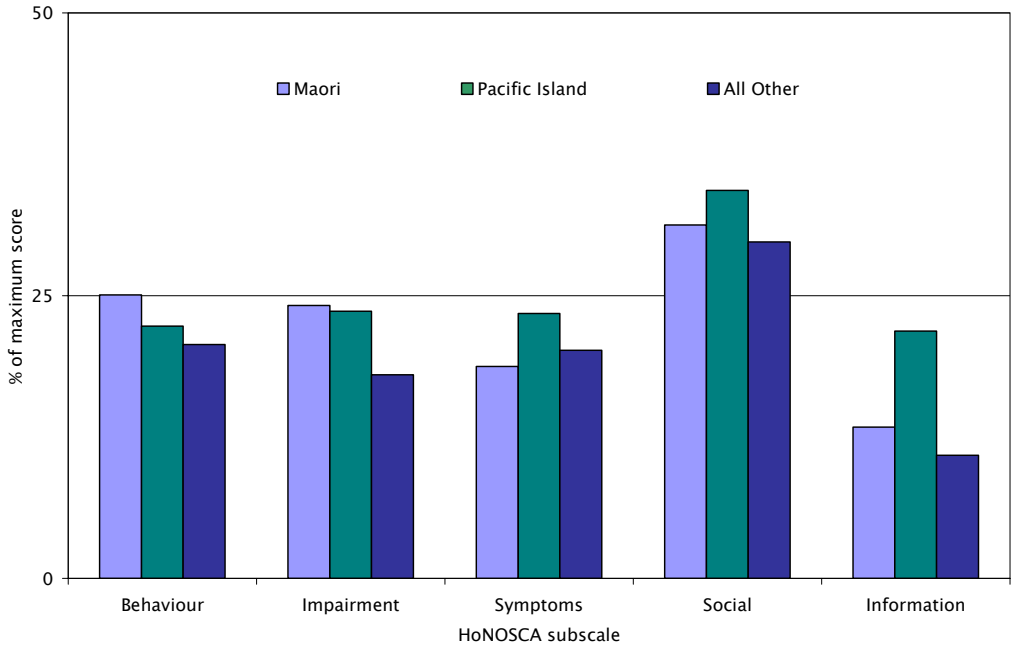


The total scores (based on items 1 to 13) for the three groups were 13.0 for Māori, 13.7 for Pacific Island and 11.9 for the All Other grouping. These differences, based on 736, 108 and 3,716 assessments respectively, were statistically significant ($F_{(2,\ 4557)} = 10.1$, $p < .001$). Pairwise comparisons revealed that the difference between Māori and All Other, and the difference between Pacific Island and All Other, were statistically significant, while the difference between the Māori and Pacific Island was not.

Figure 32 shows that, as with the HoNOS (see Figure 27 and Figure 28), the area of greatest problem for all three groups was Social problems. Differences between the three ethnicity groupings on all five HoNOSCA subscales were statistically significant. Pairwise comparisons showed that, on the Behaviour subscale, only the difference between Māori and All Other was statistically significant. On the Impairment subscale, both the Māori and Pacific Island mean scores were higher than the All Other grouping. On the Symptoms subscale, the Pacific Island mean was significantly higher than that of Māori. None of the differences on the Social subscale were significant, but on the Information subscale, all were statistically significant. The

Pacific Island mean was significantly higher than the Māori group, which was, in turn, significantly higher than that of the All Other grouping.

Figure 32. Mean HoNOSCA subscale scores in community settings, by ethnicity grouping.



8.2.2 Change as measured by the HoNOSCA

Of the 1,716 pairs of HoNOSCA assessments (see Table 10), all but 18 were obtained in community settings. Of these, 273 (16%) were of Māori, 35 (2%) of Pacific Island, 1,381 (81%) All Other and, for 9 episodes (0.5%), the ethnicity was not known. These numbers were further reduced after applying the standard criteria for including episodes in the analysis (only episodes that lasted more than three days and that had a maximum time of 91 days between assessments). Despite the very small number of Pacific Island assessments, they were retained in the analysis to maintain consistency with the analyses of the adult instruments. However, given the small numbers, the results should be interpreted with caution. The results, which appear in Table 49, show that both Māori and Pacific Island groups had very similar episode start and end scores, and both showed average improvements of around 20%. The All Other grouping had a lower (better) starting level, but made a comparable degree of improvement.

Table 49. HoNOSCA total scores at start and end of community episodes by ethnicity groupings.

Ethnicity grouping	N	Days	Start	End	% change
Māori	141	66.5	15.0	12.2	19.1%
Pacific Island	17	67.1	15.6	12.3	20.8%
All Other	734	68.2	12.8	10.6	16.7%

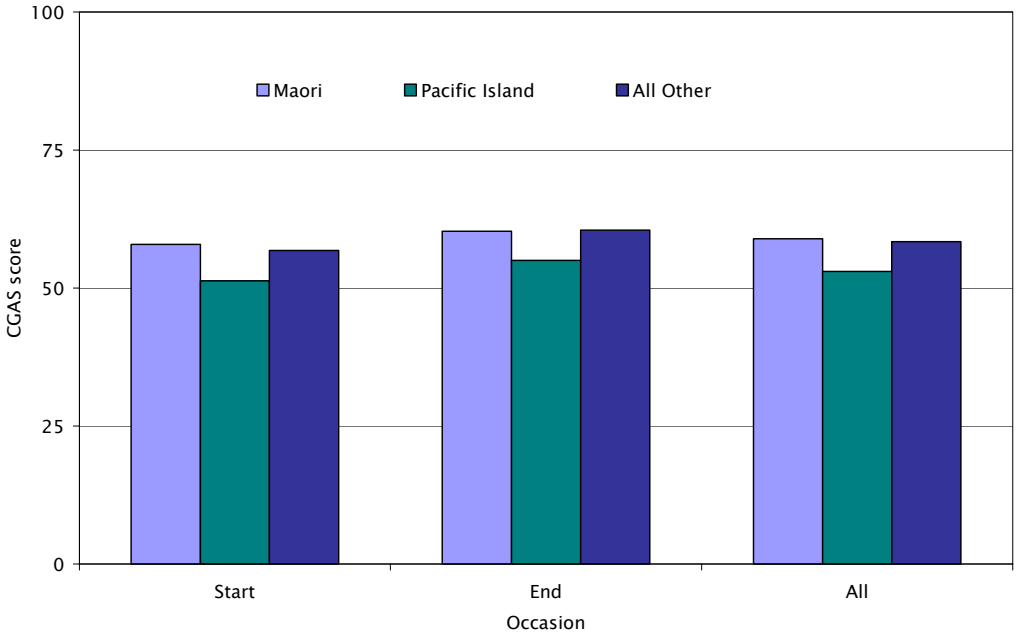
A multiple regression analysis similar to the ones in the previous sections was performed. The difference between start and end HoNOSCA total scores was predicted in terms of the start total score and ethnicity grouping. This analysis showed that greater improvements are significantly associated with higher start scores. Māori and Pacific Island changes were not significantly

different from those of the All Other grouping. In summary, the differences in changes are mainly a product of the differences in the start ratings and not related to ethnicity.

8.2.3 CGAS profile

Of the 5,160 valid CGAS ratings, all but 180 were obtained in the community. The numbers of CGAS assessments conducted with Māori or Pacific Island in inpatient settings were too small to allow meaningful analysis. As Figure 33 shows, the Pacific Island group is consistently rated lower (worse) than the other two ethnicity groupings. This was the case at both the start and end ratings.

Figure 33. Mean CGAS scores in community settings, by ethnicity grouping.



8.2.4 Change as measured by the CGAS

Of the 1,994 pairs of CGAS assessments (see Table 10) 310 (16%) were of Māori, 52 (3%) of Pacific Island, 1,620 (81%) All Other. For 12 (0.6%) episodes ethnicity grouping was not known. These numbers were further reduced after applying the standard criteria for including episodes in the analysis (only episodes that lasted more than three days and that had a maximum time of 91 days between assessments). As before, the very small number of Pacific Island assessments were retained in the analysis to maintain consistency with the analyses of the adult instruments. However, given the very small numbers, the results are not reliable. There were too few paired CGAS assessments in inpatient settings for analysis so the analysis is restricted to community episodes only. Table 50 shows that both the Māori and All Other groups improve by about 5 points, or about 10%. There was no improvement recorded for the small number of Pacific Island episodes but, given the small volume, this result is not meaningful.

Table 50. CGAS total scores at start and end of community episodes by two ethnicity groupings.

Ethnicity grouping	N	Days	Start	End	% change
Māori	167	67.0	53.5	59.0	10.3%
Pacific Island	23	67.0	50.8	50.8	0.0%
All Other	878	68.8	56.1	61.2	9.1%

8.2.5 FIHS profile

FIHS total scores were available for 425 Māori, 76 Pacific Island, and 2,271 All Other assessments. The mean scores were 3.46, 3.68 and 2.88 respectively. These differences are statistically significant ($F_{(2, 2769)} = 19.13$, $p < .001$). Pairwise comparisons showed that the Māori and Pacific Island means were not significantly different from each other, but both were significantly different from the mean for the All Other grouping.

Examination of the seven FIHS items showed that the percentage of All Others rated positive was lower than the percentage of Māori and Pacific Island in every case, and these differences were significant at the $p \leq .01$ level or higher for all items with the exception of item 4 (Problems related to primary support group, including family), which was not statistically significant. In addition, for six of the seven items, the percentage of Pacific Island rating positive was higher than the percentage of Māori.

FIHS was not required to be collected at the beginning of episodes, so there was no paired data from which to assess change.

8.3 Summary of comparison between outcomes for Māori, Pacific and Other Consumers

Adult consumers rated on the HoNOS

There were consistent differences between the three ethnicity groupings on the HoNOS and these differences were more marked in the community than in inpatient settings. Māori consumers were rated as having higher levels of severity than Pacific Island consumers who, in turn, were rated higher than the All Other grouping on the Behavioural and Social subscales. Pacific Island consumers were rated higher than Māori consumers on the Hallucinations/delusions item, and both were rated higher than the All Other grouping. Conversely, the All Others group were rated the highest on the Depression related items, on which the Pacific Island group were rated lowest.

In inpatient settings, improvements of the order of 50% of initial score were observed for all three ethnicity groupings during the course of an episode of care. Changes between the start and end of episodes were smaller in the community, and here the improvements for Māori and Pacific Island were slightly less after differences in the starting levels were taken into account.

Adult consumers rated on the LSP-16

On the LSP-16 there were consistent differences between the three ethnicity groupings, with Māori on average being rated as having the lowest levels of functioning, followed by Pacific Island and All Others. Changes during the course of a community episode were small – of the order of 5% to 8%, with little difference between the ethnicity groupings after initial levels were taken into account.

Child/Youth consumers rated on the HoNOSCA

In the HoNOSCA assessments conducted in the community, the total scores for Māori and Pacific Island consumers were significantly higher than those of the 'All Other' grouping. This stemmed largely from differences on the Behaviour and Social subscales. All three ethnicity groupings showed significant change between the start and end of episodes, ranging between 17% and 21% of initial scores. Differences between ethnicity groupings in the amount of change were not significant after initial levels were taken into account.

Child/Youth consumers rated on the CGAS

The CGAS assessments of Pacific Island consumers in the community were on average 5 to 6 points lower (worse) than those of Māori and All Others. There were too few pairs of Pacific Island CGAS assessments to draw any conclusions about the amount of change. Both of the other two groups improved by about 5 or 6 points, or 10%, although for Māori this was from a somewhat worse starting level.

Child/Youth consumers rated on FIHS

On average, Pacific Island consumers had 3.7 factors influencing health status present, Māori 3.5, and All Others 2.9. Pacific Island consumers had the highest rate in six of the seven factors, and All Others had the lowest across all seven factors. The study protocol did not require change information relating to FIHS to be collected.

9. Results: the relationship between ethnicity, deprivation and outcome profiles

In interpreting the effects of ethnicity, it is important to take account of potentially confounding variables, such as deprivation, which may differ between ethnicity groupings. This chapter examines this relationship.

'Deprivation' was captured in the study by use of the NZ CAU Index of Deprivation (Salmond C and Crampton P 2002). This is a measure of relative social deprivation derived from small geographic area New Zealand census data. It is a small-area measure of deprivation. It does not measure the relative deprivation level of each person, simply where they live. A 'decile' represents 10% of the total New Zealand population. People in the '1st decile' live in areas that have the lowest rate of social deprivation. People in the 10th or 'top decile' live in areas that have the highest rate of social deprivation. The profile of consumers using this measure was reported in the compendium casemix report.

To simplify the current analyses, the ten deciles were collapsed into five, called quintiles. Even when the ten deciles were collapsed into five, there were insufficient Pacific Child and Youth episodes in the top two (least deprived) quintiles to undertake this analysis on Child and Youth episodes (a total of 13 episodes in these quintiles). Accordingly, the results in this chapter are for adult episodes only.

The first analysis examines the relationship between HoNOS total scores obtained in the community and the deprivation index for each ethnicity grouping. The number of episodes included in the analysis is shown in Table 51 and the results in Figure 34. Table 51 shows different proportions of Māori, Pacific Island and All Others by the relative deprivation quintiles. As described in more detail in the compendium casemix report, Māori and Pacific consumers are proportionally more likely to be from deciles 7-10 than others.

Table 51. Number of observations in the HoNOS analysis.

Quintile	Māori		Pacific Island		All Other	
	No.	%	No.	%	No.	%
1-2 (least deprivation)	255	9.1%	75	10.8%	1717	14.6%
3-4	270	9.7%	113	16.3%	2115	18.0%
5-6	340	12.2%	109	15.7%	2330	19.8%
7-8	721	25.8%	149	21.4%	2975	25.3%
9-10 (most deprivation)	1209	43.3%	249	35.8%	2604	22.2%
Total	**2795**	**100.0%**	**695**	**100.0%**	**11741**	**100.0%**

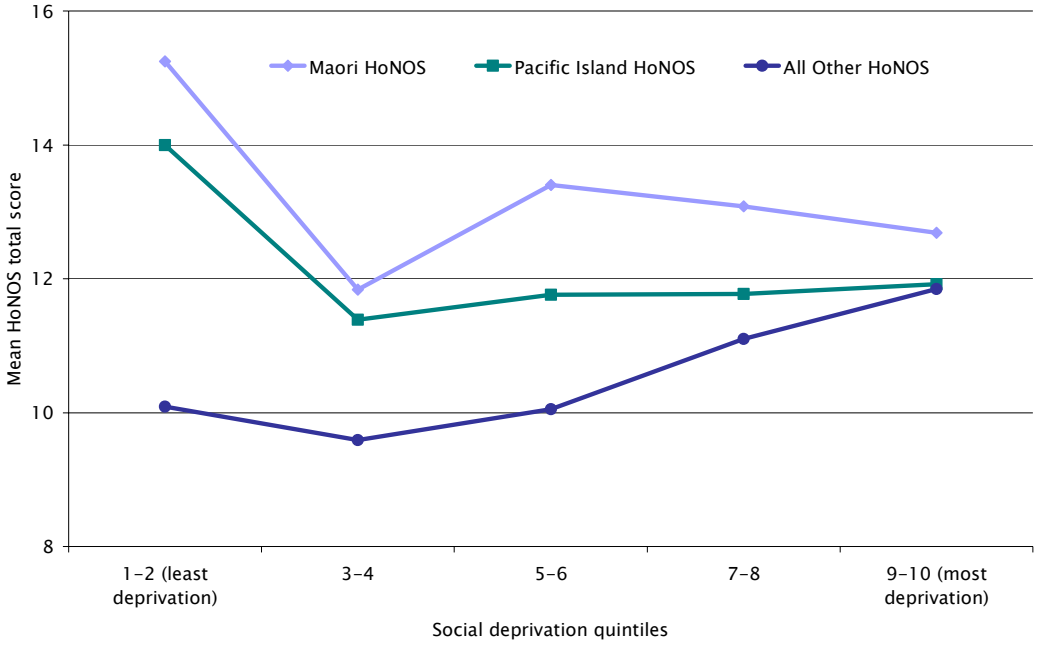
There are four key findings. First, there is a consistent pattern amongst the ethnicity groupings across all deprivation quintiles, with Māori being rated as having the highest levels of severity, the 'All Other' grouping the lowest, and Pacific Island between the two.

Second, the pattern in the average HoNOS score by quintile differs by ethnicity grouping. For both the Māori and Pacific Island groups, consumers living in areas with high socio-economic status have higher levels of severity than their counterparts living in areas of social deprivation. The pattern is reversed for the 'All Other' grouping. For this group, consumers living in areas of social deprivation have higher levels of severity than their counterparts living in areas of higher socio-economic status.

Third, the differences between the ethnicity groupings are greatest in the least deprived quintile and smallest in the most deprived quintile. The differences between the ethnicity groupings in mean HoNOS scores are greatest in the most socially advantaged areas.

Finally, consumers living in deciles 3 and 4 have significantly lower scores than those living in other areas. This applies across all ethnicity groupings.

Figure 34. HoNOS total scores by deprivation quintile and ethnicity grouping.

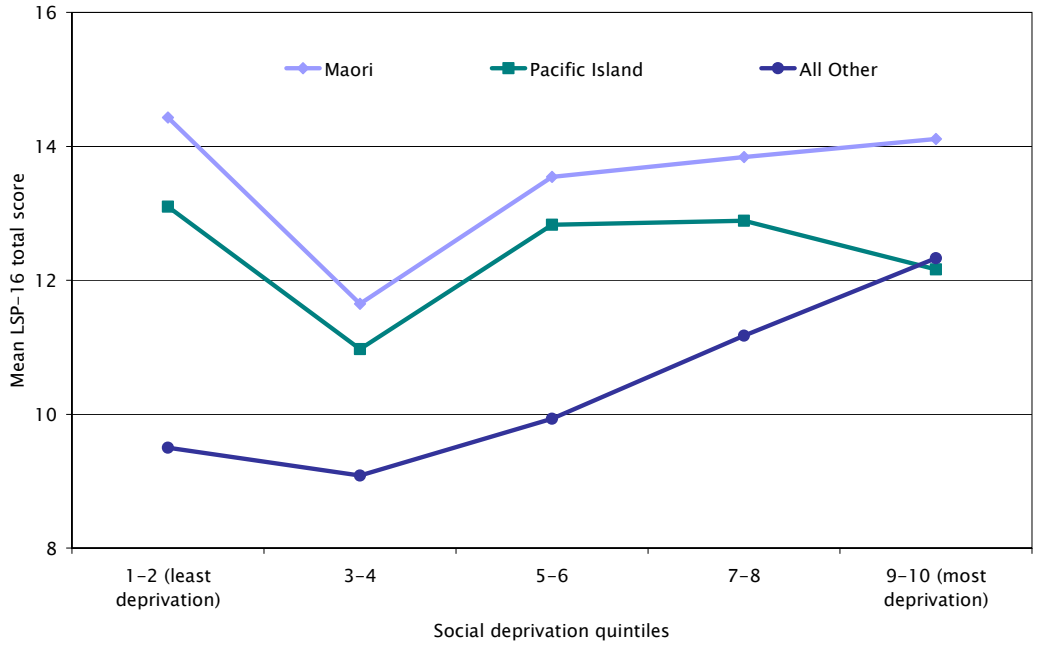


The preceding analysis was repeated for the LSP-16 with the number of episodes in the analysis shown in Table 52 and the results in Figure 35. While less clear-cut than the HoNOS results, the findings are nevertheless consistent.

Table 52. Number of observations in the LSP-16 analysis.

Quintile	Māori		Pacific Island		All Other	
	No.	%	No.	%	No.	%
1-2 (least deprivation)	264	9.3%	83	10.6%	1801	14.5%
3-4	290	10.2%	136	17.3%	2358	19.0%
5-6	349	12.3%	127	16.2%	2458	19.8%
7-8	735	25.9%	161	20.5%	3105	25.0%
9-10 (most deprivation)	1200	42.3%	277	35.3%	2679	21.6%
Total	**2838**	**100.0%**	**784**	**100.0%**	**12401**	**100.0%**

Figure 35. LSP-16 total scores by deprivation quintile and ethnicity grouping.



9.1 Summary of the relationship between ethnicity, deprivation and outcome profiles

Overall the analysis identifies a higher level of severity and lower level of functioning among Māori and Pacific consumers when compared with others – in both inpatient and community settings. The findings show improvements occurring for all groups. However this analysis raises important questions as to the reasons for severity and functioning differences between ethnic groups and this is an issue for further consultation and discussion.

This analysis has identified two key findings. Firstly, within each level of deprivation, there is a trend for Māori to have higher levels on both the HoNOS and the LSP-16 when compared with Pacific Island (who fall between the two) and with the "All other" grouping. This finding suggests that ethnicity differences in severity and function are likely impacted upon by both socioeconomic factors and issues related to ethnicity. This finding is largely consistent with that of Blakely (2004) who reported large differences in mortality between ethnic groups even after standardising for income[22].

Secondly, overall the analysis found that consumers living in areas of high relative social deprivation tend to have higher levels of severity and lower levels of functioning when compared with counterparts in areas of low relative deprivation (i.e. those with higher socioeconomic status). In contrast, analysis for Māori and Pacific peoples found the reverse pattern with higher levels of severity and poorer functioning in those from areas of least relative deprivation. It is not possible to explain these findings on the basis of the data alone. There are a few possible explanations that could be explored.

[22] Blakely T (2004) Income and ethnicity both critical in health funding: study: New research shows Māori still dying younger -- regardless of income levels. Pre-release findings of the New Zealand Census-Mortality Study. Press release 3 March 2004

One possible explanation is that issues related to ethnic or social class bias may contribute to the HoNOS and LSP-16 ratings made. There may be an 'anchor effect' in how ratings are made with the rater assigning scores relative ('anchored') to other consumers with whom they come into contact. A rater may assign lower (better) scores to a consumer who they feel comfortable and familiar with. For example, it may be that there are less Māori and Pacific responsive clinical services in areas of relative low deprivation resulting in Māori and Pacific Island consumers being rated differently in lower deprivation areas than they would in higher deprivation areas.

Ethnic and social class bias may also contribute to a second possible reason: the point at which treatment is received may differ by both ethnicity and relative deprivation. Differences in the pathways of access to care (e.g. when help is sought, when help is offered and when people are accepted into health care) may occur differently across ethnicity and socio-economic status groupings.

A possible reason may be social and cultural dislocation. Māori and Pacific Island consumers living in areas of relative social advantage may experience more social and cultural dislocation and may have less connection with family networks and community support agencies. Failure to identify problems early may be one outcome, resulting in consumers not seeking/being accepted into treatment until later in their illness.

Finally, there may be other population differences, not statistically taken into account in this analysis. For example, Māori and Pacific Island consumers with relatively low level symptoms and with higher incomes may be more likely to access services not captured in this study, such as private therapists and alternative treatments. A confounding factor may be that some mental health consumers live in areas of relative advantage because these areas are more likely to contain large houses that can be rented by consumers living in a group or shared situation.

A final possibility is that there are other statistical factors that are contributing to the findings. For example this analysis is not about the amount of change measured by the instruments but rather the mean score.

This analysis raises important issues that require further consultation with key stakeholders in order to identify both the meaning and importance of these findings.

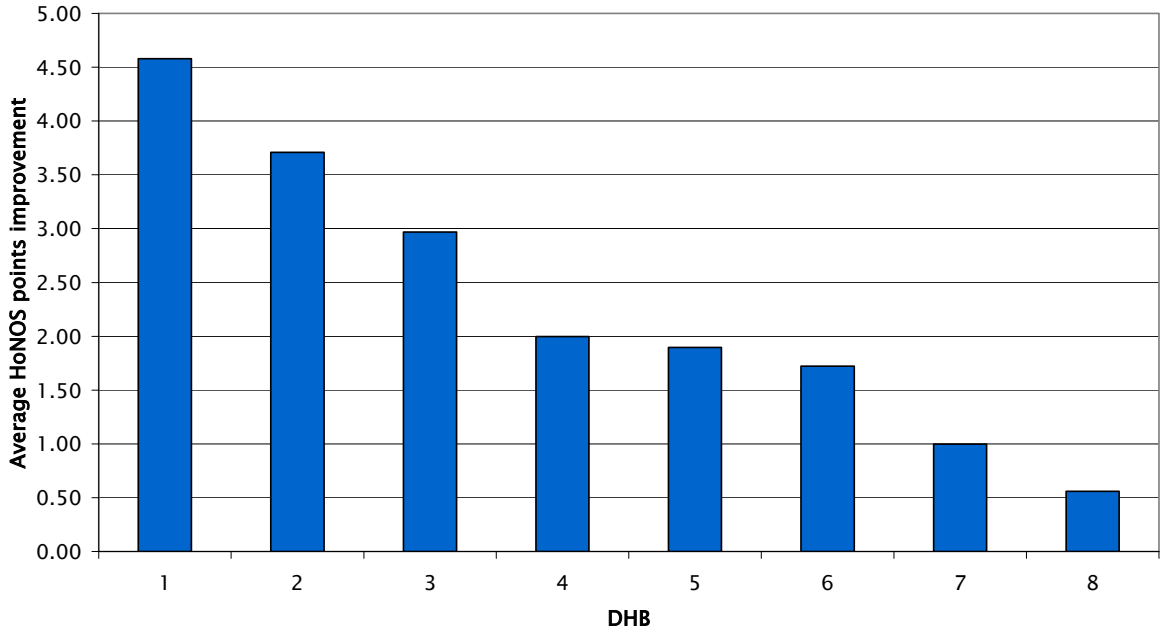
10. Results: outcomes for each casemix class

This chapter examines the relationship between outcomes and casemix. Previous chapters had suggested that the amount of change captured by the various instruments is related to the score at the start of the episode. This chapter begins by profiling outcomes by casemix class and then moves on to examine differences between the outcomes reported by each DHB, after taking into account differences in their casemix. The casemix classes in this chapter are described in more detail in the compendium casemix report.

10.1 Change as measured by the HoNOS

Figure 36 shows the average HoNOS change achieved by the 8 DHBs. DHB 1 achieved the most change, with consumers at this DHB improving by an average of over 4.5 HoNOS points. DHB 8 achieved the least change, with the average improvement being just 0.6 of a HoNOS point.

Figure 36. Average HoNOS improvement by DHB.



There are two possible reasons for the difference between DHB 1 and DHB 8. One possible reason is that DHB 1 provides better mental health care. Consumers at this DHB achieve better outcomes because DHB 1 delivers more effective care. The other possible reason is that the mix of consumers at DHB 1 is different to the mix at DHB 8. Consumers at DHB 1 achieve better outcomes because, as a group, they are different to those seen by DHB 8.

What is obvious from the results above is that comparing the performance of the various DHBs is only meaningful if the types of consumers they see are taken into account. Variation in health care outcomes is a fact of life and measurement tools are required that help to understand the different sources of this inevitable variation.

One obvious difference between the DHBs is that they have a different mix of community and inpatient episodes. But there are also other possible differences such as differences in the age profile, ethnicity profile or problem severity profile. Factors such as these need to be taken into account to make meaningful comparisons.

The measurement tool that is used in this chapter to control for differences in the mix of consumers is the casemix classification developed in the compendium study. This classification assigns consumer episodes of care to a 'casemix class' based on the characteristics of the consumer. Consumers who are similar are assigned to the same class. Consumers who are different are assigned to different classes. When outcomes results are standardised in this way to take into account the mix of consumers, any remaining differences can be attributed to differences between the DHBs. This is akin to an insurer making 'risk-adjustments'.

There are 29 different classes for adult episodes in the New Zealand mental health classification, 16 for inpatient episodes (classes numbered from AI01 to AI16) and 13 for community episodes (classes numbered AC01 to AC13). They are shown in Table 53. This table shows the average improvement in HoNOS scores for each of the casemix classes. It is sorted based on the amount of improvement achieved. The table shows, by class, the average score at the start of episodes and the average score at the end. The number of episodes included in the analysis is also shown. Episodes were only included in this analysis if they had both complete casemix data and complete outcome data. Insufficient outcome data were available to report on one class.

Class AI10 showed the most improvement. This class is for Pacific Island consumers who, at the beginning of their episode of care, were experiencing moderate to severe hallucinations and who were treated as an involuntary patient. The focus of their care during the episode was 'acute'. On average, consumers in this class improved by 12 HoNOS points. At the other end of the spectrum, consumers in Class AC02 showed no change in their HoNOS score, indicating no improvement. This class is for community consumers who received voluntary treatment and whose care was classified as having an Intensive Extended focus.

Table 53. Average improvement in HoNOS scores by casemix class.

NZCL	Class description	No of episodes included in analysis	Mean HoNOS start	Mean HoNOS end	Mean HoNOS improve-ment
AI10	Mod. to severe hallucinations, Acute FoC, Involuntary, Pacific	38	19.2	6.8	12.4
AI09	Mod. to severe hallucinations, Acute FoC, Involuntary, Māori	109	19.1	7.0	12.1
AI11	Mod. to severe hallucinations, Acute FoC, Involuntary, All other	202	17.2	7.2	10.0
AI12	Mod. to severe hallucinations, not Acute FoC	68	18.3	8.8	9.5
AI08	Mod. to severe hallucinations, Acute FoC, Voluntary	113	16.3	7.5	8.7
AI02	No or mild hallucinations, age <65, Acute FoC, with impairments	242	15.7	8.3	7.5
AI03	No or mild hallucinations, age <65, not Acute FoC, Māori	20	13.4	6.2	7.2
AI06	No or mild hallucinations, age 65-74	35	13.6	7.7	5.9
AI01	No or mild hallucinations, age <65, Acute FoC, No impair	248	10.9	6.6	4.3
AI05	No or mild hallucinations, age <65, not Acute FoC, All other	58	13.5	9.2	4.2
AI16	Adult Inpatient, ongoing, involuntary, All other	173	15.5	12.1	3.5
AC09	Involuntary, age <21	37	17.8	14.6	3.2
AI07	No or mild hallucinations, age >74	50	12.4	9.5	3.0
AI15	Adult Inpatient, ongoing, involuntary, Pacific	40	12.1	9.3	2.8
AI13	Adult Inpatient, ongoing, Voluntary	53	12.4	9.9	2.5
AI14	Adult Inpatient, ongoing, involuntary, Māori	133	13.0	10.7	2.3
AC04	Voluntary, Functional Gain FoC, Pacific Island	24	11.6	9.4	2.2
AC03	Voluntary, Functional Gain FoC, Māori	107	12.1	10.1	2.0
AC13	Involuntary age >71	51	12.7	10.7	1.9
AC05	Voluntary, Functional Gain FoC, European/Other	924	11.0	9.1	1.9
AC08	Voluntary, Other FoC, European/Other	1884	11.0	9.6	1.5
AC06	Voluntary, Other FoC, Māori	447	14.9	13.8	1.1
AC10	Involuntary, age 21-71, Māori	235	13.7	13.2	0.5
AC07	Voluntary, Other FoC, Pacific Island	120	12.2	11.8	0.4
AC11	Involuntary, age 21-71, Pacific Island	48	15.8	15.4	0.4
AC12	Involuntary, age 21-71, European/Other	514	12.4	12.2	0.3
AC02	Voluntary, Intensive Extended FoC	274	13.9	13.9	0.0
AI04	No or mild hallucinations, age <65, not Acute FoC, Pacific	insufficient cases for analysis			
All classes		6247	12.8	10.2	2.3

This table sorted by amount of change.

Figure 37 shows these results graphically for the inpatient classes. The left axis shows the average score at the start and the end of the episode. The right axis shows the average improvement in each class. Figure 38 shows the same results for the community classes.

Figure 39 presents the profile of inpatient episodes in another way. It shows the proportion of episodes in each class who improved, stayed the same and got worse. The proportion of inpatients whose HoNOS score improved varies from 98% in Class AI12 to 45% in Class AI15. Over 25% of consumers having 'ongoing' (long stay episodes) were rated worse at the end of their episode than they were rated at the start.

Figure 40 shows the same profile for community episodes. The proportion of community consumers whose HoNOS score improved varies from 65% in Class AC03 to 40% in Class AC11. About 30% of consumers in most classes were rated worse at the end of their episode than they were rated at the start. But over 50% of consumers in Class AC11 were rated worse at the end of their episode than they were rated at the start. However, this class is very small and the results should be treated with caution (see Table 53 above for the number of observations in each class).

Figure 37. Improvement on the HoNOS by inpatient class.

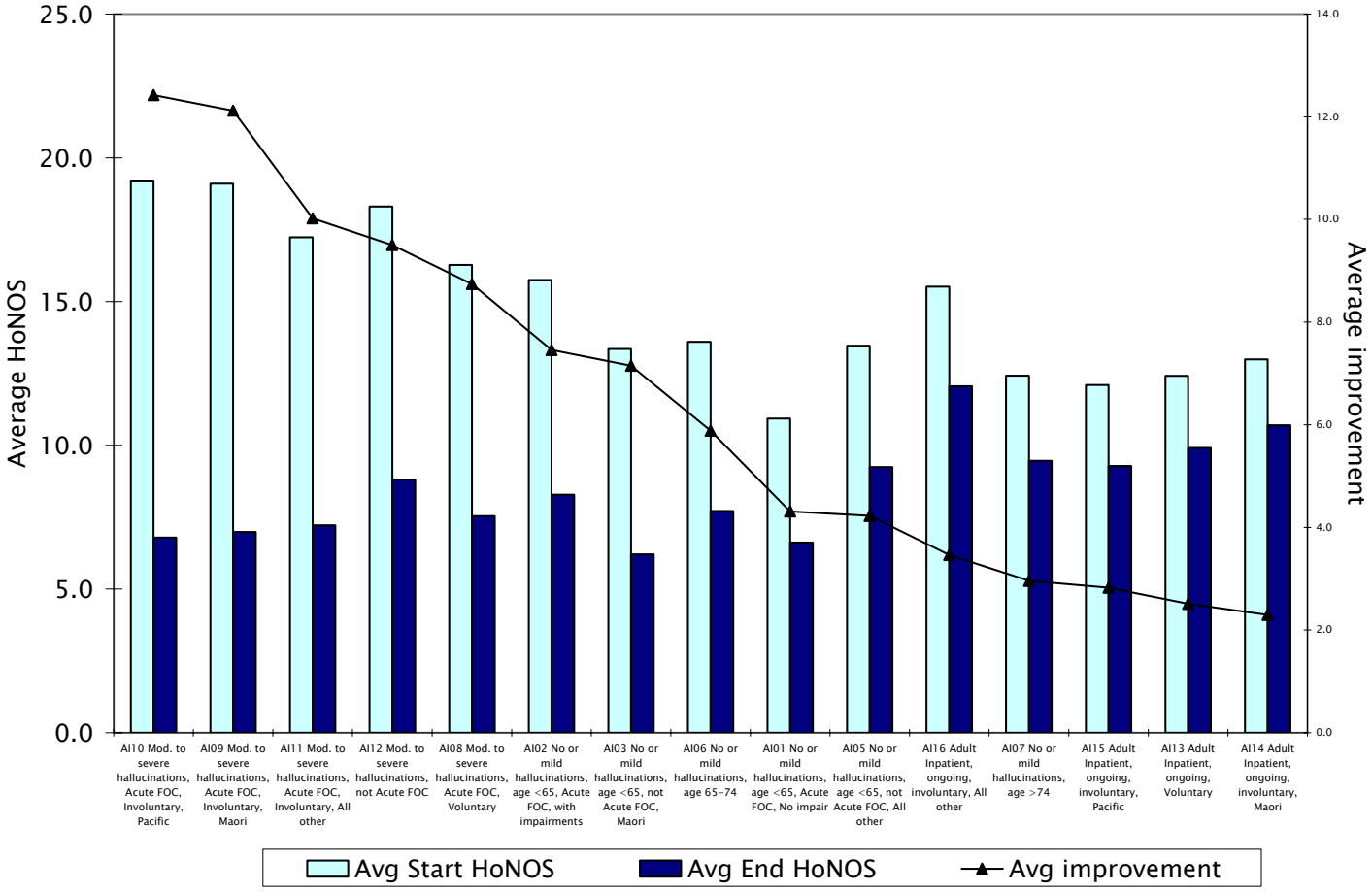


Figure 38. Improvement on the HoNOS by community class.

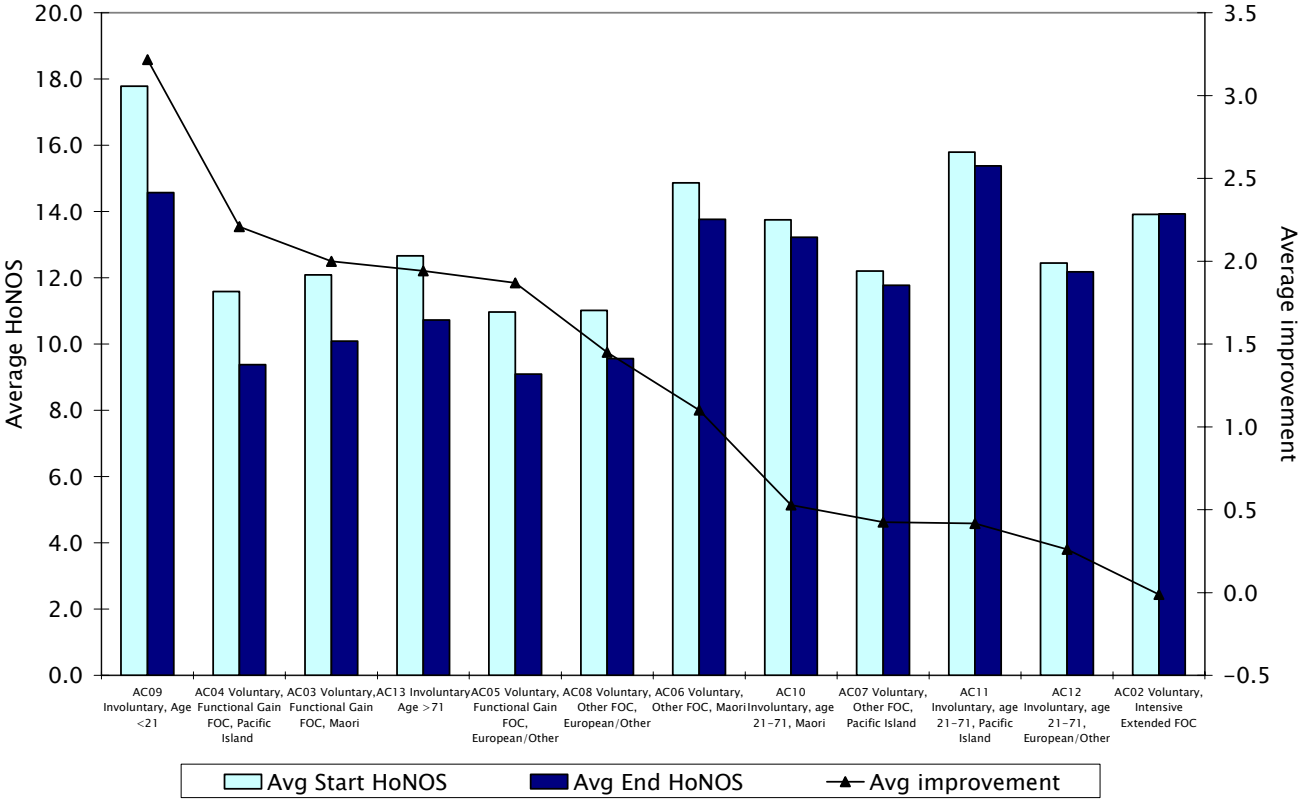


Figure 39. Summary of change profile on the HoNOS for each adult inpatient class.

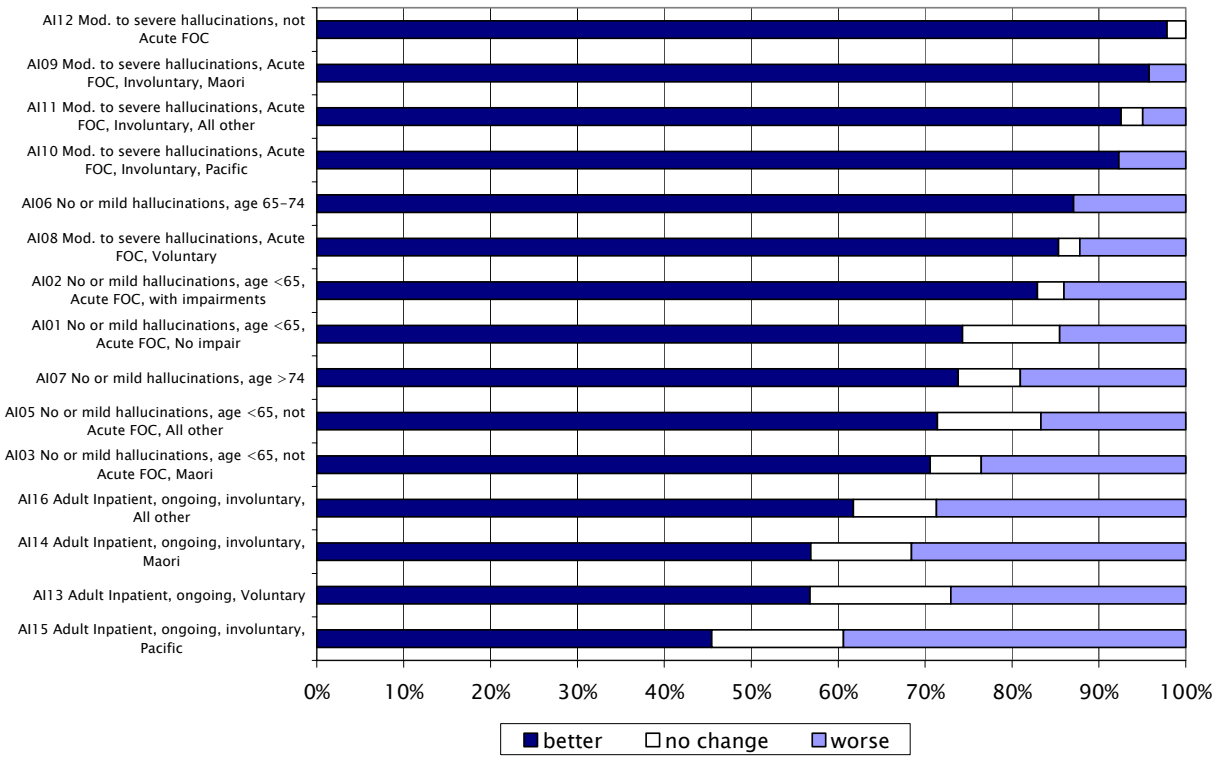
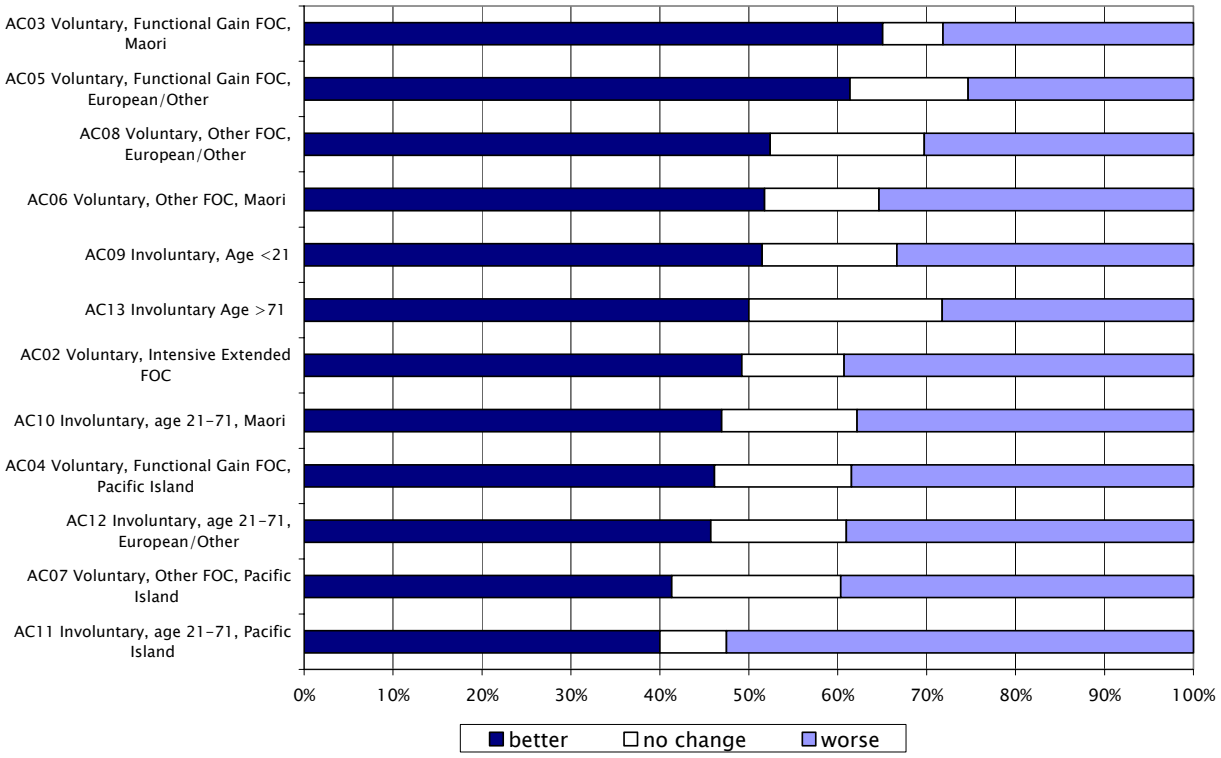


Figure 40. Summary of change profile on the HoNOS for each adult community class.



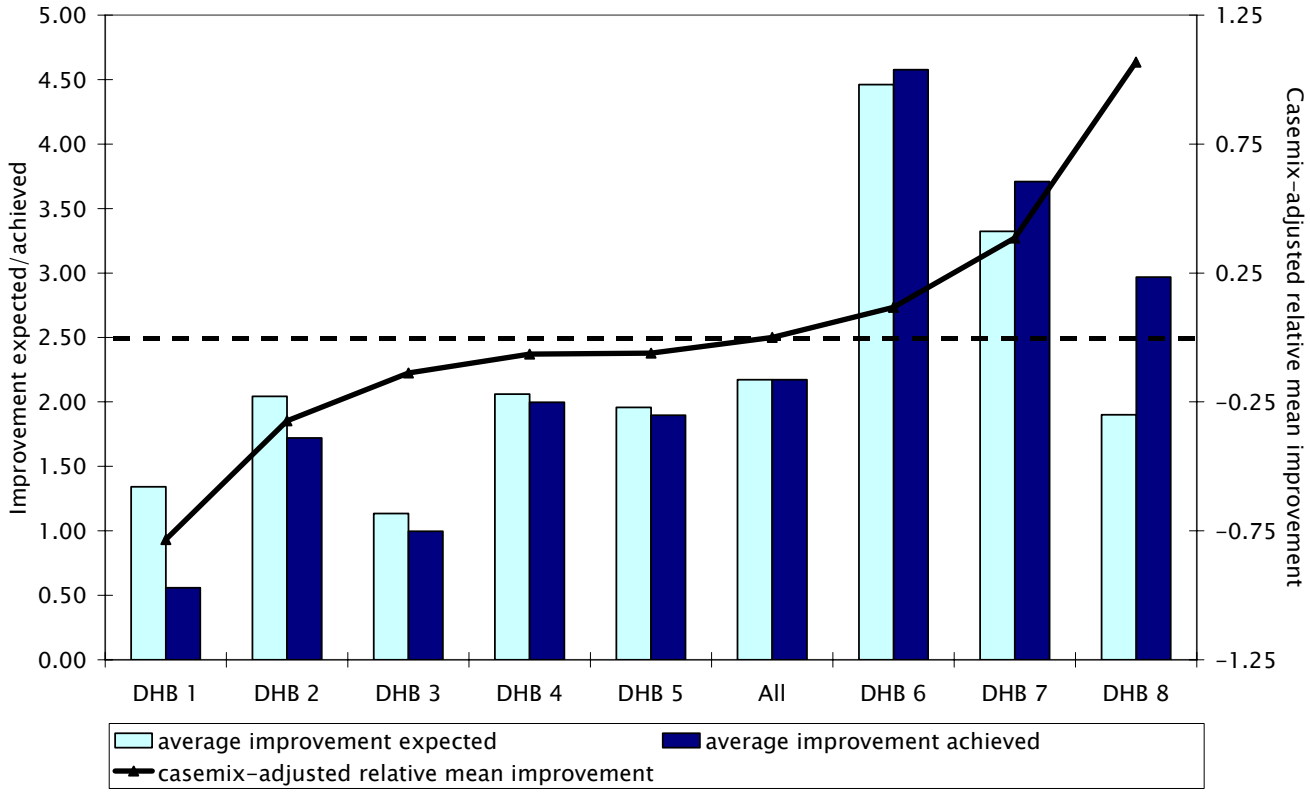
10.1.1 Casemix-adjusted HoNOS outcomes by DHB

As discussed above, a comparison between health services on some statistics (such as the change in HoNOS score) is only possible if the health services being compared treat similar consumers. Differences between the amount of improvement on an instrument such as the HoNOS may indicate differences between health services (in this case, DHBs). Alternately, they may simply reflect the unique mix of consumers in each DHB. To overcome this difficulty, it is possible to statistically control for casemix. This is achieved by adjusting measures such as HoNOS change so that the comparison is only made between similar types of consumers. In this case, the statistic of interest is the Casemix-Adjusted Relative Mean Improvement (CARMI) score for each DHB[23].

Figure 41 shows the casemix-adjusted relative mean HoNOS improvement score for each DHB. Mathematically, this is achieved by subtracting the actual HoNOS change of each episode at each DHB from the average HoNOS change of episodes in the same casemix class. These differences are then averaged across the DHB. A casemix-adjusted relative mean improvement of, say, 1 point would indicate that, on average, consumers improved by 1 HoNOS point more than similar consumers in the national study. A negative casemix-adjusted relative mean HoNOS improvement score indicates that consumers achieved less improvement than similar consumers in the national study.

[23] This term was coined by Janette Green, Senior Research Fellow (Applied Statistics) in her work in developing casemix-adjustment methodologies for use in the medical rehabilitation sector.

Figure 41. Improvement on the HoNOS by DHB.



Using this measure, the DHB that achieved the most HoNOS improvement was DHB 8. After adjusting for casemix, DHB 8 achieved an average of 1.1 points more improvement than the study average. Based on the casemix of DHB 8, the expected improvement (based on the national study average) was an average of 1.9 points on the HoNOS. DHB 8 achieved an average improvement of 2.97 points. In contrast, DHB 1 averaged 0.8 point less improvement than the study average for its casemix.

10.2 Change as measured by the LSP-16

Table 54 shows the average improvement in LSP-16 scores for each of the casemix classes. Due to the small volume of LSP-16 assessments undertaken in an inpatient setting, only community classes are included.

As before, the table is sorted based on the amount of improvement achieved. The table shows, by class, the average score at the start of episodes and the average score at the end. The number of episodes included in the analysis is also shown. Episodes were only included in this analysis if they had both complete casemix data and complete outcome data.

As with the HoNOS, Class AC09 showed the most improvement. On average, consumers in this class improved by 1.3 LSP-16 points. At the other end of the spectrum, consumers in Class AC02 and AC11 showed no change over the course of the episode.

Table 54. Average improvement in LSP-16 scores by casemix class.

NZCL	Class description	No of episodes included in analysis	Mean LSP-16 start	Mean LSP-16 end	Mean LSP-16 improve-ment
AC09	Involuntary, Age <21	41	18.5	17.2	1.3
AC03	Voluntary, Functional Gain FoC, Māori	115	13.8	12.9	0.9
AC04	Voluntary, Functional Gain FoC, Pacific Island	47	9.6	8.8	0.7
AC05	Voluntary, Functional Gain FoC, European/Other	1121	10.2	9.5	0.7
AC07	Voluntary, Other FoC, Pacific Island	186	11.8	11.1	0.7
AC06	Voluntary, Other FoC, Māori	501	14.2	13.6	0.7
AC08	Voluntary, Other FoC, European/Other	2231	10.9	10.3	0.6
AC13	Involuntary age >71	59	12.6	12.1	0.5
AC10	Involuntary, age 21-71, Māori	274	14.8	14.4	0.4
AC12	Involuntary, age 21-71, European/Other	641	14.2	13.9	0.3
AC02	Voluntary, Intensive Extended FoC	341	15.0	15.0	0.0
AC11	Involuntary, age 21-71, Pacific Island	77	14.8	14.9	0.0
All classes		5634	12.1	11.5	0.6

Table sorted by amount of change

Figure 42. Improvement on the LSP-16 by community class.

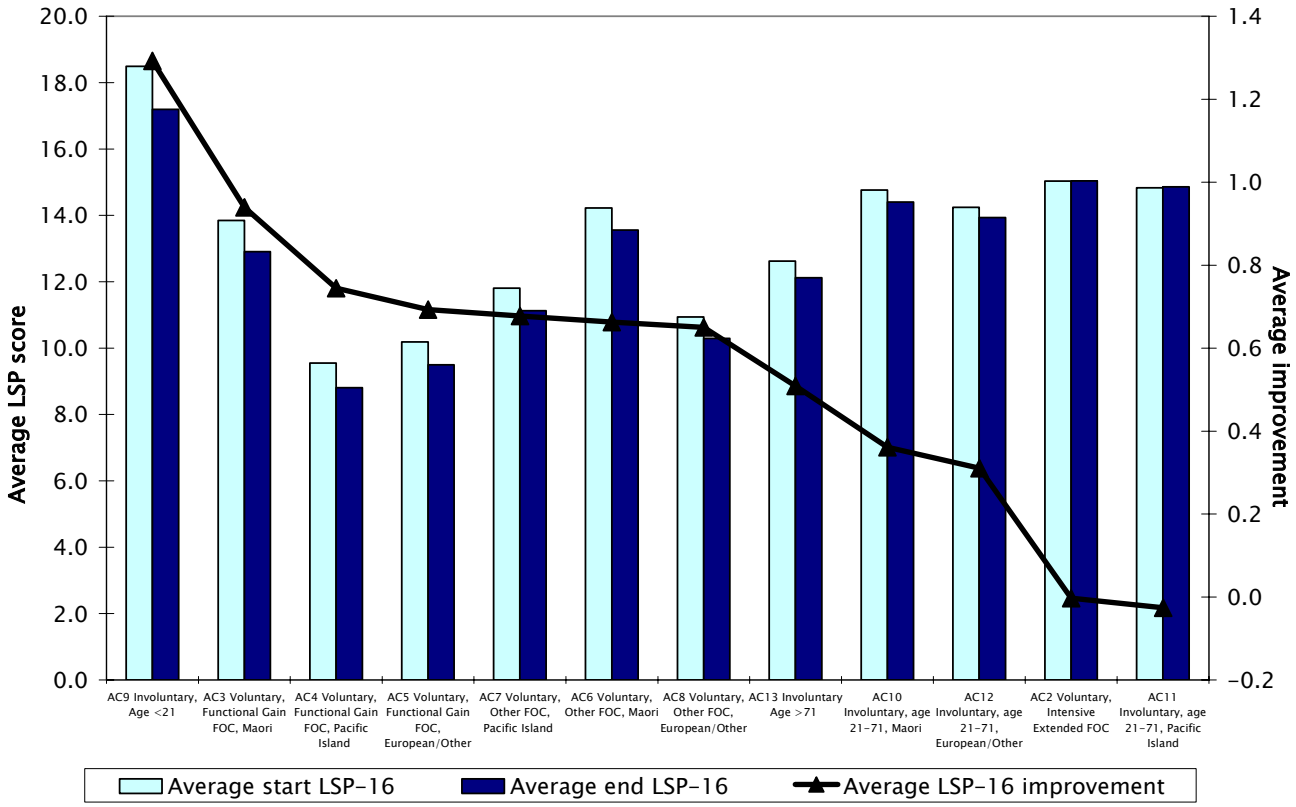
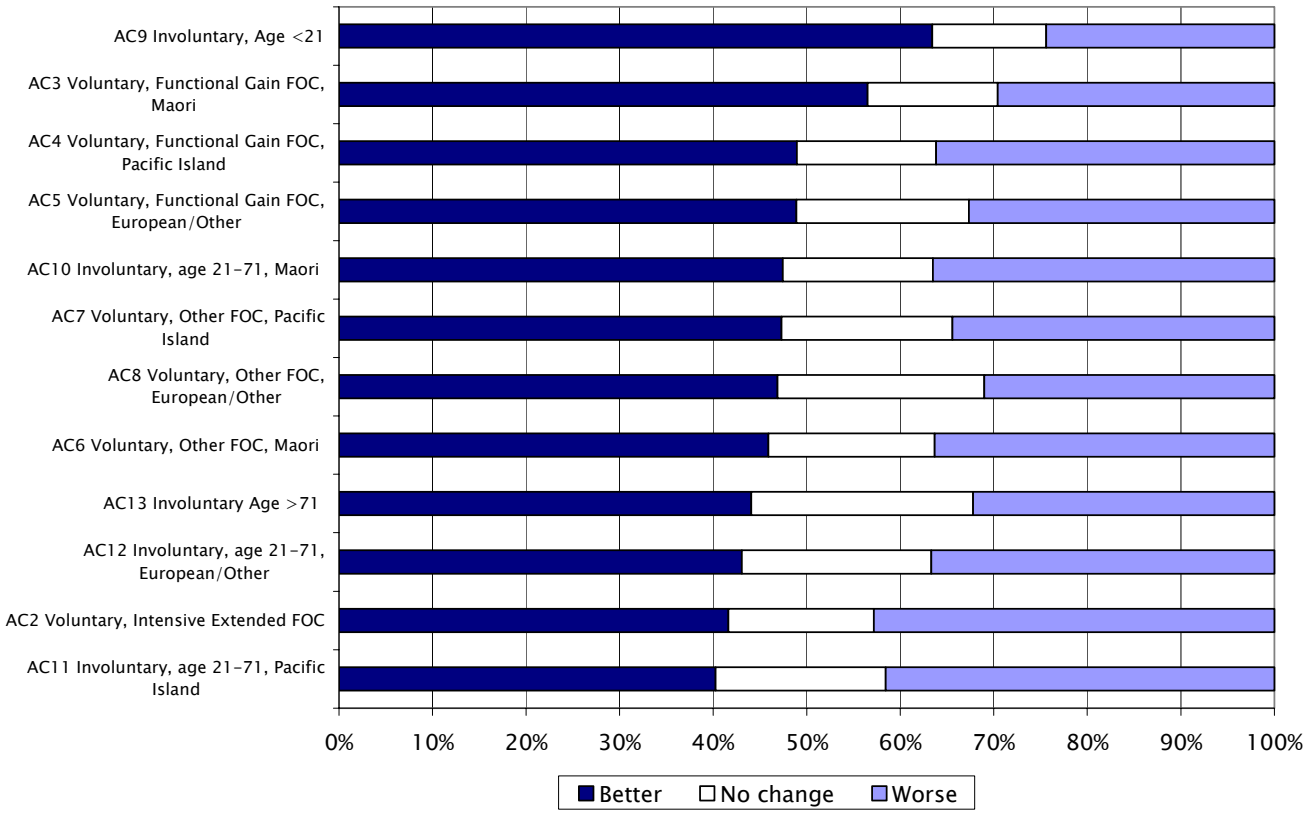


Figure 42 shows these results graphically for each class. The left axis shows the average score at the start and the end of the episode. The right axis shows the average improvement in each class.

Figure 43 presents the profile of community episodes in another way. It shows the proportion of episodes in each class that improved, stayed the same and got worse. The proportion of community consumers whose LSP-16 score improved varies from 63% in Class AC09 to 40% in Class AC11. There are two classes in which over 40% of consumers were rated worse at the end of their episode than they were rated at the start (Classes AC02 and AC11).

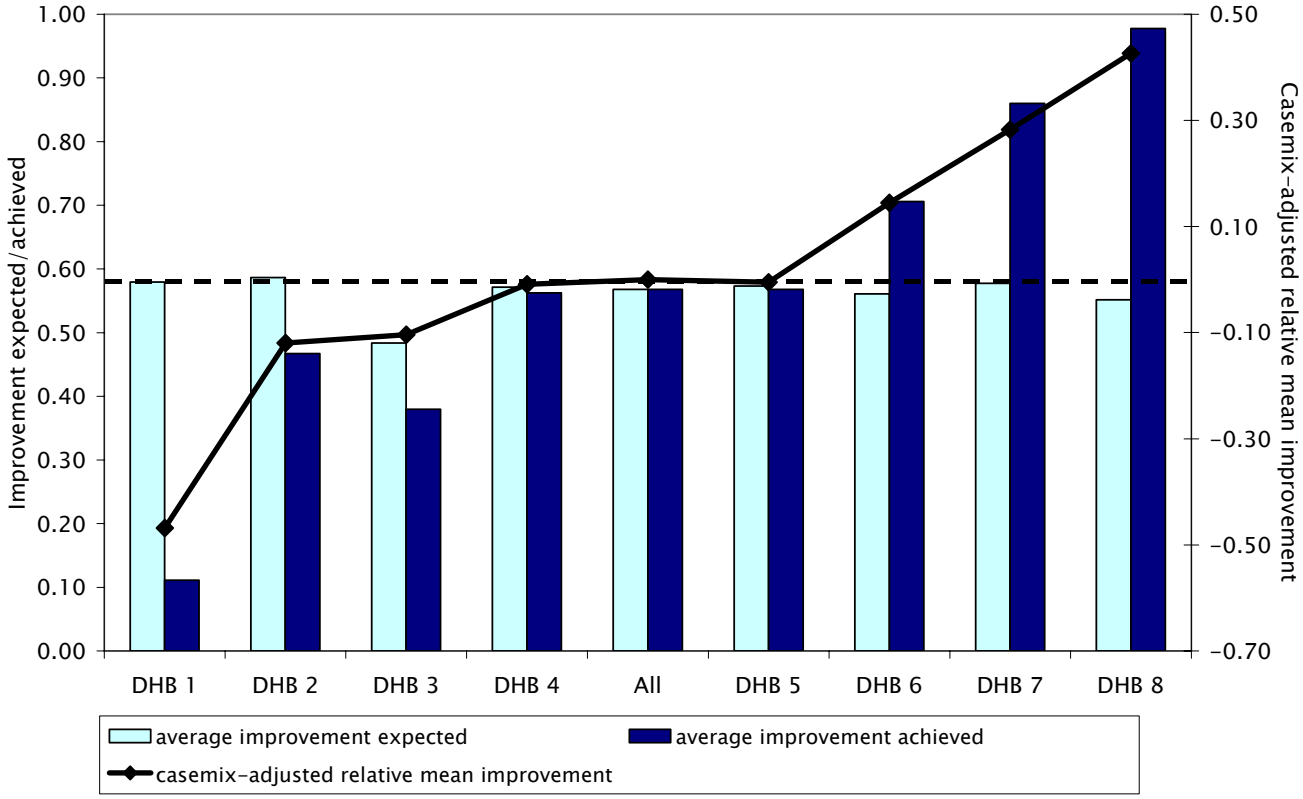
Figure 43. Summary of change profile on the LSP-16 for each adult community class.



10.2.1 Casemix-adjusted LSP-16 outcomes by DHB

Figure 41 shows the casemix-adjusted relative mean LSP-16 improvement score for each DHB (see page 93 for an explanation of this statistic). Using this measure, the DHB that achieved the most LSP-16 improvement was DHB 8. After adjusting for casemix, DHB 8 achieved an average of 0.43 points more improvement than the national study average. Based on the casemix of DHB 8, the expected improvement (based on the study average) was an average of 0.55 points on the LSP-16. DHB 8 achieved an average improvement of 0.98 points. In contrast, DHB 1 averaged 0.47 points less improvement than the study average for its casemix. Two DHBs (DHB 4 and DHB 5) achieved exactly the average improvement expected. As with the HoNOS, DHBs are numbered in this figure based on their ranking on this measure. DHB 1 in the LSP analysis may not be the same DHB as DHB 1 in the HoNOS analysis.

Figure 44. Improvement on the LSP-16 by DHB.



10.3 Change as measured by the HoNOSCA

The need to take into account the mix of consumers when making outcome comparisons was discussed in Section 10.1 on page 87 and the method used to achieve this was outlined on page 92. These same approaches apply equally to child and youth episodes.

The New Zealand mental health classification has 13 classes for child and youth episodes, 9 for community episodes and 4 for inpatient episodes. The community classes are shown in Table 53. This table also shows the average improvement in HoNOSCA scores for each of the community casemix classes. The inpatient classes are excluded from analysis due to the small number of episodes in each class. The table is sorted based on the amount of improvement achieved on the HoNOSCA. The table shows, by class, the average score at the start of episodes and the average score at the end. The number of episodes included in the analysis is also shown. Episodes were only included in this analysis if they had both complete casemix data and complete outcome data.

Class CC05 showed the most improvement. This class is for youth who, at the beginning of their episode of care, were experiencing hallucinations or delusions. On average, consumers in this class improved by 6 HoNOSCA points. At the other end of the spectrum, consumers in Class CC03 showed a slight increase in their HoNOSCA score, indicating no improvement. This class is for children who have no school attendance problems and a low HoNOSCA total score.

Table 55. Average improvement in HoNOSCA scores by casemix class.

NZCL	Class description	No of episodes included in analysis	Mean HoNOSCA start	Mean HoNOSCA end	Mean HoNOSCA improve-ment
CC5	Youth, problem hallucinations	54	22.6	16.7	5.9
CC6	Youth, problem self-injury	78	20.4	14.8	5.6
CC2	Child with school problem	108	21.4	17.3	4.1
CC7	Youth, problem school attendance	114	18.4	15.6	2.8
CC4	Child, no school problem, high HoNOSCA total	496	14.4	12.4	2.0
CC8	Youth, with emotional problem	93	13.0	11.5	1.5
CC9	Youth, none of these problems	178	9.1	8.7	0.4
CC3	Child, no school problem, low HoNOSCA total	124	4.8	5.3	-0.4
All classes		1245	14.3	12.1	2.1

Table sorted by amount of change

Figure 45 shows these results graphically for each class. The left axis shows the average score at the start and the end of the episode. The right axis shows the average improvement in each class.

Figure 46 presents the profile of the classes in another way. It shows the proportion of episodes in each class that improved, stayed the same and got worse. The proportion of consumers whose HoNOSCA score improved varies from 68% in Class CC06 to 33% in Class CC03. The profile of Class CC03 is evenly split, with one third of consumers improved, one third stayed the same and one third got worse. In contrast, only 12% of consumers in CC05 got worse.

Figure 45. Improvement on the HoNOSCA by community class.

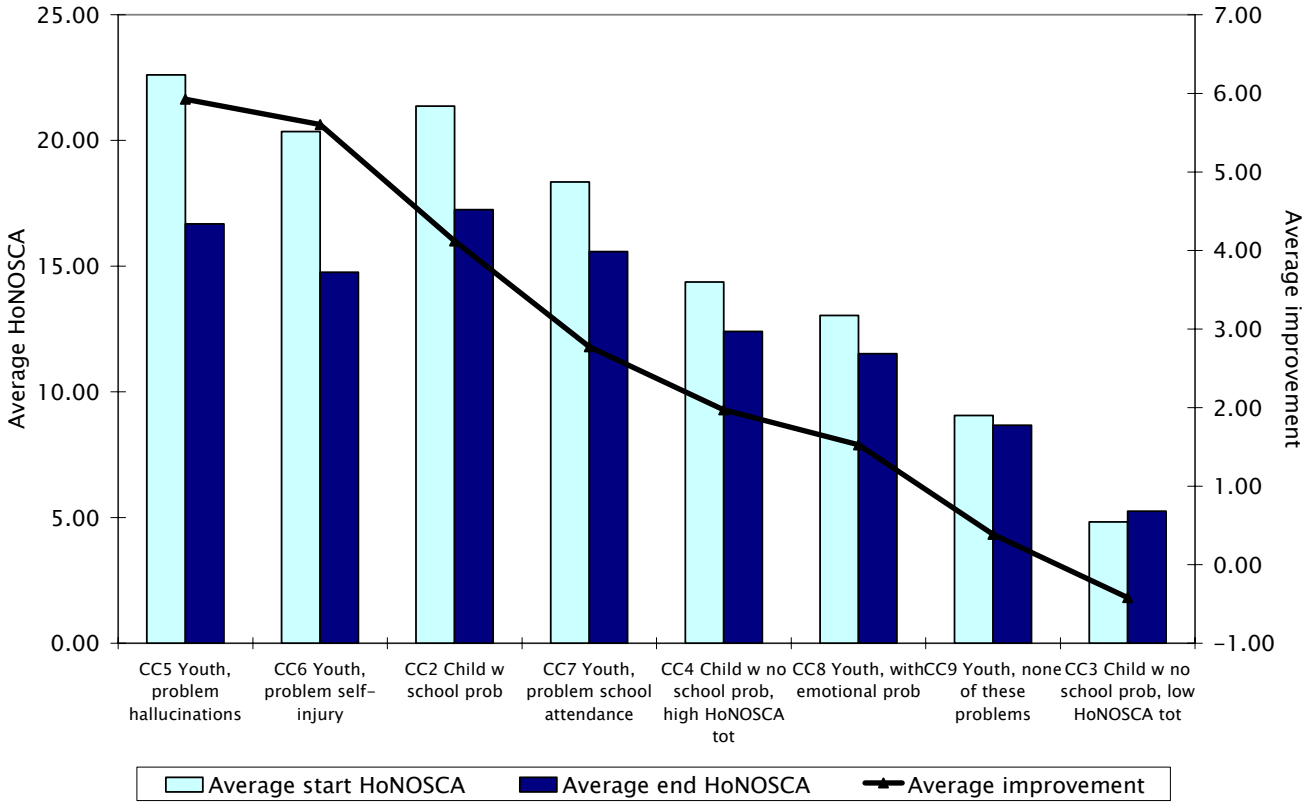
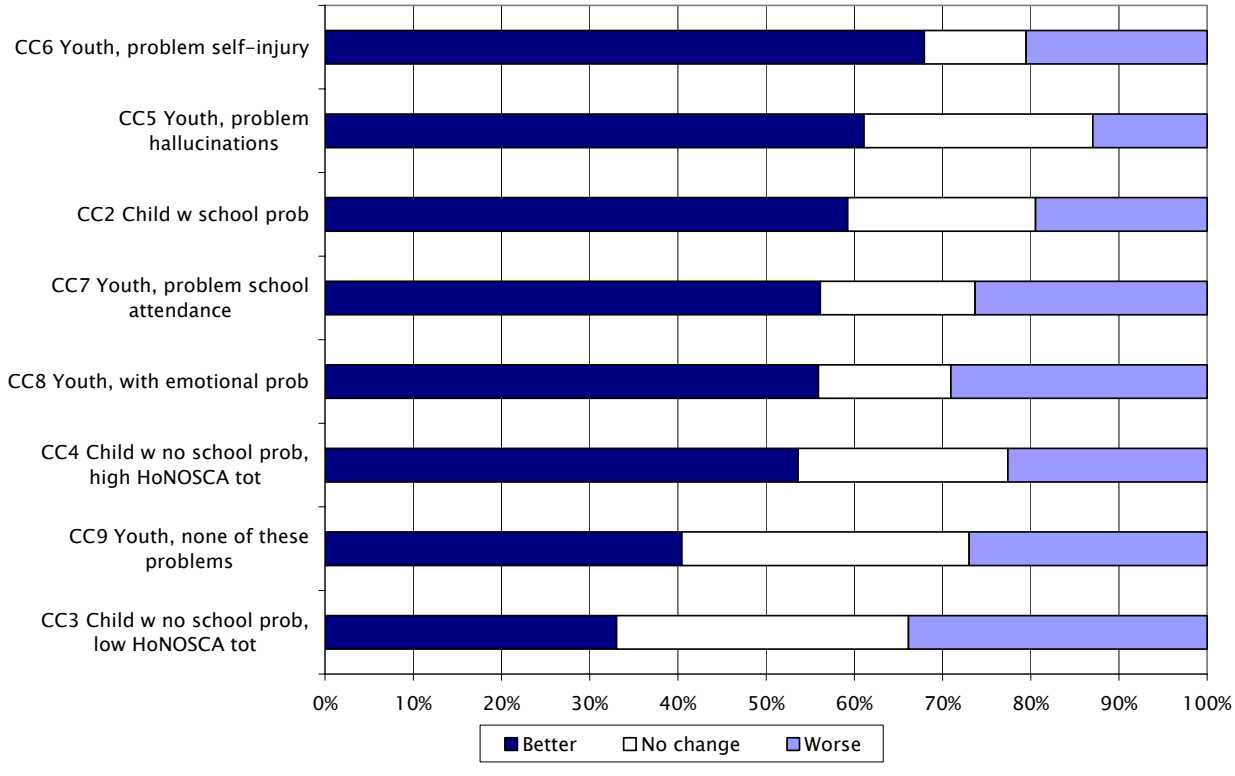


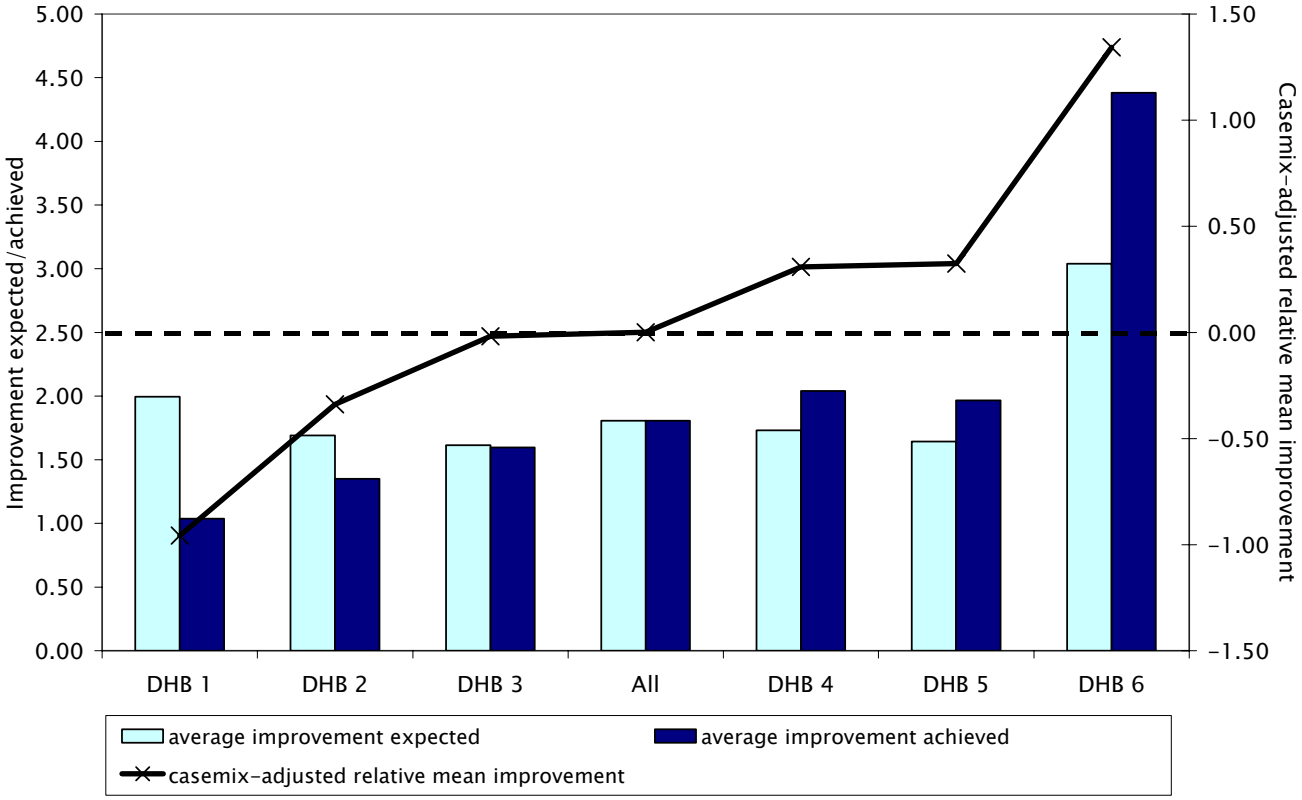
Figure 46. Summary of change profile on the HoNOSCA for each community class.



10.3.1 Casemix-adjusted HoNOSCA outcomes by DHB

Figure 47 shows the casemix-adjusted relative mean HoNOSCA improvement for each DHB (see page 92 for an explanation of this statistic). Two DHBs had insufficient episodes for analysis. Using this measure, the DHB that achieved the most HoNOSCA improvement was DHB 6. After adjusting for casemix, DHB 6 achieved an average of 1.4 points more improvement than the study average. Based on the casemix of DHB 6, the expected improvement (based on the study average) was an average of 3.0 points on the HoNOSCA. DHB 6 achieved an average improvement of 4.4 points. In contrast, DHB 1 averaged 1 point less improvement than the study average for its casemix.

Figure 47. Improvement on the HoNOSCA by DHB.



10.4 Change as measured by the CGAS

Table 53 shows the average improvement in HoNOSCA scores for each of the casemix classes. Unlike the HoNOSCA, there were sufficient inpatient episodes with both casemix and CGAS data for them to be included in the analysis. However, as shown in the table, the volumes are small in the inpatient classes. The table is sorted based on the amount of improvement achieved on the CGAS. The table shows, by class, the average score at the start of episodes and the average score at the end. The number of episodes included in the analysis is also shown. Episodes were only included in this analysis if they had both complete casemix data and complete outcome data.

Class CY01 showed the most improvement. On average, consumers in this class improved by 11 CGAS points. At the other end of the spectrum, consumers in Class CC09 showed the least improvement (4.5 points).

Table 56. Average improvement in CGAS scores by casemix class.

NZCL	Class description	No of episodes included in analysis	Mean CGAS start	Mean CGAS end	Mean CGAS improve-ment
CYI01	Inpatient, high cost diagnoses	22	40.3	51.5	11.1
CYI02	Inpatient, mid cost diagnoses	23	43.7	53.6	10.0
CYI04	Inpatient, ongoing episodes	10	38.2	47.2	9.0
CC06	Youth, problem self-injury	106	51.4	58.9	7.5
CYI03	Inpatient, low cost diagnoses	14	48.9	55.5	6.6
CC05	Youth, problem hallucinations	80	46.5	52.4	5.8
CC08	Youth, with emotional problem	122	53.7	59.0	5.4
CC07	Youth, problem school attendance	160	52.4	57.2	4.8
CC04	Child, no school problem, high HoNOSCA total	585	55.9	60.0	4.1
CC03	Child, no school problem, low HoNOSCA total	154	67.2	70.9	3.7
CC02	Child with school problem	136	50.8	54.2	3.4
CC09	Youth, none of these problems	246	62.8	65.8	2.9
All classes		1658	55.8	60.3	4.5

Table sorted by amount of change

Figure 48 shows these results graphically for each class. The left axis shows the average score at the start and the end of the episode. The right axis shows the average improvement in each class.

Figure 49 presents the profile of the classes in another way. It shows the proportion of episodes in each class that improved, stayed the same and got worse. The proportion of consumers whose CGAS score improved varies from 80% in Class CY04 to 46% in Class CC09.

Figure 48. Improvement on the CGAS by casemix class.

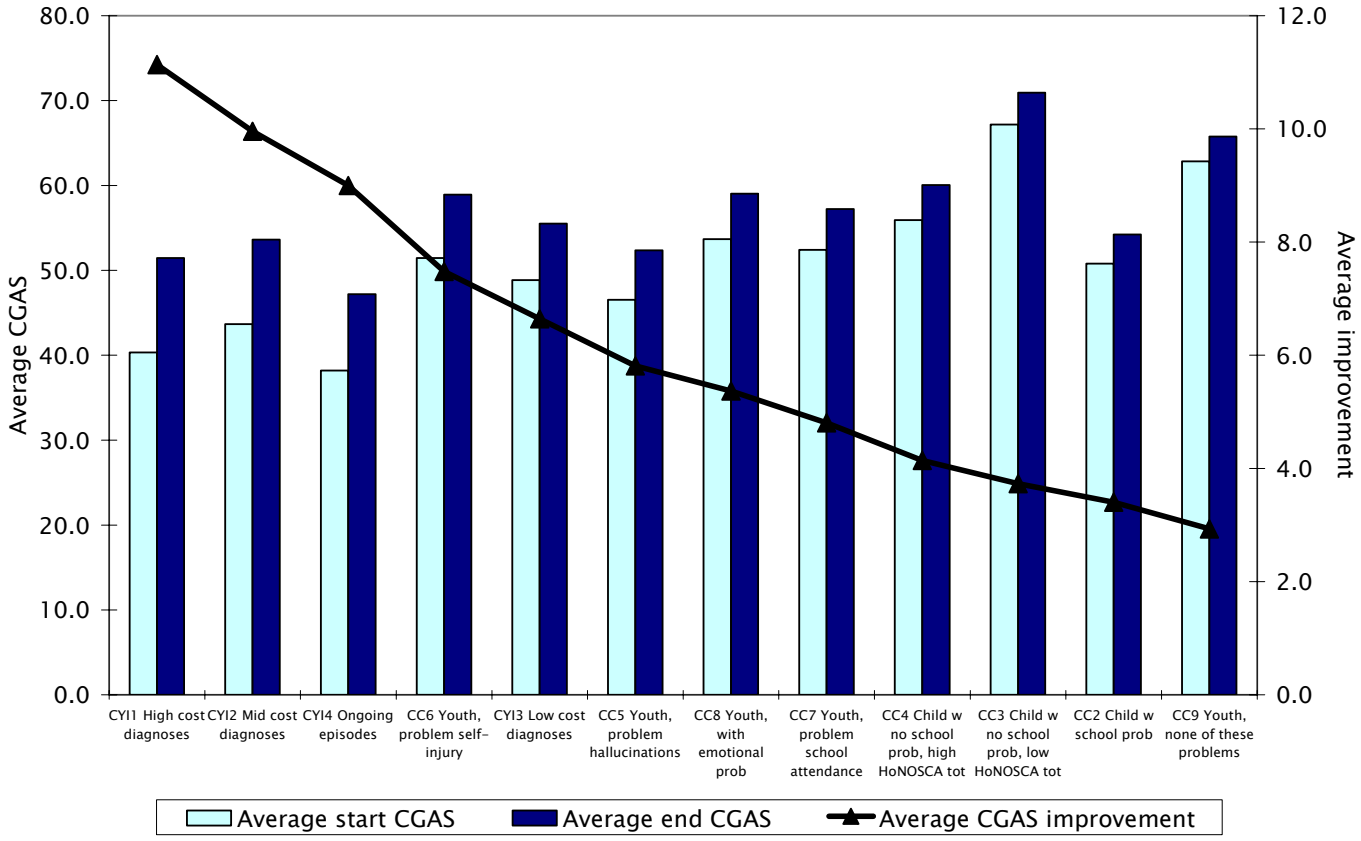
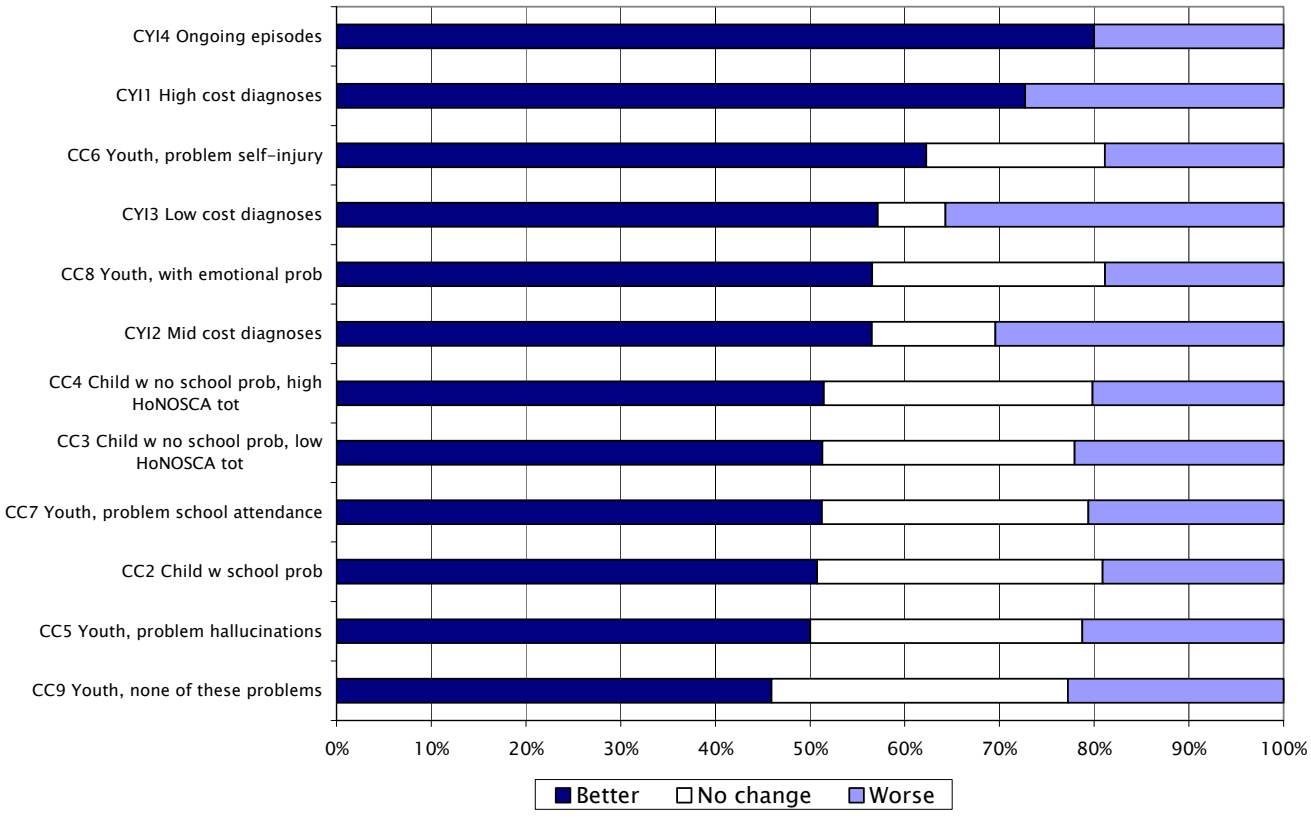


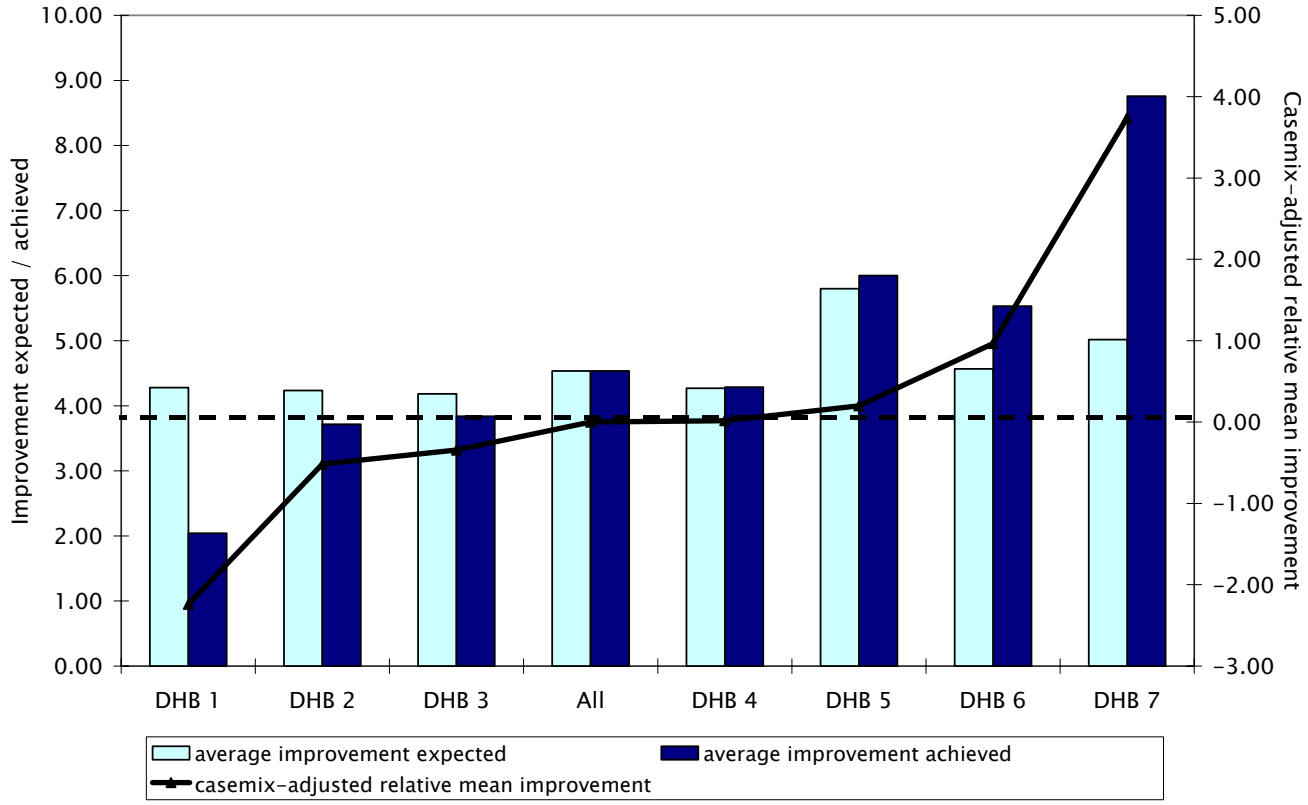
Figure 49. Summary of change profile on the CGAS for each casemix class.



10.4.1 Casemix-adjusted CGAS outcomes by DHB

Figure 47 shows the casemix-adjusted relative mean HoNOSCA improvement for each DHB (see page 92 for an explanation of this statistic). One DHB is excluded. Using this measure, the DHB that achieved the most CGAS improvement was DHB 7. After adjusting for casemix, DHB 7 achieved an average of 3.7 points more improvement than the national study average. Based on the casemix of DHB 7, the expected improvement (based on the study average) was an average of 5.0 points on the CGAS. DHB 7 achieved an average improvement of 8.8 points. In contrast, DHB 1 averaged 2.2 points less improvement than the study average for its casemix.

Figure 50. Improvement on the CGAS by DHB.



10.5 Summary of the relationship between outcomes and casemix

The results reported in this chapter demonstrate that there is an important relationship between casemix and outcomes. Consumer outcomes differ between the casemix classes that were developed in the compendium study. In that study, the goal was to define a set of classes based on their cost. This outcomes analysis has demonstrated that the classes are also useful as a way of measuring and understanding outcomes. The classes, and the average amount of change on each relevant instrument, are summarised in Figure 51, Figure 52 and Figure 1Figure 53.

Figure 51. Change profile of the adult inpatient classes.

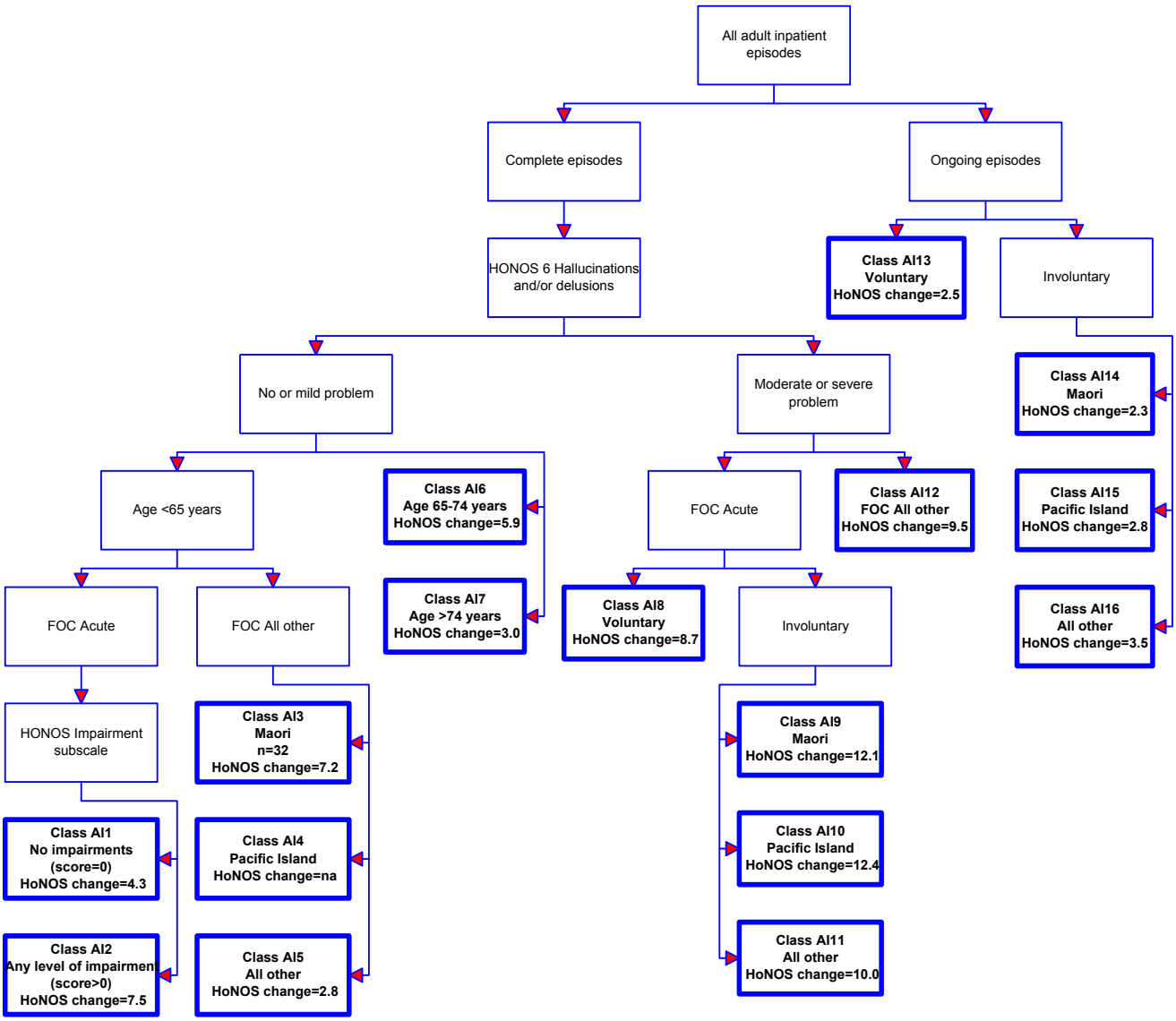


Figure 52. Change profile of the adult community classes.

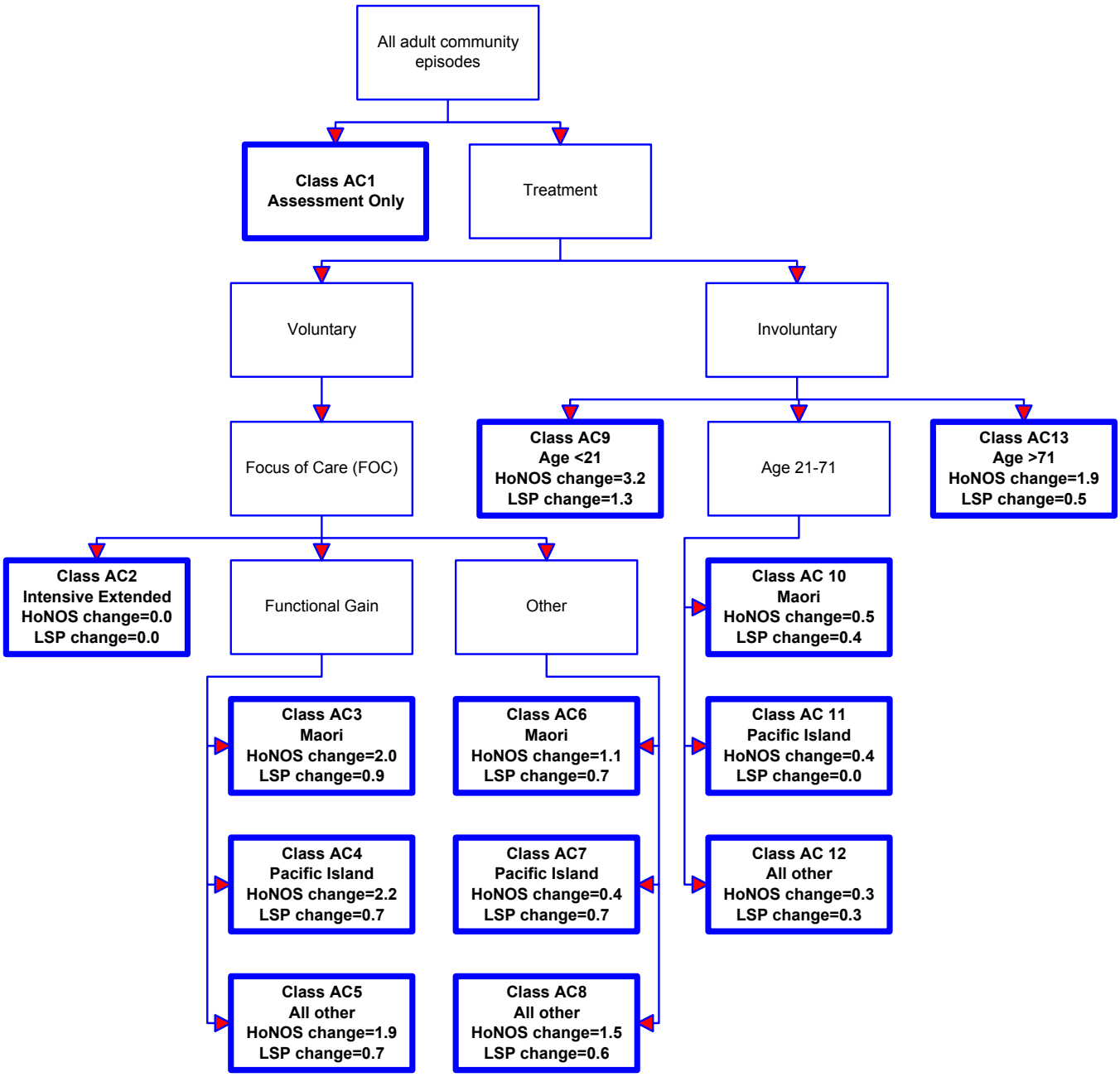
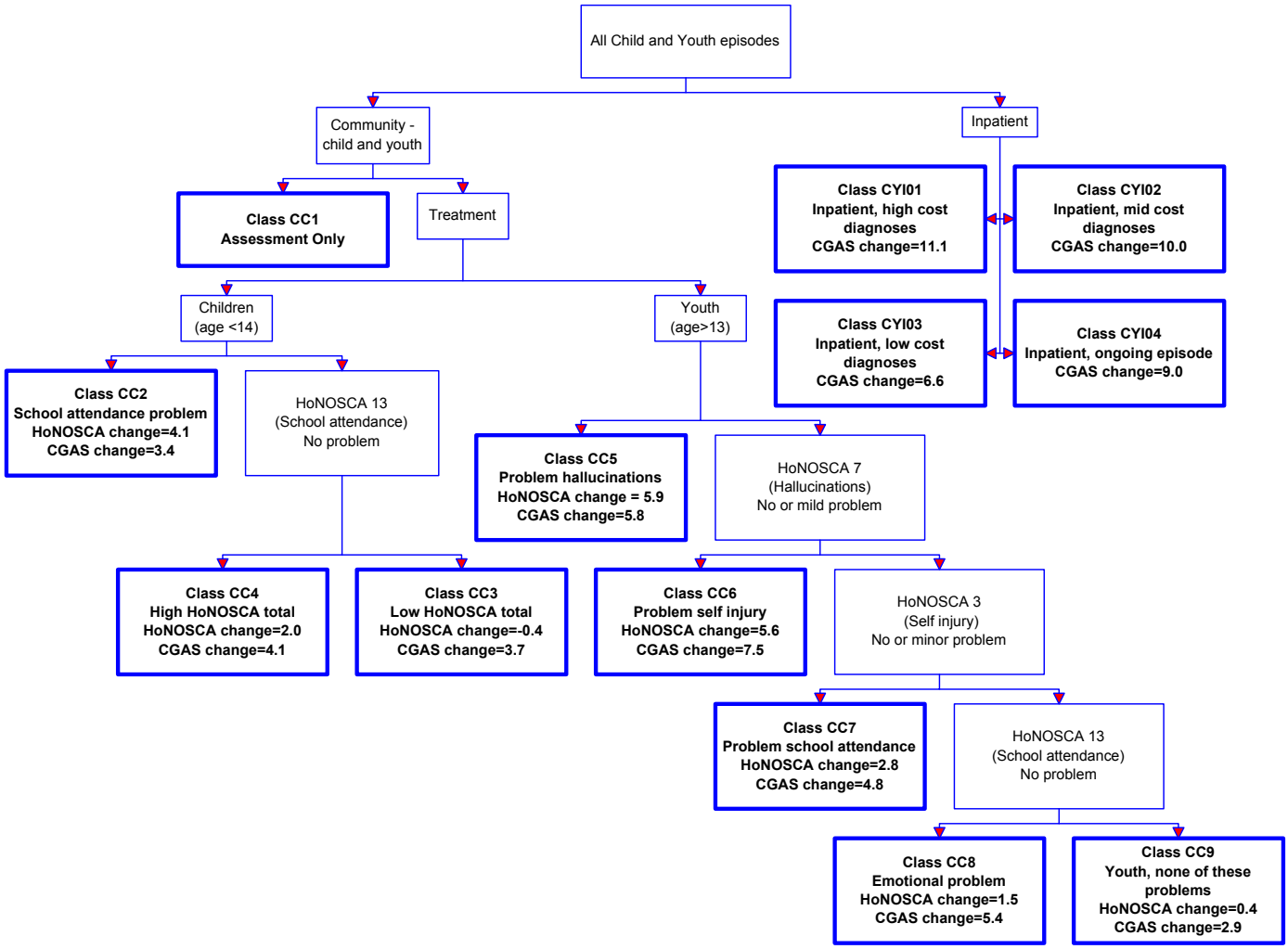


Figure 53. Change profile of the child and youth classes.



These figures show how an episode of care is assigned to a casemix class. The casemix classes are shown in bold. The amount of change for each class is included in each box. Details of the average start and end scores of each class can be found in Table 53, Table 54, Table 55 and Table 56 above.

The Casemix-Adjusted Relative Mean Improvement (CARMI) score as described in Chapter 10 (page 92) was used in this analysis to control for these differences. The CARMI is a way that the outcomes achieved by different mental health services can be fairly compared for benchmarking and quality improvement purposes because it takes into account the unique mix of consumers at each DHB.

11. Discussion of key findings and the implications for policy and practice

11.1 Compliance, completion and technical performance of the outcome measures

In total, six clinical measures were collected in this study:

- HoNOS / HoNOS65+ (Health of the Nation Outcome Scales)
- LSP-16 (an abbreviated version of the Life Skills Profile)
- Focus of Care (FoC)
- HoNOSCA (Health of the Nation Outcome Scales for Children and Adolescents)
- Children's Global Assessment Scale (CGAS)
- Factors influencing health status (FIHS)

Of these, four are outcome measures (HoNOS, LSP-16, HoNOSCA and CGAS) and two are measures intended to help explain the outcomes achieved (FoC and FIHS). A key question raised in the compendium casemix report is whether these measures are appropriate for routine collection. There are several issues.

The technical performance of all instruments was satisfactory and consistent with the technical performance reported in previous Australian studies. There are no technical reasons why any of the measures are not appropriate for routine collection.

Both the compliance and completion results for the HoNOS, the HoNOSCA and the CGAS measures suggest good data quality and clinical acceptability. The overall pattern of clinical ratings on these measures was similar to that found in the various Australian studies, suggesting that these measures can be used for international benchmarking purposes by the two countries.

Clinicians participating in the study continually provided feedback that they did not regard the LSP-16 as an acceptable measure. Likewise, consumer representatives also objected strongly to this measure. For this reason, the LSP-16 was discounted as a contender variable in the casemix classification. The poor compliance on this measure (22% of consumers were not rated at all on this measure) is consistent with this feedback. However, when it was completed, the completion rate of the LSP-16 was actually better than that achieved with the HoNOS.

Data quality on the Focus of Care measure was reasonable. However, this analysis found that only the Intensive Extended category of the Focus of Care measure was reliably associated with different ratings on the two adult outcome measures. As already reported in the compendium casemix report, the FoC variable works differently in the New Zealand context than it does in Australia. It is not clear whether these differences are due to differences in clinical practice or due to differences in the way that the definitions are interpreted in the two settings.

The results achieved using the various measures are related. For example, consumers with low ratings on the HoNOS also tend to have low ratings on the LSP-16. This also applies to the child and youth measures. Each of the instruments is capturing different, but related, domains and all are relevant to measuring consumer outcomes. None appear to be redundant.

The final, but important, comment on the instruments is that they are all clinician-rated instruments. In this regard, the study is no different to much of routine practice in which the assessment of outcomes is undertaken by health professionals as the 'experts'. It is important to recognise, however, that this approach is limited. The views of consumers and family/carers view are required for any comprehensive assessment of consumer outcomes in mental health care.

11.1.1 Implications for policy and practice

- From a technical perspective, the four consumer outcome measures – the HoNOS, the LSP-16, the HoNOSCA and the CGAS – appear suitable for routine collection in New Zealand.
- From a clinical and a consumer perspective, the appropriateness and acceptability of the LSP-16 warrants further consideration. If, after review, the LSP-16 is still not considered to be acceptable, an alternate measure of function will be required as the only current measure is subscale items in the HoNOS. Both mainstream and mental health specific measures of function should be considered.
- Further clinical consultation and research on the Focus of Care measure appears warranted with perhaps some refinement of the definitions of all classes – Acute, Functional Gain, Intensive Extended and Maintenance.
- Ongoing training in the measures adopted for routine use is critical and remains a major task as part of the workforce outcomes development and training program.
- The clinician-rated measures reported here need to be complemented by both consumer-rated and family/carer-rated measures.

11.2 Consumer outcomes

The results (summarised in Table 57) demonstrate that each of the instruments is sensitive to change over the course of an episode of mental health care.

Table 57. Summary of percentage change from episode start to episode end with each measure.

Measure	Average percentage change from episode start to episode end
HoNOS in inpatient settings	50%
HoNOS in community settings	11%
LSP-16 in community settings	6%
HoNOSCA in community settings	17%
CGAS in community settings	9%

In inpatient settings, the HoNOS was the only instrument that was used to measure change. It showed large changes, of the order of 50% of the initial total score, between the start and the end of episodes. Smaller changes, around 11%, were found in the community. Overall changes between start and end of episodes in the community on the LSP-16 were smaller than on the HoNOS – of the order of 6%. HoNOSCA assessments in the community indicated an overall change of just over two points, or about 17%. On average, CGAS assessments in the community improved five points, or 9%, between starts and ends of episodes.

Overall, the amount of change reported is related to the starting point. Consumers rated at the start as having a high level of severity or a low level of function generally show more improvement than those rated as having low severity or high function. This is not surprising, as those with high severity/low function have more room for improvement. But the implication is important – differences in the mix of cases between mental health services result in different mental health services achieving different outcomes. This needs to be taken into account when benchmarking the outcomes achieved by different mental health services.

The results suggest an important relationship between casemix and outcomes. Consumer outcomes differ between the NZ-CAOS casemix classes. The goal of the NZ-CAOS study was to define a set of classes based on their cost. This outcomes analysis has demonstrated that the classes are also useful as a way of classifying and understanding outcomes. The Casemix-Adjusted Relative Mean Improvement score as described in Chapter 10 (page 92) is a way that the outcomes achieved by different mental health services can be fairly compared for benchmarking and quality improvement purposes because it takes into account the unique mix of consumers at each DHB.

11.2.1 Implications for policy and practice

- The consumer outcome measures employed in this study are sensitive to change and appear suitable for use in benchmarking the performance of New Zealand mental health services.
- However, for benchmarking purposes, the outcomes achieved by different mental health services need to be casemix-adjusted to take into account differences between ethnicity groups and differences in the consumer profiles at the start of episodes.
- The Casemix-Adjusted Relative Mean Improvement (CARMI) score as described in Chapter 10 (page 92) is a way that the outcomes achieved by different mental health services can be fairly compared for benchmarking and quality improvement purposes because it takes into account the unique mix of consumers at each DHB.

11.3 Outcomes, ethnicity and social deprivation

Consistent with the results in the compendium casemix report, this study found differences between Māori, Pacific Island and other adult consumers. Unlike the casemix results, this outcomes analysis also suggests differences between child and youth consumers. The overall pattern is that Māori and Pacific Island consumers are rated as having higher levels of symptom severity and lower levels of functioning than other consumers.

However, the study also found that, overall, there are no ethnicity differences in the amount of change achieved during an episode of mental health care. Any differences are due to differences in the ratings recorded at the start of an episode of care. Once these differences are taken into account, all three high-level ethnicity groupings – Māori, Pacific Island and All Other – achieve about the same level of improvement. This is an important finding.

This study has identified a complicated set of relationships between ethnicity, social deprivation and consumer outcomes that cannot be readily explained. The overall finding is that consumers living in areas of relatively high levels of social deprivation have higher levels of severity and lower function than their counterparts living in areas of higher socio-economic status. But this is not the case for either the Māori or Pacific Island groups, where this pattern is reversed.

The study also found a weak, but definite, relationship between ethnicity, deprivation and both severity and functioning. Differences between ethnicity groupings in both severity and functioning are more marked when deprivation is least and less when deprivation is greatest.

Due to the large sample size, most differences in this analysis were statistically significant. In some cases, the differences are also clinically significant. However, the validity of these findings is open to question as they may be affected by, for example:

- Cultural misunderstandings on both sides;
- Language difficulties, and
- Traditional and cultural dynamics.

Taking a broader view, there are three possible reasons for the differences reported in this study:

- The results reflect real differences between consumers (for example, Māori and Pacific Island consumers may be entering the service system later in their episode of illness than non-Māori and non-Pacific Island consumers, and, when they do, have higher levels of system severity and lower levels of function than those who entered the service system earlier in their illness) and/or
- The results reflect problems with the cultural competency of the clinical workforce and/or
- The results reflect problems with the measures.

It was beyond the scope of this study to explain the reasons for the differences reported but there is nothing in the data to suggest the third possible explanation. Further consultation with the sector and further research is required to understand the reasons for the differences. The epidemiological study into the incidence and prevalence of different mental health problems in New Zealand populations being undertaken as one of the four priority areas defined under the Mental Health Research and Development Strategy (MHR&DS) should also inform this issue.

Irrespective of the reasons, the issue of cultural competency remains a challenge for the mental health sector and is already being addressed as part of the workforce outcomes development and training program. This forms a critical and long-term agenda for mental health services in New Zealand.

11.3.1 Implications for policy and practice

- Whether or not cultural competency has materially affected the outcomes and case mix results presented here, the issue of cultural competency remains a challenge for the mental health sector and will require ongoing attention as part of the workforce outcomes development and training program.
- Further work is required to understand the reasons for the differences in the profiles of Māori, Pacific Island and other consumers and the relationship between ethnicity, deprivation and consumer outcomes. The national epidemiological study may shed some light on this as would further research and consultation with consumers, carers and clinicians.
- It is suggested that this report be complimented by further reports that discuss possible interpretations of the data and implications for the relevant stakeholder groups.

12. References

Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990; 107: 238-246.

Blakely T (2004) *Income and ethnicity both critical in health funding: study: New research shows Mäori still dying younger -- regardless of income levels.* Pre-release findings of the New Zealand Census-Mortality Study. Press release March 2004.

Buckingham B, Burgess P, Solomon S, Pirkis J and Eagar K (1998) *Developing a Casemix Classification for Mental Health Services*. Commonwealth Department of Health and Family Services, Canberra.

College Research Unit, *Health of the Nation Outcome Scales: Report on Research*. 1996 London: Royal College of Psychiatrists.

Eagar K (2000) *An overview of health outcomes measurement in Australia*. Mental Health Outcomes Research in Aotearoa: Proceedings of the September 2000 Mental Health Research and Development Strategy Outcomes Conference and Other Developments in Mental Health Outcomes Research. Health Research Council of New Zealand: Auckland. p 142-148.

Gaines P, Bower A, Buckingham W, Eagar K, Burgess P. & Green J. (2003) *New Zealand Mental Health Classification and Outcomes Study: Final Report.* Health Research Council of New Zealand: Auckland.

Gaines P, Bower A & Buckingham W (2001*) Mental Health Classification and Outcomes Study: Study Resource Manual.* Health Research Council of New Zealand. Auckland.

Gowers S, Bailey-Rogers S, Shore A, Levine W. The Health of the Nation Outcome Scales for Child & Adolescent Mental Health (HoNOSCA). *Child Psychology & Psychiatry Review* 2000; 5: 50-56.

Rosenthal R, Rubin DB. A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* 1982; 74: 166-169.

Salmond C and Crampton P (August 2002) *NZDep2001 Index of Deprivation*. Department of Public Health, Wellington School of Medicine and Health Sciences, Wellington, New Zealand.

Stedman T, Yellowlees P, Mellsop G, Clarke R, Drake S, *Measuring consumer outcomes in mental health*. Canberra, A.C.T: Department of Health and Family Services, 1997.

Sulloway FJ. *Born to rebel: birth order, family dynamics, and creative lives*, New York: Vintage. 1997.

Tabachnick, B. G. & Fidell, L. S. *Using Multivariate Statistics*. 1996 New York: Harper Collins.

Trauer T, *Analysis of outcome measurement data from the four Victorian "Round One" agencies*. Unpublished report, 2003.

Trauer T, Callaly T, Little J, Shields R, Smith J. Health of the Nation Outcome Scales (HoNOS): results of the Victorian field trial. *British Journal of Psychiatry* 1999; 174: 380-38.

Trauer T. The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health* 1999; 8: 499-509.

13. Appendix Description of the consumer outcome measures

13.1 Health of the Nation Outcome Scales (HoNOS)

The HoNOS is a 12-item clinician-completed instrument developed by the Research Unit of the Royal College of Psychiatrists in the United Kingdom in the mid-1990s. It was designed specifically for use with people with a mental illness and is widely regarded as a general measure of severity of mental health disorder.

The 12 scales are:

1. Overactive, aggressive, disruptive or agitated behaviour
2. Non-accidental self-injury
3. Problem drinking or drug-taking
4. Cognitive problems
5. Physical illness or disability problems
6. Problems associated with hallucinations and delusions
7. Depressed mood
8. Other mental and behavioural problems
9. Problems with relationships
10. Problems with activities of daily living
11. Problems with living conditions
12. Problems with occupation and activities

The clinician rates the consumer on each of the scales in terms of their assessment and knowledge of the person's situation over the recent period, usually defined as the previous two weeks. All available information is used in making the ratings.

Each scale is rated using the following key:

0 No problem
1 Subclinical problem
3 Mild problem
4 Moderate problem
5 Severe problem
9 Don't know / unable to rate (not used in computing summary scores)

Assessments may be used as a profile of the 12 scales, or the total score, or subscale scores. The subscales identified by the Royal College are Behaviour (scales 1 to 3), Impairment (scales 4 and 5), Symptoms (scales 6 to 8), and Social (scales 9 to 12). However, an alternative set has been proposed (see Chapter 4 of this report).

Clinicians need to be trained in the use of the HoNOS. For trained staff the completion of a HoNOS rating takes about 5 minutes.

A different version of the HoNOS – the HoNOS65+ – was used to rate older consumers. This version includes the same items but the rating instructions have been modified to make specific provision for problems more common in older people (eg, night wandering). For the purposes

of the results reported throughout this report, HoNOS and HoNOS65+ data were grouped together.

13.2 Abbreviated Life Skills Profile (LSP-16)

The Life Skills Profile (LSP) is a 39-item instrument developed in Australia in 1989. It was designed as a measure of disability and functioning in persons with mental disorder. The 16-item version used in the CAOS study was developed in the mid-1990s for use in the Australian casemix study (Mental Health Classification and Service Costs, MH-CASC). More recently, another short form (LSP-20) has been developed.

The 16 items are:

1. Does this person generally have any difficulty with initiating and responding to conversation?
2. Does this person generally withdraw from social contact?
3. Does this person generally show warmth to others?
4. Is this person generally well groomed (e.g. neatly dressed, hair combed)?
5. Does this person wear clean clothes generally, or ensure they are cleaned if dirty?
6. Does this person generally neglect her or his physical health?
7. Is this person violent to others?
8. Does this person generally make and/or keep up friendships?
9. Does this person generally maintain an adequate diet?
10. Does this person generally look after and take her or his own prescribed medication (or attend for prescribed injections on time) without reminding?
11. Is this person willing to take psychiatric medication when prescribed by a doctor?
12. Does this person cooperate with health services (e.g. doctors and/or other health workers)?
13. Does this person generally have problems (e.g. friction, avoidance) living with others in the household?
14. Does this person behave offensively (includes sexual behaviour)?
15. Does this person behave irresponsibly?
16. What sort of work is this person generally capable of (even if unemployed, retired or doing unpaid domestic duties)?

The clinician rates the consumer on each of the scales in terms of their assessment and knowledge of the person's situation over the past three months. All available information is used in making the ratings.

Each scale is rated using the following key:

0 No problem
1 Mild problem
2 Moderate problem
3 Severe problem
9 Don't know / unable to rate (not used in computing summary scores)

Assessments may be used as a profile of the 16 scales, or the total score, or subscale scores. The subscales are Withdrawal (items 1, 2, 3 and 8), Self Care (items 4, 5, 6, 9 and 16), Compliance (items 10 to 12), and Antisocial (scales 7, 13 to 15).

Clinicians need little or no training in the use of the LSP-16. The completion of an LSP-16 rating takes about 5 minutes.

13.3 Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA)

The HoNOSCA is a 15-item clinician-completed instrument developed in the United Kingdom in the late 1990s. Modelled on the HoNOS, it was designed specifically for use with children and adolescents with a mental illness and is widely regarded as a general measure of severity of mental health disorder.

The 15 scales are:

1. Disruptive, antisocial or aggressive behaviour
2. Problems with over-activity, attention or concentration
3. Non-accidental self injury
4. Alcohol, substance or solvent misuse
5. Problems with scholastic or language skills
6. Physical illness or disability problems
7. Problems associated with hallucinations, delusions, or abnormal perceptions
8. Problems with non-organic somatic symptoms
9. Problems with emotional and related symptoms
10. Problems with peer relationships
11. Problems with self care and independence
12. Problems with family life and relationships
13. Poor school attendance
14. Problems with lack of knowledge or understanding about the nature of the child or adolescent's difficulties
15. Problems with lack of information about services or management of the child or adolescent's difficulties

The clinician rates the consumer on each of the scales in terms of their assessment and knowledge of the person's situation over the recent period, usually defined as the previous two weeks. All available information is used in making the ratings.

Each scale is rated using the following key:

0 No problem
1 Subclinical problem
3 Mild problem
4 Moderate problem
5 Severe problem
9 Don't know / unable to rate (not used in computing summary scores)

Assessments may be used as a profile of the 15 scales, or the total score (based on scales 1 to 13), or subscale scores. Clinicians need to be trained in the use of the HoNOSCA. For trained staff the completion of a HoNOSCA rating takes about 5 minutes.

13.4 Children's Global Assessment Scale (CGAS)

The CGAS is used as the measure of level of functioning for patients seen by specialist child and adolescent mental health services. The instrument was developed in the USA to provide a global measure of severity of disturbance in children and adolescents. The measure provides a single global rating only, on a scale of 1-100. Clinicians assign a score, with 1 representing the most functionally impaired child, and 100 the healthiest. The CGAS contains detailed behaviourally oriented descriptions at each anchor point that depict behaviours and life situations applicable to children and adolescents.

13.5 Factors Influencing Health Status (FIHS)

The Factors Influencing Health Status measure is a checklist of 'psychosocial complications' based on the problems and issues identified in the chapter of ICD-10 regarding Factors Influencing Health Status. It was developed specifically as part of the MH-CASC project (see above). The factors are:

1. Maltreatment syndromes
2. Problems related to negative life events in childhood
3. Problems related to upbringing
4. Problems related to primary support group, including family circumstances
5. Problems related to social environment
6. Problems related to certain psychosocial circumstances
7. Problems related to other psychosocial circumstances

The purpose of these items is to identify the degree to which the child or adolescent has 'complicating psychosocial factors' that require additional clinical input during the episode of care. They are important in understanding variations in outcomes, and are based on advice by clinicians that children or adolescents seen by specialist mental health services may present in the context of a range of circumstances which influence the person's health status but are not in themselves a current illness or injury. For example, the child may be severely affected by a history of sexual abuse but does not have a formal psychiatric diagnosis.

The FIHS comprises a simple checklist, requiring the clinician to indicate whether one or more factors are present. The seven categories of ICD codes included in the scale were selected on the basis of advice from clinicians about the most frequently occurring factors. The FIHS yields a total score ranging from 0 to 7.