

THE RELIABILITY AND VALIDITY OF NURSING HOME STANDARDS

John Braithwaite
Valerie Braithwaite
Diane Gibson
Miriam Landau
Toni Makkai

Canberra
Department of Health, Housing and Community Services
June 1991

CONTENTS

LIST OF FIGURES	iv
LIST OF TABLES	iv
ACKNOWLEDGMENTS	v
EXECUTIVE SUMMARY	vi
1. INTRODUCTION	1
The Nursing Home Regulation in Action Project	4
Types of Reliability and Validity	5
2. INTERNAL CONSISTENCY AND INTER-RATER RELIABILITY	9
Internal Consistency	9
Inter-Rater Reliability	11
Planning the Reliability Study	12
The Choice of Independent Rater	13
The Sampling Frame	14
The Level of Agreement Between Raters	15
Variation in Reliability Within the Sample	23
Changing the Rating Categories	24
Reasons for Disagreements	25
Summary	29
3. FACE AND CONTENT VALIDITY	31
Comprehensiveness	31
Clarity of the Standards	33
Difficulty in Rating the Standards Consistently	38
Desirability and Practicality of the Standards	39
Summary	42
4. CONCURRENT VALIDITY	43
Validation Against a Global Assessment of Quality	43
Director of Nursing Agreement With Team Ratings	46
Summary	51
5. COMPARISON WITH AMERICAN STUDIES	53
BIBLIOGRAPHY	65

LIST OF FIGURES

Figure 4.1:	Per cent of overall agreement of directors of nursing with the rating given them by the standards monitoring team	48
Figure 4.2:	Per cent of director of nursing agreement with the standard monitoring teams' rating of 'not met'	49
Figure 4.3:	Per cent of director of nursing agreement with the standard monitoring teams' rating of 'met in part'	49

LIST OF TABLES

Table 1.1:	Objectives and outcome standards for Australian nursing homes	1
Table 2.1:	Item-total correlations for the 31 standards	10
Table 2.2:	Per cent of overall agreement between the team and the independent rater	16
Table 2.3:	Ratings given by team and independent rater – after negotiation	19
Table 2.4:	Reasons for differences between team ratings and independent ratings at the initial rating stage	26
Table 3.1:	Evaluating the face and content validity of the standards	34
Table 4.1:	Correlation of standards monitors ratings with a rating of how the quality of care in the nursing home compares with other nursing homes	45

ACKNOWLEDGMENTS

We wish to acknowledge the support and cooperation of Australian nursing home industry associations, unions, consumer groups, state and federal governments, residents, staff and proprietors – all have given unflagging assistance to the completion of this ambitious study. Of 51 nursing homes and 30 standards monitoring teams asked to cooperate with the inter-rater reliability study reported herein, there was only one case where special circumstances led a nursing home to decline to cooperate in the study. In our larger random sample of nursing homes for the interview study, the refusal rate was only 4 per cent. This remarkable response is indicative of the level of support we have enjoyed.

The study is part of a larger study which has benefitted from substantial funding from a variety of sources – the Department of Community Services and Health, the Australian National University, the Australian Research Council, the American Bar Foundation, the Japanese-American Education Foundation, the Florida Department of Health, the University of Queensland and the University of Delaware. We gratefully acknowledge all of this support.

We would like to thank Professor David Ermann, Professor John Gardiner, Dr Kathryn Malec and Dr Colleen Phillips for critical comments on an earlier draft of the paper.

Finally, we wish to particularly thank Coralie Friend and Nancy Walke who did an excellent job on the most important part of this work, the inter-rater reliability study. It was a difficult job, a tense one at times, and throughout an organizationally difficult endeavour. They both handled all these difficulties with consummate skill, patience, ethics and professionalism.

EXECUTIVE SUMMARY

This study evaluates the reliability and validity of ratings of nursing home compliance with 31 Australian government standards. While doubts are raised about some standards, in general very strong support is found for the reliability of the standards and limited tests also support their validity. In all, the data are confronted with 19 reliability and validity tests.

Highlights at the level of total compliance scores obtained by adding compliance ratings for all 31 standards are:

- (a) Inter-rater reliability coefficients ranging from .93 to .96.
- (b) Validation against a global team rating of "quality of care" compared to other nursing homes of .64.
- (c) Validation against self-assessed ratings by the director of nursing of .88.

The validation results are very encouraging indeed and reliability is much stronger than that found in any of the American studies of the reliability of nursing home ratings by government inspectors.

While the study strongly indicates intra-state reliability, it does not demonstrate inter-state reliability. Indeed there are grounds for suspecting that inter-state reliability may be weak. The data suggest that inter-state reliability is achievable using these standards, even if it is not achieved.

While the study concludes that the 31 standards comprehensively cover the domain of quality of nursing home life, the standards monitoring process does not comprehensively collect the evidence necessary for valid ratings of some of the standards. The question of how to improve the comprehensiveness of data collection will be addressed in the final report of this consultancy.

For 29 of the 31 standards, over 95 per cent of directors of nursing thought they were clear. For all of the standards, at least 95 per cent of both directors of nursing and standards monitors thought that the standards were desirable. For 24 of the standards, over 90 per cent of both directors of nursing and standards monitors thought they were practical.

Some standards that a minority of either directors of nursing, proprietors or standards monitors objected to as unclear, undesirable or impractical turned out to be standards that appeared satisfactory, and in some cases outstanding, in terms of correlations with total compliance scores, inter-rater reliabilities, validation against the global "quality of care" assessment, and validation against the self-assessments of the director of nursing. Moreover, since the overwhelming majority of directors of nursing, proprietors, and standards monitors also agreed that all of these standards were clear, capable of being rated consistently, desirable and practical, there is no reason to reject any of these standards. Standards in this category were:

- 1.1 Residents enabled to receive appropriate medical care by a practitioner of their choice.
- 1.2 Residents enabled and encouraged to make informed choices about their individual care plans.
- 2.2 Residents are enabled and encouraged to maintain control of their financial affairs.
- 2.3 Residents have maximum freedom of movement.
- 3.1 The nursing home has policies developed in consultation with the residents to ensure freedom of choice.
- 4.1 Homelike environment.
- 7.1 Resident's right to take risks.

In other words, the above are all standards with some minor face validity problems. They are not, however, shown to have substantive problems of reliability and validity. Indeed some of them are among the very strongest standards on these criteria. They are standards that require more educative effort, not change to the standards themselves. Departmental officers point out that special educative effort has been made on some of these standards. The lesson may be that there are standards that require constant educative effort, otherwise clarity and commitment to them will slip back again.

Standards which do, on the other hand, have a substantive question-mark over their reliability and validity are:

- 1.3 All residents as free from pain as possible. (*Failure to systematically collect evidence for this standard*).
- 1.5 Residents enabled to maintain continence. (*A particularly strong standard in most respects. However, unreliability caused when teams do not require individualized continence management programs for a met rating*).

1.9 Sensory losses identified and corrected. *(In general a solid standard, but unreliability as a result of some raters emphasizing aspects of the environment (particularly lighting) while others do not).*

5.4 Nursing home free from undue noise. *(Reliable between raters on the day, but nursing homes complain teams can strike a bad day, explaining some unimpressive validation results for this standard).*

5.6 Nursing home practices support the resident's right to die with dignity. *(Profound problems with evidence-gathering to rate this standard validly).*

In no case do the substantive question-marks which hang over these latter standards seem devastating or fatal. There is indeed no standard that does not perform well on more than a few of our reliability and validity tests. Nevertheless, these latter standards are the ones where reform should be seriously considered. The consultants would expect to be making recommendations on the need for changes with regard to these standards in their final report, after interested parties have had an opportunity to comment on this report.

A troubling cause of unreliability in New South Wales arose where a single problem caused more than one outcome not to be met. New South Wales standards monitoring staff felt it was unfair to allow one problem to cause more than one standard to be put out of compliance. However, inconsistency arose from arbitrary choice of which of the adversely affected standards would and would not be rated met.

1. INTRODUCTION

Following a series of public scandals and parliamentary enquiries (McLeay Report, 1982; Giles Report, 1985) about the standard of care provided in nursing homes and hostels in Australia the Commonwealth government began a major reform of residential aged care in 1986. In 1987 a Commonwealth/State Working Party on Nursing Home Standards was convened. Its purpose was to consult with all parties in the nursing home industry and to formulate a set of workable standards that could be used to assess the quality of care in Australian nursing homes.

The Working Party settled on 31 Standards which had to be satisfied if seven objectives were to be met. These standards are listed by objective in Table 1.1. The purpose of this report is to evaluate the reliability and validity of these standards. To complement the general prescriptions in each of the standards, the *Nursing Home Standards Monitoring Guidelines* (1987) set out a number of more detailed things to "look for" under each standard. However, this document emphasises that these are only guidelines, 'not hard and fast rules, and as such, standards monitoring staff are expected to use their judgment in the application of the guidelines' (p.2).

Table 1.1: Objectives and outcome standards for Australian nursing homes

Objective 1:	Health care: Residents' health will be maintained at the optimum level possible
Standard 1.1	Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed
Standard 1.2	Residents are enabled and encouraged to make informed choices about their individual care plans
Standard 1.3	All residents are as free from pain as possible
Standard 1.4	All residents are adequately nourished and adequately hydrated.
Standard 1.5	Residents are enabled to maintain continence.
Standard 1.6	Residents are enabled to maintain, and if possible improve, their mobility and dexterity.
Standard 1.7	Residents have clean healthy skin consistent with their age and general health.
Standard 1.8	Residents are enabled to maintain oral and dental health.
Standard 1.9	Sensory losses are identified and corrected so that residents are able to communicate effectively.
Objective 2:	Social independence: Residents will be enabled to achieve a maximum degree of independence as members of society.
Standard 2.1	Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts.
Standard 2.2	Residents are enabled and encouraged to maintain control of their financial affairs.
Standard 2.3	Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.
Standard 2.4	Provision is made for residents with different religious, personal and cultural customs.
Standard 2.5	Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.

continued

Objective 3: Freedom of choice: Each resident's right to exercise freedom of choice will be recognised and respected whenever this does not infringe on the rights of other people.

- Standard 3.1 The nursing home has policies which have been developed in consultation with residents and which:
– enable residents to make decisions and exercise choices regarding their daily activities
– provide an appropriate balance between residents' rights and effective management of the nursing home
– and are interpreted flexibly taking into account individual resident needs
- Standard 3.2 Residents and their representatives are enabled to comment or complain about conditions in the nursing home.

Objective 4: Homelike environment: The design, furnishings and routines of the nursing home will resemble the individual's home as far as reasonably possible.

- Standard 4.1 Management of the nursing home is attempting to create and maintain a homelike environment.
- Standard 4.2 The nursing home has policies which enable residents to feel secure in their accommodation.

Objective 5: Privacy and dignity: The dignity and privacy of nursing home residents will be respected.

- Standard 5.1 The dignity of residents is respected by nursing home staff.
- Standard 5.2 Private property is not taken, lent or given to other people without the owner's permission.
- Standard 5.3 Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private.
- Standard 5.4 The nursing home is free from undue noise.
- Standard 5.5 Information about residents is treated confidentially.
- Standard 5.6 Nursing home practices support the resident's right to die with dignity.

Objective 6: Variety of experience: Residents will be encouraged and enabled to participate in a wide variety of experiences appropriate to their needs and interests.

- Standard 6.1 Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.

Objective 7: Safety: The nursing home environment and practices will ensure the safety of residents, visitors and staff.

- Standard 7.1 The resident's right to participate in activities which may involve a degree of risk is respected.
- Standard 7.2 Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors.
- Standard 7.3 Residents, visitors and staff are protected from infection and infestation.
- Standard 7.4 Residents and staff are protected from the hazards of fire and natural disasters.
- Standard 7.5 The security of buildings, contents and people within the nursing home is safeguarded.
- Standard 7.6 Physical and other forms of restraint are used correctly and appropriately.

*Source: Commonwealth/State Working Party (1987) **Living in a Nursing Home** (Canberra: Australian Government Publishing Service).*

The standards are rated by standards monitoring teams of generally two or three members (never fewer than two and rarely more than three), one of whom is always a registered nurse. These are mostly employees of the Commonwealth Department of

Community Services and Health, though in New South Wales state government teams have done standards monitoring visits, and in other jurisdictions state or local government officers have regularly joined Commonwealth officers on the teams. For some of the New South Wales homes in our sample, directors of nursing from the industry also acted as team members. The team usually spends a day (but sometimes longer) to gather the information it needs to formulate initial ratings on the 31 standards. It then returns to base to assemble this information; it meets as a team to discuss the positives and negatives under each standard this information implies and to agree on an interim team rating. Quite often this process reveals a need to gather further information from the nursing home. This done, the team returns to the nursing home to discuss their preliminary findings with management of the facility. The nursing home will often provide further information at this point to support their reactions to the preliminary findings which, in turn, may involve the team in further data collection at the facility to check these claims. On return to base, the team will then despatch a final written report with final ratings on the 31 standards. Since it is these ratings which are the focus of concern in this report, we will not discuss the follow-up action that is then taken in an attempt to get the nursing home into compliance with the standards (see Braithwaite *et al.*, 1990).

The philosophy underlying the new standards was that they should be outcome-oriented and resident-centred. This was an attempt to make a sharp break with the nursing home inspection practices of the past. Prior to 1987, nursing home inspections in Australia were very much based on verifying inputs (for example, staffing levels, availability of clean linen, presence of fire alarms), the presence of which was assumed to guarantee a capacity to deliver adequate care. There was not a systematic evaluation of whether quality care was actually delivered and whether residents enjoyed a satisfactory quality of life. The new standards were intended to shift the regulatory process in just this direction.

Moreover, outcomes were to be assessed by placing considerable reliance on what residents said about their quality of life and care during resident interviews. Associated with the shift to a resident-centred methodology was a new prominence for resident rights and the social aspects of care to complement the former preoccupation with the medico-nursing aspects of care. The "soft" nature of some of the new standards this introduced, the apparent subjectivity of concepts like privacy and dignity, caused many critics from the industry and the health care professions, and indeed the Department of Community Services and Health's own legal advisers, to question the reliability of any inspection process based on such standards. Hence the importance of a systematic empirical assessment of the reliability of the standards.

The Nursing Home Regulation in Action Project

The reliability and validity study reported in this document is part of an on-going, large scale assessment of the contribution of the standards monitoring process to quality of care and quality of life for nursing home residents. The *Nursing Home Regulation in Action Project* commenced in 1987. It incorporates quantitative and qualitative approaches to the collection of data on the standards monitoring process, a longitudinal research component, and cross-national comparisons in an attempt to gain a thorough understanding of the regulatory process in Australian nursing homes.

The qualitative component of the study has involved observations of standards monitoring events at nursing homes in all Australian states and territories, in the United States of America and in the United Kingdom. Interviews have been conducted at different stages of the evolution of the Australian process with standards monitors, their superiors in state departments and in state and head offices of the Commonwealth department, proprietors, directors of nursing, nursing home staff, residents, industry associations, unions and consumer groups. Training courses, rating sessions, nursing home staff meetings, residents' meetings and a variety of other meetings and conferences were also attended.

The major quantitative component comprises two stages of data collection. The first wave data collection was completed in 1990. This involved a structured interview with directors of nursing after the standards monitoring process was completed, the ratings on that visit, and a questionnaire completed by the standards monitoring team in each of 410 nursing homes in Queensland, New South Wales, Victoria and South Australia. The second phase involves a follow-up questionnaire sent to directors of nursing on completion of a second visit by the standards monitoring team, and should be finalized by late 1991.

The analysis of reliability and validity presented in this report draws on both the quantitative (Chapters 2, 3, 4 and 5) and qualitative (Chapter 5) components of the project as described above, preliminary findings for which are reported elsewhere (Braithwaite *et al.*, 1990). The report also presents data from an inter-rater reliability study of 50 homes in New South Wales and Victoria (Chapter 2). Independent raters visited the nursing home on the same day as the standards monitoring team; agreement between the independent rater and the team was then measured at three points in time, yielding three measures of inter-rater reliability. Further details of the methodology adopted are presented in Chapter 2.

Types of Reliability and Validity

The reliability of a measure refers to whether one gets the same result when the measure is administered under varying conditions of administration. The validity of a measure refers to the degree to which the measure measures what it is supposed to measure. A measure can be reliable without being valid (for example, a ruler with an incorrect scale will give the same invalid length each time it is used); but a measure cannot be valid without being reliable. In this study, 19 different tests have been made of the reliability and validity of the 31 standards. While this is a more comprehensive approach to the reliability and validity of regulatory decision-making than has been attempted before in the literature for any domain of regulation, it is still far from an exhaustive approach. Three of our tests are of what we will define as internal consistency, six are of inter-rater reliability, eight of face and content validity and two of concurrent validity.

Internal consistency: Internal consistency concerns whether the different components of a measure are all contributing to measuring the same thing. To assess internal consistency, we can look at the results of factor analyses of the compliance ratings given on the 31 standards. Another way to approach the internal consistency of the total compliance score from the 31 standards is to split the standards into two equal groups of standards to ascertain whether scores on the two split halves of the measure are highly correlated. An alpha reliability coefficient in effect takes the average of the split-half correlations for all possible ways of splitting the standards into two groups. We can calculate an alpha reliability coefficient for the scale we obtain by adding together all 31 standards, or subsets of those standards. Finally, we look at the correlation of each standard with the total compliance score for all other standards combined to diagnose whether there are standards which measure something unrelated to the others.

Inter-rater Reliability: By far the greatest reliability doubt in the industry is that the ratings you get depend on which team you get. Some teams are perceived as tougher than others; some are perceived as more sophisticated than others; and also objectively teams vary in their size, experience and disciplinary backgrounds. The most important part of our study, therefore, was to put an extra nurse into 50 nursing homes at the same time as the visit of the government team. This nurse independently rated the nursing home on the 31 standards. This enabled the calculation of an inter-rater reliability coefficient. The correlation coefficient between the government team's rating and the reliability rater was calculated at three points in time. First, the reliability coefficient was calculated after the team had met and settled on its initial rating. This was the only fully independent reliability

test – where each side reached its conclusions without being subject to influence from the other side. Then the team met with the independent rater so that both sides could swap the information they had independently collected and could argue with each other as to why each disagreed with the other's ratings. Both sides then had the opportunity to change their ratings and a second inter-rater correlation coefficient was calculated on the revised ratings. A third coefficient was then calculated after the team had an opportunity to visit the nursing home again to discuss the ratings with the director of nursing and/or proprietor. Teams passed on to the independent rater additional information supplied by the nursing home; on occasion this resulted in either the team or the independent rater or both changing ratings. Hence we have three separate inter-rater reliability coefficients calculated at three different points in time during the regulatory process. In fact, there are six coefficients because we asked both sides to rate each standard in two different ways at each point in time. The details of this will be discussed below.

Face and Content Validity: Face validity is usually regarded as the least important form of validity. It addresses whether the measure makes sense to those who are being assessed. For regulatory measurement, nevertheless, face validity has unusual importance because regulatory ratings which make no sense to the industry will be denied legitimacy by the industry, subjected to legal challenge, and will usher in an adversarial regulatory culture. 'Content validation involves essentially the systematic evaluation of the test content to determine whether it covers a representative sample of the behaviour domain to be measured' (Anastasi, 1982: 131). The first approach to evaluating the content validity of the 31 standards is to assess their *comprehensiveness* by relying on the qualitative fieldwork aspect of our study. The authors observed standards monitoring visits, team discussions and negotiations with nursing homes to see if problems of poor quality care or poor quality of life for nursing home residents were being discovered (or neglected) that could not be comfortably fitted into the standards. Discussions were also held with all key participants in the industry – nursing home proprietors, directors of nursing, staff, residents, consumer groups, unions, professional groups and government regulators – to solicit their opinions on ways in which they felt the standards were insufficiently comprehensive.

The face and content validity of the standards was also assessed in a more quantitative way by asking directors of nursing whether each of the 31 standards was clear to them (and whether standards monitors thought the standards were easy to rate) and if not, why not. Directors of nursing and standards monitoring teams were also asked whether they thought each of the standards was desirable and practical, and if not, why not. In a more

unstructured way, proprietors of our sample nursing homes were also asked which standards they were critical of, and why. We will argue that these tests lead us to an understanding of both the face and content validity of the standards.

Concurrent Validity: Concurrent validity is concerned with whether a rating correlates positively with other measures of the same or a similar concept assessed at the same point in time. Predictive validity, the other form of criterion-related validation, assesses whether the ratings successfully predict scores on a related test at a different point in time. No test of predictive validity is included in this study. The first approach to concurrent validation was to correlate the ratings on each of the 31 standards with a global assessment of the quality of care in the nursing home by the team. The second was to correlate team ratings with self-ratings of the home's compliance with each of the standards by the director of nursing.

Conclusion

The reason we test reliability and validity in so many different ways is that there is no perfect measure against which to test the standards. There does not exist in the world some perfectly reliable and valid measure of nursing home quality of care or quality of life. So we test the standards against many imperfect measures. In doing so, we make no assumptions that the ratings of independent raters or of directors of nursing are necessarily better or worse than those of standards monitoring teams. We only make the assumption that a failure of these alternative measures to correlate with team ratings would put their validity in doubt. We also assume that repeatedly finding positive correlations against alternative measures with different sources of error means that there is quite a lot of true variance in the standards – that the repeated positive correlations cannot be explained by shared error variance.

Finally, it must be said that reliability and validity is not everything with regulatory standards, not even the most important thing. The most important role of standards is as mechanisms for focusing dialogue between the government and the industry, between consumer advocates and the industry, between directors of nursing and staff, between consultants and staff, between residents and staff and between educators and students about how to improve the quality of life in nursing homes.

To illustrate, the consultants suspect on the basis of the data presented herein that standard 5.6, "Nursing home practices support the resident's right to die with dignity", is

not rated validly and perhaps cannot be rated validly. This does not mean that we think the standard should be scrapped. Indeed we see great virtue in the existence of this standard as a reminder to nursing homes that this is an important issue. We see virtue in standards monitoring teams underlining with directors of nursing that this is an issue that they cannot put off, that they must be able to account for how they attend to it systematically and sensitively (for example, through staff training). Indeed, it might be better to be satisfied with this kind of dialogue than to pursue, in a myopic quest for consistency, tight documentation of something that cannot be documented.¹

Hence it is important to keep a report which deals exclusively with the reliability and validity of regulatory standards in perspective. The reliability and validity of standards is an important issue, but it is only a small part of what is required to evaluate comprehensively a regulatory process.

¹There are so many issues to be sensitive to here, that it would be difficult to cover them with any set of documentation guidelines, as illustrated by the following comment from a director of nursing: "It's the usual Australian custom to remove jewellery from a deceased person. When one of our residents who was a Greek lady died we removed the wedding ring from her finger and the family were devastated."

2. INTERNAL CONSISTENCY AND INTER-RATER RELIABILITY

This chapter is divided into two major sections. The first section assesses the internal consistency of the standards. That is to say whether different standards are all contributing to measuring a global concept called "compliance". This is undertaken by using factor analytic methods and reliability coefficients. The second section focuses on the inter-rater reliability of the standards. The data used in this section are taken from a study designed specifically to address this issue. The first part of this section describes the methodology used to collect the data; then follows an examination of the extent of agreement on the standards between the raters and the standards monitoring teams; finally the reasons for disagreement between the independent raters and the teams in their assessments of the nursing homes are discussed.

Internal Consistency

In this report, we will only summarize the work we have done on the internal consistency of the standards as this has been presented in more detail in our first report (Braithwaite *et al.*, 1990). What that report shows is that the standards fall a long way short of being pure measures of the objectives in which they are grouped. When a factor analysis is undertaken on the whole set of 31 standards, the three-factor structure obtained does not correspond to the seven objectives under which the standards are listed in Table 1.1. Information about compliance with a single standard tends to tell us relevant things about how well the nursing home is meeting more than one, and often several, of the seven objectives. Therefore, it does not make sense to form measures of compliance with the objectives by adding ratings for all the standards listed under each objective in Table 1.1.

Rather than clustering neatly around different objectives, all standards tended to be loosely interrelated and could be summated to give an overall compliance score. Put another way, standards tend to be related to other standards not only from the same objective but also from other objectives. Table 2.1 shows the item-total correlations for the 31 standards – the correlation between the rating on that standard and the sum of the ratings on the other 30 standards. All standards have credible item-total correlations – the lowest being 0.28 for standard 2.5, "Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens."

Table 2.1: Item-total correlations for the 31 standards

	Item-total correlation
Objective 1: Health care	
1.1 Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed.	.54
1.2 Residents are enabled and encouraged to make informed choices about their individual care plans	.64
1.3 All residents are as free from pain as possible.	.40
1.4 All residents are adequately nourished and adequately hydrated.	.60
1.5 Residents are enabled to maintain continence	.68
1.6 Residents are enabled to maintain, and if possible improve, their mobility and dexterity.	.56
1.7 Residents have clean healthy skin consistent with their age and general health.	.54
1.8 Residents are enabled to maintain oral and dental health.	.45
1.9 Sensory losses are identified and corrected so that residents are able to communicate effectively.	.47
Objective 2: Social independence	
2.1 Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts	.53
2.2 Residents are enabled and encouraged to maintain control of their financial affairs.	.37
2.3 Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.	.56
2.4 Provision is made for residents with different religious, personal and cultural customs	.49
2.5 Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.	.28
Objective 3: Freedom of choice	
3.1 The nursing home has policies which have been developed in consultation with residents and which:	
– enable residents to make decisions and exercise choices regarding their daily activities	
– provide an appropriate balance between residents' rights and effective management of the nursing home	
– and are interpreted flexibly taking into account individual resident needs.	.70
3.2 Residents and their representatives are enabled to comment or complain about conditions in the nursing home	.67
Objective 4: Homelike environment	
4.1 Management of the nursing home is attempting to create and maintain a homelike environment	.65
4.2 The nursing home has policies which enable residents to feel secure in their accommodation	.60
Objective 5: Privacy and dignity	
5.1 The dignity of residents is respected by nursing home staff.	.61
5.2 Private property is not taken, lent or given to other people without the owner's permission.	.60
5.3 Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private	.56
5.4 The nursing home is free from undue noise	.45
5.5 Information about residents is treated confidentially.	.45
5.6 Nursing home practices support the resident's right to die with dignity.	.41
Objective 6: Variety of experience	
6.1 Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.	.50
Objective 7: Safety	
7.1 The resident's right to participate in activities which may involve a degree of risk is respected.	.59
7.2 Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors	.62
7.3 Residents, visitors and staff are protected from infection and infestation.	.57
7.4 Residents and staff are protected from the hazards of fire and natural disasters.	.50
7.5 The security of buildings, contents and people within the nursing home is safeguarded.	.39
7.6 Physical and other forms of restraint are used correctly and appropriately.	.52

It is interesting and important to note from Table 2.1 that some of the standards which we will see in the next chapter to be more controversial standards are found among the five standards which are the best predictors of overall compliance – 1.2, resident participation in individual care plans; 1.5, maintaining continence (also concerned with individualized care); 3.1, policies developed in consultation with residents to enable freedom of choice and flexibly take account of individual needs; 3.2, residents enabled to comment and complain; and 4.1, homelike environment. What the standards which are the best predictors of compliance have most in common is a focus on individualization (as opposed to institutionalization) and resident participation.

A standard way of assessing the internal consistency of a set of items is to calculate an alpha reliability coefficient (Aiken, 1985: 88). Essentially, what this does is calculate the average split-half reliability coefficient for all possible ways of splitting the 31 standards into two equal groups. The issue is whether the compliance score we get from one half of the standards is the same as that we get from the other half. If it is, the overall measure of compliance is internally consistent. The alpha coefficient for the total compliance score is .90, indicating that we do have this kind of internal consistency.

The important conclusion from the work on internal consistency is that no standard correlates so highly with any other standard as to make it redundant, while all standards have sufficient correlation with the other standards to make it sensible to add ratings on all standards to give a total compliance score. Henceforth in this report we will therefore evaluate both the reliability and validity of each standard and the reliability and validity of the total compliance score. The total compliance score is calculated by summing each of the individual ratings resulting in a scale from 0 to 31. Where 0 means no compliance and 31 means complete compliance. In calculating this scale, a met rating on a particular standard attracts a score of 1, a met in part or action required rating a score of 0.5 and a not met or urgent action required rating a score of 0.

Inter-Rater Reliability

A frequent comment from directors of nursing and proprietors is that the ratings you get depends on whether the standards monitoring team that visits your home is a tough one or a soft one, a knowledgeable one or an inexperienced one. We sought to assess inter-rater reliability by generating independent ratings of the same home at the same time as the official standards monitoring team was in the nursing home.

Planning the Reliability Study

Inter-rater reliability studies are extremely rare in the literature on regulatory inspectorates. In fact, nursing home inspection in the United States and Australia are the only areas where we are aware of any having been conducted. There are two reasons for this. First, government agencies are normally fearful of conducting studies which might show that their judgements about compliance with the law are arbitrary and capricious. Second, they are difficult to do – expensive, logistically a nightmare, and intrusive for the organizations being inspected. Consequently, it is normally impossible to get anyone to fund a reliability study and difficult to get cooperation from either the government inspectorate or the industry.

In our initial grant application to the Department of Community Services and Health, we did not raise the desirability of conducting an inter-rater reliability study. We felt that it might be so threatening to them that it might cause rejection of the whole idea of the study. Our plan was, once the study was approved, to work on persuading the Department that an inter-rater reliability study was needed to really put the standards to the test. But we did not push our luck by asking the department to fund it. Instead, we offered to submit grant applications to outside funding agencies such as the Australian Research Council, who we felt would see the scientific value of the study. Three unsuccessful applications later, we found that this confidence was misplaced; part of the problem was that reviewers and funding agencies did not believe we could successfully complete the study. So we proceeded with a scaled-down version of the study with Australian National University funds. Once this commitment was made, the Department of Community Services and Health came to the party to top this up with their funds, so the study could proceed at the level we had initially proposed.

An important feature of making the study acceptable to the industry was to have the independent rater in the nursing home on the same day as the official standards monitoring team. In most cases, the independent rater spent some extra time in the home before the arrival or after the departure of the team, including occasional spillover to a second day. However, most of the rater's time in the home was the same time as the team was in the home, so the rater did not cause a great deal of extra disruption to the home. From a methodological point of view, there are both positives and negatives to this approach. The negative is that guidelines had to be enforced to prevent the team and the independent rater from communicating with each other in any way about the standards monitoring visit during the visit. While this turned out to be less of a problem than expected, there doubtless

were occasions when the team noticed that the independent rater was spending a lot of time checking something in one particular room, for example, and this was a cue to pay attention to this room. However, the methodological disadvantages of having the independent rater visit the home the next day seem more profound. First, the next day the independent rater would in some senses see a different nursing home – a nursing home coping with a hot day rather than a cold day, a different menu in the kitchen, even some different staff. But most crucially, it would be a nursing home one day after a standards monitoring visit, a nursing home that had perhaps discovered from the team the previous day that its toilets were insufficiently clean – that had insisted that its cleaning staff rectify this problem the next morning. In fact, many of the problems of non-compliance discovered in standards monitoring visits are rectified within 24 hours. It follows that different results from the reliability rater the next day may not reflect unreliability but rather that the nursing home's compliance had actually changed.

The Choice of Independent Rater

A second way of making the reliability study minimally disruptive was to use a single independent rater rather than an entire second team of independent raters. This also had the advantage of keeping down the cost of the study. These two advantages of a single independent rater are counterbalanced by a considerable methodological disadvantage. If one finds disagreement between the two raters (the team and the individual), one cannot be sure that this is not because the team has more eyes and ears for discovering problems than the single rater. Happily and surprisingly, it turned out that we found high agreement between the two raters. Hence we did not suffer the effects of this methodological limitation. That is, we did not have to write a conclusion that our finding was unreliability, but that we could not be sure that this conclusion was right. We are not in the position of having to say that there is no way of knowing whether our conclusion of unreliability is because we relied on the judgments of a single rater. This is because our conclusion is one of reliability rather than unreliability.

Some compensation for the "one set of eyes and ears" deficit was made by allowing the independent rater to stay in the home a little longer than the team – to stay, in fact, until she was satisfied that she had all the information she needed to feel confident about her ratings. Moreover, we were then able to check whether any unreliability was due to the independent rater seeing things that happened at a time after the official team had left the home.

Finding suitable independent raters was difficult. The independent rater had to be a registered nurse so that she could rate all standards, including those that only nurses were qualified to rate. She had to have been fully trained as a standards monitor and to have had experience at the job of standards monitoring. But the most difficult problem of all was that she had to be someone with a flexible timetable that could accommodate erratic part-time work – who could fit in with the ever-changing movements of standards monitoring teams.

It was decided to conduct the reliability study in the wider Sydney and Melbourne regions as defined in our larger study (Braithwaite *et al*, 1990). We quickly found a independent rater in New South Wales with these qualifications. She was a former director of nursing from the industry who conducted training courses for directors of nursing on the new standards. As part of her preparation for this work, she had taken the training course for standards monitoring teams. At a later date, when the Department of Community Services and Health suffered an acute staffing shortage, she was seconded to work for a period with the department as a standards monitor. She therefore had the advantage of being both experienced as a standards monitor, yet independent of the department, and respected by the industry for her independent status.

A person with this independent status and the relevant qualifications did not exist in Melbourne. Instead, a senior registered nurse in a managerial position with the standards monitoring program who had both extensive experience in aged care, and in acting as a standards monitor, was used as the independent rater. Hence, the official team and the reliability rater were not equal. The team had the advantage of two or three heads being better than one and of extra sets of eyes and ears on site. The independent rater had the advantage of being able to stay in the facility longer (should she feel a need to) and the advantage of being rather more experienced and qualified for the job than the average team member.

The Sampling Frame

Twenty-five nursing homes were included in the New South Wales sample, 25 in Victoria. Half the homes were required to be above the median size of homes in that state (51 beds in NSW, 32 in Victoria) and half below the median. The independent raters were also required to meet quotas of non-profit and for-profit homes in proportion to their numbers in each state. They were also required to go on visits with all active Commonwealth teams in their state, to do at least two visits with each team, and to strive for an approximately equal number of visits with each team. This in fact became a messy criterion as resignations, illnesses and rotations between teams undermined the integrity of the definition of when a

team became a different team. For all the messiness, the selection was done in a way that leaves us confident that the independent raters spread themselves fairly evenly across all standards monitors operating in both states for the period of the study. In New South Wales, if we count as a separate team every case in which even one member of the team changed, there were 16 teams in the study, though in reality there were 9 core teams in the study, with others occasionally being added to this core. On this criterion, there were 14 different teams in the Victorian study from 9 core teams. Overall, then we have data from 50 homes for 30 teams which were in some way different from all the other teams. The visits occurred during the period November 1989 to August 1990.

Recourse to quota rather than random sampling was necessary because the Department already had its scheduling of visits driven by the proportionate stratified sampling of nursing homes for our wider study. Even had this not been true, a random sample selected especially for this study would have been impossible because of the inevitability of two sampled homes being visited at the same time by different teams and the impossibility of the independent rater always being able to drop the other commitments in their lives in response to the timetabling of the teams. We will see later that reliability results do not vary significantly by type of nursing home or state. Also in our wider study of nursing homes, randomly sampled homes were not found to be different on compliance-related variables from the non-randomly selected homes (Braithwaite *et al.*, 1990). There are therefore no strong reasons for suspecting that sampling methodology should have masked effects on these reliability results.

Within the discretion allowed by their quotas, independent raters were instructed to be especially on guard against a bias toward "easy" nursing homes. They were told, "if you have to err, err on the side of homes which are more likely to be problem homes, because these will be homes which give you more opportunities to disagree with the team". As it turned out, independent raters did err quite significantly on the side of homes with more problems. While the average number of met ratings for all homes in our wider study was 23, the average number of met ratings for all homes in the inter-rater reliability study was 18.

The Level of Agreement Between Raters

Agreement between the independent rater and the team was measured at three points in time. After the team had completed its visit, the team would meet (normally the next day) to discuss as a team the positives and negatives observed on each standard and to agree on initial ratings. Soon after, they would meet with the independent rater to compare initial

ratings. The per cent of agreement between these initial (totally independent) ratings are provided in the first column of Table 2.2. Then they would discuss why they obtained different ratings when their ratings were in fact different. After this discussion, one side or the other would sometimes decide that their ratings were wrong. On occasion the combining of their information even caused them both to think that they were wrong. This generated the "after conferring" ratings, our second measure of agreement which is represented in the second column of Table 2.2. After the team had been back to the nursing home for a further visit to advise the nursing home of their ratings (giving the nursing home an opportunity to provide further information which might change them), the team passed this information on to the independent rater. Both sides then had the opportunity to change their ratings again in light of the feedback from the nursing home. These final ratings were the basis for calculating agreement in the third column in Table 2.2. At each stage, the reasons for disagreement and changes of heart were recorded.

Table 2.2: Per cent of overall agreement between the team and the independent rater^a (n=50)

	Standard	Initially	After Conferring	After Negotiation
1.1	Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed	84	90	92
1.2	Residents are enabled and encouraged to make informed choices about their individual care plans	90	92	90
1.3	All residents are as free from pain as possible	90	94	94
1.4	All residents are adequately nourished and adequately hydrated	90	92	94
1.5	Residents are enabled to maintain continence	78	86	86
1.6	Residents are enabled to maintain, and if possible improve, their mobility and dexterity	90	96	96
1.7	Residents have clean healthy skin consistent with their age and general health	98	98	98
1.8	Residents are enabled to maintain oral and dental health	96	96	96
1.9	Sensory losses are identified and corrected so that residents are able to communicate effectively.	84	86	88
2.1	Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts.	90	94	96
2.2	Residents are enabled and encouraged to maintain control of their financial affairs.	94	94	96
2.3	Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.	94	98	94
2.4	Provision is made for residents with different religious, personal and cultural customs.	94	98	96
				Continued

	Standard	Initially	After Conferring	After Negotiation
2.5	Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.	90	94	98
3.1	The nursing home has policies which have been developed in consultation with residents and which: regarding their daily activities; – provide an appropriate balance between residents rights and effective management of the nursing home; – and are interpreted flexibly taking into account individual resident needs.	88	92	92
3.2	Residents and their representatives are enabled to comment or complain about conditions in the nursing home.	84	94	90
4.1	Management of the nursing home is attempting to create and maintain a homelike environment.	88	94	94
4.2	The nursing home has policies which enable residents to feel secure in their accommodation.	86	90	92
5.1	The dignity of residents is respected by nursing home staff.	92	98	98
5.2	Private property is not taken, lent or given to other people without the owners permission.	96	98	98
5.3	Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private.	88	94	94
5.4	The nursing home is free from undue noise.	94	96	92
5.5	Information about residents is treated confidentially.	90	96	96
5.6	Nursing home practices support the residents right to die with dignity.	96	98	98
6.1	Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.	92	94	94
7.1	The residents right to participate in activities which many involve a degree of risk is respected.	94	96	94
7.2	Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors.	82	92	90
7.3	Residents, visitors and staff are protected from infection and infestation.	92	98	96
7.4	Residents and staff are protected from the hazards of fire and natural disasters.	94	94	96
7.5	The security of buildings, contents and people within the nursing home is safeguarded.	98	98	98
7.6	Physical and other forms of restraint are used correctly and appropriately.	92	92	90

^a Overall agreement means the team and the independent rater gave exactly the same rating. These were for ratings made using the new format of met, action required and urgent action required. The per cent agreement was identical under the old format (met, met in part, not met) with the exceptions of: standard 5.6, 94 96 96; and standard 7.2, 84 90 88.

A high level of overall agreement was recorded for all standards. Not surprisingly, this level of agreement increased slightly after conferring. Receipt of negotiation feedback from the nursing home only made a slight difference – on some standards increasing agreement slightly, on others reducing it slightly. Table 2.3 shows the actual ratings given by the team and the independent rater after negotiation. It shows that the independent raters are slightly more reluctant to give met ratings. From Tables 2.2 and 2.3 we can see that the standard with the lowest level of agreement was 1.5, "Residents are enabled to maintain continence". The difference is explained by the independent rater taking a tougher position on what is required for met ratings. Independent raters expected stronger evidence of an individualized continence management program before a met rating could be given; dry sheets and regular toileting were not enough.

There were four other standards with higher levels of disagreement. These were 1.1, 1.9, 3.2 and 7.2.

Standard 1.1 requires that "residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed". The main sources of disagreement here were one of the raters observing evidence of transcription of medication and the other either not seeing it or choosing to overlook it.

Standard 1.9 concerns identification of sensory losses so residents can communicate effectively. Differences in ratings on this standard were mainly explained by some raters giving more weight than others to aspects of the environment, notably lighting levels that adversely affect those suffering from sensory loss.

Standard 3.2 refers to the residents or their representatives being enabled to comment or complain about conditions in the nursing home. Differences in this rating were mainly due to some raters emphasizing the positive aspects of what had been done (that is, complaints resolved) whereas others preferred to rate according to what had *not* been done. There were also disagreements when there was just one case of a resident saying they were "afraid to complain".

Standard 7.2 requires that "Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors". Generally, differences on the rating of this standard were due to one of the raters picking up a problem such as missing 'wet floors' signs, missing hand-rails, missing call-bells, and also differing perceptions of the seriousness of such faults.

One standard frequently criticised for inconsistency by people from the industry during our fieldwork was the homelike environment standard (4.1). This standard, however, is rated quite consistently – with 88 per cent initial agreement rising to 94 per cent after conferring.

Table 2.3: Ratings given by team and independent rater – after negotiation (N=50)

	Met %	Action Required %	Urgent Action Required %
1.1 Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed			
Rating by team	36	26	38
Rating by independent rater	30	30	40
Per cent agreement ^a	83	92	100
1.2 Residents are enabled and encouraged to make informed choices about their individual care plans			
Rating by team	62	30	8
Rating by independent rater	62	32	6
Per cent agreement	94	87	75
1.3 All residents are as free from pain as possible			
Rating by team	92	2	6
Rating by independent rater	86	6	8
Per cent agreement	93	100	100
1.4 All residents are adequately nourished and adequately hydrated			
Rating by team	42	44	14
Rating by independent rater	40	44	16
Per cent agreement	90	95	100
1.5 Residents are enabled to maintain continence			
Rating by team	26	60	14
Rating by independent rater	16	70	14
Per cent agreement	54	97	100
1.6 Residents are enabled to maintain, and if possible improve, their mobility and dexterity			
Rating by team	58	38	4
Rating by individual rater	58	38	4
Per cent agreement	97	95	100
1.7 Residents have clean healthy skin consistent with their age and general health			
Rating by team	90	10	0
Rating by individual rater	88	12	0
Per cent agreement	98	100	–
1.8 Residents are enabled to maintain oral and dental health			
Rating by team	68	32	0
Rating by individual rater	64	36	0
Per cent agreement	94	100	–
1.9 Sensory losses are identified and corrected so that residents are able to communicate effectively.			
Rating by team	72	28	0
Rating by individual rater	60	40	0
Per cent agreement	83	100	–
2.1 Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts.			
Rating by team	74	26	0
Rating by independent rater	74	26	0
Per cent agreement	97	92	–

Continued

		Met	Action Required	Urgent Action Required
2.2	Residents are enabled and encouraged to maintain control of their financial affairs.			
	Rating by team	78	10	12
	Rating by independent rater	76	14	10
	Per cent agreement	97	100	83
2.3	Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.			
	Rating by team	82	18	0
	Rating by independent rater	78	22	0
	Per cent agreement	93	89	—
2.4	Provision is made for residents with different religious, personal and cultural customs.			
	Rating by team	80	20	0
	Rating by independent rater	76	24	0
	Per cent agreement	95	100	—
2.5	Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.			
	Rating by team	90	8	2
	Rating by independent rater	88	10	2
	Per cent agreement	98	100	100
3.1	The nursing home has policies which have been developed in consultation with residents and which: regarding their daily activities; — provide an appropriate balance between residents rights and effective management of the nursing home; — and are interpreted flexibly taking into account individual resident needs.			
	Rating by team	70	26	4
	Rating by independent rater	64	28	8
	Per cent agreement	91	92	100
3.2	Residents and their representatives are enabled to comment or complain about conditions in the nursing home.			
	Rating by team	62	32	6
	Rating by independent rater	56	40	4
	Per cent agreement	87	100	67
4.1	Management of the nursing home is attempting to create and maintain a homelike environment.			
	Rating by team	40	54	6
	Rating by independent rater	38	56	6
	Per cent agreement	90	96	100
4.2	The nursing home has policies which enable residents to feel secure in their accommodation.			
	Rating by team	64	30	6
	Rating by independent rater	62	34	4
	Per cent agreement	94	93	67
5.1	The dignity of residents is respected by nursing home staff.			
	Rating by team	46	32	22
	Rating by independent rater	46	30	24
	Per cent agreement	100	94	100
5.2	Private property is not taken, lent or given to other people without the owners permission.			
	Rating by team	48	34	18
	Rating by independent rater	46	36	18
	Per cent agreement	96	100	100

Continued

	Met	Action Required	Urgent Action Required
5.3 Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private.			
Rating by team	36	34	30
Rating by independent rater	30	38	32
Per cent agreement	83	100	100
5.4 The nursing home is free from undue noise.			
Rating by team	70	30	0
Rating by independent rater	62	38	0
Per cent agreement	89	100	-
5.5 Information about residents is treated confidentially.			
Rating by team	70	20	10
Rating by independent rater	66	22	12
Per cent agreement	94	100	100
5.6 Nursing home practices support the residents right to die with dignity.			
Rating by team	82	18	0
Rating by independent rater	84	16	0
Per cent agreement	100	89	-
6.1 Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.			
Rating by team	42	56	2
Rating by independent rater	36	62	2
Per cent agreement	86	100	100
7.1 The residents right to participate in activities which many involve a degree of risk is respected.			
Rating by team	80	20	0
Rating by independent rater	78	22	0
Per cent agreement	95	90	-
7.2 Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors.			
Rating by team	16	46	38
Rating by independent rater	14	44	42
Per cent agreement	75	87	100
7.3 Residents, visitors and staff are protected from infection and infestation.			
Rating by team	36	46	18
Rating by independent rater	34	44	22
Per cent agreement	94	96	100
7.4 Residents and staff are protected from the hazards of fire and natural disasters.			
Rating by team	36	26	38
Rating by independent rater	34	24	42
Per cent agreement	94	92	100
7.5 The security of buildings, contents and people within the nursing home is safeguarded.			
Rating by team	68	28	4
Rating by independent rater	70	26	4
Per cent agreement	100	93	100
7.6 Physical and other forms of restraint are used correctly and appropriately.			
Rating by team	50	38	12
Rating by independent rater	46	40	14
Per cent agreement	88	89	100

a This is the per cent agreement by the independent rater with the teams rating.

Another common criticism is that the new social and residents' rights standards are "mushy" and more subjective than traditional medico-nursing or more tightly specified structure/process standards. Therefore, it is argued that reliability on them is unattainable. The data do not support this conclusion. Allegedly "soft" standards such as 5.1, "The dignity of residents is respected by nursing home staff"², are rated with impressive reliability.

The immediate suspicion one would harbour in seeing the exceptionally high agreement in Table 2.2 is that this is to be explained in the following way. In an overwhelming majority of cases both teams and independent raters give nursing homes met ratings. As a statistical artifact, it follows therefore that met-met agreement will be very common. Say we have 90 per cent getting mets on a particular standard. In the 90 per cent of cases where one of the raters gives a met rating, there will therefore be a 90 per cent chance that the other rater will agree on a met on purely statistical grounds.

Table 2.3 shows, however, that agreement is not merely statistical in this way. Surprisingly, while the percentage agreement of the independent rater when the team gives a met rating is 93 per cent (n=933) when it is an "action required" rating, it increases to 95 per cent (n=461), and when it is "urgent action required", agreement increases further to 97 per cent (n=156). The reasons why agreement on mets is slightly lower is that independent raters are somewhat tougher. Hence independent raters are slightly less likely to agree with teams when teams give met ratings.

While it is true that across all standards action required and urgent action required are even more reliably rated than mets, at the level of the reliability of individual standards, caution is warranted. There are five standards for which more than 40 of the nursing homes were given met ratings by the team – 1.3, 1.7, 2.3, 2.5 and 5.6. For these individual standards, while we can be impressed by the reliability of met ratings, we cannot be confident about the reliability of the other ratings, and we would be wise to harbour the doubt that the only reason there is so much agreement on mets is that mets are almost always given on these standards.

² Note this standard also is strongly validated (Table 7, Chapter 4; Figures 1-3, Chapter 4) and has a strong item-total correlation (Table 1, Chapter 2).

Overall, the inter-rater reliability coefficient for the initial total compliance score is .93; for the compliance score after conferring it rises to .96, and after negotiation it remains at .96. These are outstandingly high reliability coefficients.

Variation in Reliability Within the Sample

There was no significant difference in reliability between New South Wales and Victorian teams. The correlation between team and independent rater initial ratings for New South Wales was .93, while for Victoria it was .96. For ratings after negotiation, the coefficients for the two states were respectively .95 and .98. There were also no notable differences in reliability for small versus large homes and non-profit versus for-profit homes.

While intra-state reliability is high in both New South Wales and Victoria, we have not established that there is consistency between states in the way standards are rated. It would be a very expensive research project to fly independent raters from one state to another as they were required to join official teams. What we have established is that there is nothing about the nature of these standards that makes inter-state reliability impossible. Our fieldwork suggests, however, that there are some differences of note between states in how heavily they weigh different aspects of the standards and how tough they are in general. The data also show considerable variation between states in the proportion of ratings that are mets on particular standards (see Braithwaite *et al.*, 1990). Of course we cannot be sure that this does not reflect real differences between states in the level of compliance with those standards.

It must be said, however, that most of the processes which would seem to have delivered the high intra-state reliability reported here are intra-state processes – in particular, regular meetings of standards monitors to discuss tricky ratings, dialogue on how to rate specific homes with team members rotated from other teams within the state, and training courses. Perhaps what is needed now is more replication of these successful intra-state processes at an inter-state level. In the early years of a new program there is actually merit in allowing different states to experiment with their own ways of solving problems. The time is now here to pull together the best lessons from that diverse experience in an effort to get inter-state convergence on the best solution. Options that are being considered by the Department of Community Services and Health to improve inter-state reliability are: (a) more workshops on consistency for standards monitors from different states; (b) experimenting with standards monitors from one state conducting monitoring visits with colleagues from other states; (c) central office staff attendance at standard review days with

states; and (d) asking the "tougher" state on a particular standard to liaise with the "softer" state (or vice versa) to prepare a paper on how to achieve consistency on the standard.

Changing the Rating Categories

One unfortunate aspect of the study was that it commenced just before the Department made a slight change to the three rating categories to be used. Until the end of 1989 nursing homes were rated as having "met", "met in part" or "not met" the standards. These categories were defined as follows:

Met – The outcome indicated by the standard is met for residents in the nursing home. Furthermore, there is an understanding of the intent of the standard.

N.B. It should be noted that there is no inconsistency if the nursing home is still able to improve performance in an area where the standard is considered to be met.

Met in part – For a standard to be judged "met in part", efforts should have been made to meet the standard accompanied with knowledge of the intent of the standard. The standard should be considered met in part where:

(i) the standard is fully met for the majority of residents, and the remaining residents are not considered to be suffering abuse, neglect, denial of rights and/or other significant detriment as a result of the standard not being met in full; or

(ii) where for all of the residents the considerations are substantially satisfied.

Not met – For a standard to be judged "not met", either lack of knowledge and substantial lack of practices or procedures to ensure the attainment of the standard should be in evidence. Where one or more residents are suffering abuse, neglect, denial of rights and/or other significant detriment as a result of a standard not being complied with, the standard is to be considered "not met".

Following its *Standards Monitoring Review* in 1989, the department revised the definition of the standards to achieve two objectives: simplification and making it more palatable to give, and get, not met ratings. To the latter end, not met and met in part were replaced with "urgent action required" and "action required". These new ratings had a more constructive, future-oriented tone to them rather than a backward-looking, exam-result, quality. It can be seen below that the definitions for "action required" and "urgent action required" were virtually identical to the old "met in part" and "not met" definitions, but simplified:

Met – The team considers that residents are experiencing the quality of life and care described in the standard. This does not necessarily mean there is not room for improvement or that the home could not operate more efficiently.

Action required – EITHER the standard is fully met for the majority of residents and the other residents are not experiencing neglect, abuse, denial of rights or any other significant detriment OR substantially met for all residents and the home is taking action to address those minor concerns identified.

Urgent action required – For one or more residents there is an identified abuse, neglect, denial of rights and/or other significant detriment.

To check what difference this rather minor change made, both teams and independent raters were asked to fill out separate ratings under the two sets of definitions. Only 283 of the 9,300 ratings changed. 53 per cent of these changes were in the direction of tougher ratings with the new standards, 47 per cent in the direction of softer ratings. The only changes in level of agreement of the independent rater with the team for Table 2.2 were with standards 5.6 and 7.2 (see note ^a under Table 2.2).

Overall, all inter-rater reliability coefficients for total compliance were unchanged (to two decimal places) by moving from the new to the old rating categories. There may be social desirability bias in these results as raters knew that senior management of the program had said that they did not expect the change of rating labels to make a difference to the ratings actually given.

Another minor issue was the occasional use (peculiar to Victoria under the old rating categories) of the category "met with room for improvement". These were coded "met" for purposes of the foregoing analyses. Treating "met with room for improvement" as a separate category did not change any of the inter-rater reliability coefficients for total compliance nor materially affect any other results.

Reasons for Disagreements

Table 2.4 shows the sources of disagreement when different ratings were given by the team and the independent rater. The first five categories all relate to different information being collected by the team versus the independent rater; they account for the majority of the sources of disagreement. In three cases the independent rater was given confidential information by the director of nursing that was not provided to the team. Since part of the reason why the information was voluntarily given would seem to have been that information given to the independent rater would not adversely affect official ratings, the independent rater did not share this information with the team. The independent rater simply advised the team that she disagreed because of confidential information in her possession. It remained a submerged source of disagreement at all three stages where agreement was calculated.

Table 2.4: Reasons for differences between team ratings and independent ratings at the initial rating stage (n=146)

	N	Per cent
<i>Differences in the collection of information</i>		
1. One rater saw documentation the other missed	5	3
2. One rater got extra information from residents	11	8
3. One rater got extra information from staff, director of nursing or proprietor	11	8
4. One rater got extra information from visitors	4	3
5. One rater saw an incident or observed a feature of the home that the other missed	39	27
<i>Differences in the interpretation of standards</i>		
6. One rater required more instances of a problem to justify a lower rating (to constitute a pattern)	5	3
7. One rater required more to be done for higher rating	38	26
8. Raters gave different weightings to the different aspects of a standard	27	19
9. Disagreement as to whether this aspect of alleged poor care was covered by the standard	6	4

The source of disagreement when different information was collected was less often different documentation that was seen or different information that was gained in interviews with residents, staff or relatives; it was more often a particular incident or other direct observation that one of the raters had seen but the other had not. Standards under the privacy and dignity objective were the most common source of disagreement here. One rater would see a resident being changed, toileted or showered within sight of other residents and the other rater would not see this incident. A single incident of this kind is viewed as serious and normally results in the nursing home losing its met rating. A staff indiscretion such as the following is another example, as expressed in the words of the independent rater:

Standard 5.5: I rated this standard "met" but agreed with the team that it should be "urgent action required" when a team member told me that:

- a staff member was heard to loudly discuss a resident's mental condition
- the same staff member was heard discussing the medial diagnosis of a resident.

Both incidents happened in the corridor of the home where residents and visitors could overhear.

Only five of these disagreements were due to the independent rater seeing an incident that occurred at a time when the team was not in the nursing home; 34 were a result of the luck of being in the right place at the time when an incident occurred.

Only five disagreements were accounted for by raters disagreeing on how many instances of a particular problem are needed to constitute a pattern of poor quality care

sufficient to warrant an action required or an urgent action required rating. We had expected more disagreement than this on whether one incident was enough to put a standard out, and if not one, how many. The following extract from the report of an independent rater is an example of the type of disagreement that did occur:

Standard 7.6: The team observed an isolated incident of inadequate documentation of restraint. [The independent rater] did not change the rating as the evidence suggested that policy was adequate and every effort was made to ensure compliance with the standard. The incident observed by the team was unusual circumstances and may have been related to the pressures of the team's visit.

A much more common source of disagreement was one rater requiring more to be done to meet the standard. In 15 of these cases, the heart of this disagreement was that one rater required an input to be in place while the other was satisfied so long as the outcome for the resident was regarded as satisfactory. Here are some examples of these types of disagreements, expressed in the words of the independent rater:

Standard 5.3: I rated the standard as "action required" because the shower unit and the bathroom did not have locks or indicators on the doors. The team said they witnessed staff knocking on the doors and requesting permission to enter. During the visit neither the bathroom nor shower were not occupied; however, in any event I do not believe the team observation warranted a met classification, as the potential for resident embarrassment is high.

Standard 1.3: The rater was concerned that there was no assessment or evaluation of the severity of pain, its causes, and the effectiveness of pain relief measures. The team rated this as met since there was evidence that if a person complained of pain, medication was given.

Standard 1.2: The team rated this standard "action required" because they did not believe that care plans were regularly reviewed. I disagreed because nurses notes indicated review and although care plans were not dated they were obviously used by the "hands on" staff. All staff were very aware of residents' needs. This Department cannot insist that any one specific document is used by facilities. We can only insist that each resident has a "plan of care". The format used is up to the home.

Standard 7.2: The team rated this standard as "action required". The hot water/mixing valve system did not comply with the [state] Department of Health regulations. [The independent rater] rated as met as the existing system did not in any way compromise the safety of the residents – water was coming through the system at an approved 43 degrees Centigrade. There are regular temperature checks and maintenance of the unit. All other aspects were satisfactory. Later a Commonwealth team member confided to the rater that the rating was made in the interest of harmonious Commonwealth/State relations. Without a [state] representative on the team the rating would have almost certainly have been met.

Six of the disagreements under the "requiring more to be done" heading were for standard 1.5, "Residents are enabled to maintain continence". The following account of the independent rater typifies this source of disagreement:

Standard 1.5: The team rated this standard "met" but I disagreed because there were no initial medical/nursing assessments of the residents who are admitted to the home incontinent. There is no educational program for staff or residents. Any assessment is "hit and miss" and nearly everyone is toileted 2nd hourly. The team agreed and changed their rating to "action required".

There were 27 cases of disagreement over how much weighting to give one aspect of a standard over another. For example, on standard 1.4, if the food served is sufficient in quantity and nutritional quality, how much weighting should one give to the fact that the coffee is served cold? One of the interesting things that happened here points up the advantage of team dialogue. The following are cases where the team and the rater had made the same observations, but the independent rater admitted to neglecting the weighing of this aspect of her observations when it came to making her rating.

Standard 3.1: The team was concerned that residents were given limited food choices and that the menu was repetitious. Consultation with the residents appeared not to take place. Information gathered by the team was the same as mine. I did not take the information above into consideration when I rated the standard. I agree with the team [now] that the rating should be "urgent action required" and not "action required".

Standard 5.1: The team pointed out that "food wastage" charts pinned to the notice board in the corridor was undignified: the charts had residents' names and were on full view to everyone. I had noted the charts, but did not consider them when rating the standard. I agreed that the standard should be rated "action required".

A much less common problem was disagreement over whether a particular type of poor care or poor quality of life outcome was covered by the particular standard. There were only 6 cases of this type of disagreement. The following is an example concerning standard 1.5, "Residents are enabled to maintain continence."

Standard 1.5: Same evidence collected. [The independent rater] rated standard as met. There was evidence of assessment, monitoring and review of continence problems. Toilets were accessible and appropriate aids supplied. Team rated "met in part" as they were concerned that one resident had to pay for her continence pads. The team commented that they hoped to move to "met" at the negotiation. [The independent rater] did not change rating as the issue of who pays for continence aides is not addressed in this standard.

In New South Wales, a particularly important source of disagreement under this heading was an instruction from the Sydney office of the Department not to downgrade a nursing home more than once for the same thing. Unreliability arises when one rater chooses to put out standard A while the other chooses to put out standard B for the same incident. This source of unreliability is one reason why official Commonwealth policy is to put both standard A and B out of compliance if a single incident causes both to be outcomes that are not met.

Summary

1. The internal consistency of the 31 standards is sufficient to justify adding scores from all standards (without any deletions) to form a total compliance score. The alpha reliability coefficient for this measure was .90.
2. The standard with the lowest correlation with the total compliance score is 2.5, "Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens."
3. The standards with the highest correlation with the total compliance score were: 1.2, resident participation in individual care plans; 1.5, maintaining continence; 3.1, policies developed in consultation with residents to enable freedom of choice; and 4.1, homelike environment. What the standards which are the best predictors of compliance have most in common is a focus on individualization (as opposed to institutionalization) and resident participation.
4. An inter-rater reliability study was conducted by placing an independent rater in the nursing home at the same time as the standards monitoring team for 50 nursing home visits. Across 3,100 instances where the team and the independent rater made independent ratings, they gave exactly the same rating on a three-point scale 94 per cent of the time.
5. Inter-rater reliability coefficients increased from .93 for initial ratings to .96 after conferring and .96 after negotiation.
6. Independent raters agreed with team ratings of not met/urgent action required ratings 97 per cent of the time.
7. Not only the "met/not met" divide, but also the "met in part/not met" divide (under the old categories) and the "action required/urgent action required" divide are rated reliably. Change from the old to the new rating categories makes minimal difference to results.
8. Inter-rater reliabilities were much the same in Victoria as in New South Wales, for small homes as for large homes and in non-profit homes as in for-profit homes.
9. The standard with the highest level of disagreement between raters was 1.5, "Residents are enabled to maintain continence". Others where unreliability frequently arose were: 1.1, residents able to receive appropriate medical care by a doctor of their choice; 1.9, sensory losses identified and corrected; 3.2, residents and representatives able to complain; 7.2, safe environment. Even these standards where the greatest difficulties were encountered are rated quite reliably.
10. The homelike environment standard (4.1), which has been subjected to considerable criticism in the industry for its alleged unreliability, was in fact rated quite reliably.
11. The "soft" social and human rights standards, such as 5.1, "The dignity of residents is respected by nursing home staff", are rated quite reliably.

12. Most disagreements between raters are caused by one rater collecting information that the other missed. The most common kind of difference in information collected is the direct observation of incidents – such as residents being changed or toileted in view of other residents.
13. The second most common source of disagreement was one rater requiring more for a standard to be met than the other rater. An especially common form of this disagreement was one rater requiring an input to be present while the other was satisfied if the outcome was perceived to be met.
14. The third most common source of disagreement was raters giving different weighting to the importance of different aspects of a standard.
15. Disagreements over whether an aspect of poor care should be included under a particular standard and disagreement over how many instances of a particular problem were needed to justify a lower rating were surprisingly uncommon.
16. A troubling cause of unreliability in New South Wales arose where a single problem caused more than one outcome not to be met. New South Wales standards monitoring staff felt it was unfair to allow one problem to cause more than one standard to be put out of compliance. However, inconsistency arose from arbitrary choice of which of the adversely affected standards would and would not be rated met.
17. While these data show strong evidence of inter-rater reliability within states, it does not demonstrate inter-state reliability. The data show that there is no reason why inter-state reliability cannot be achieved with these standards; but they do not show that it is actually achieved. There are in fact reasons to suspect that while fairly good intra-state reliability is currently being achieved, inter-state reliability is not.

3. FACE AND CONTENT VALIDITY

Face validity is especially important in the domain of business regulation because regulations that do not make sense to those being regulated will be denied legitimacy, will be resisted (politically, administratively and in the courts) and will not voluntarily be complied with. A standard can have face validity without having content validity – without measuring accurately the things that it is necessary to measure in order to assess this aspect of quality of life. The first issue we must address is the comprehensiveness of the standards. Then we will examine their clarity, the difficulties they pose for raters, their desirability and their practicality.

Comprehensiveness

At one level, the 31 standards score very well on the comprehensiveness criterion. The authors have accompanied standards monitoring teams on 37 visits, observing team members debate among themselves how to rate a particular home. We have not observed it to be a problem that team members have a concern about an adverse quality of life outcome for residents but cannot find a standard to write this up under. When they raise a problem with their colleagues on the team, they can almost always quickly identify a standard or standards that cover that problem. When our American colleague Professor David Ermann went out with Australian teams in three states viewing the process through his American eyes, there were not issues that he could identify that could not be fitted into the Australian standards. When the Department conducted its *Standards Monitoring Review* in 1989, extensive consultations with the industry, unions and consumers took place. While there was a minority view that the Department should give guidance to the industry by supplying comprehensive lists of inputs relevant to each of the outcomes, suggestions about outcomes that were important and not covered by the standards were not raised.

The cynic might say that the standards are so broad and vague that it is hardly surprising that they cover everything in principle. The real question, the cynic asks, is whether in practice teams collect all the data they need to rate validly all these outcomes. Our own view is that there are quite serious limitations on the extent to which they do. In the following paragraphs we will seek to put some flesh on the bare bones of this assertion.

A common assertion about nursing homes from industry participants in all the countries we have visited for this research is that a competent person can identify a nursing home with poor standards within an hour of looking around. This assertion has not sat well

with American researchers who have consistently failed to find high reliability in the ratings of American inspection teams (see Chapter 5). It may be, however, that both the genius and the limitation of the Australian process is that it engages with the nursing home only at that rather broad level of the quality of care it provides, and it is at this broad level that reliability can be achieved, as in this oft-repeated conventional wisdom of the industry.

Once the team follows protocols that cause it to dig deeper, reliability may become problematic. Yet in failing to dig deep, the process may in some senses be reliable but not valid. Consider standard 2.2, "Residents are enabled and encouraged to maintain control of their financial affairs." This is assessed by a team member asking the person responsible for managing resident accounts (and for liaison with guardians or relatives who manage accounts) to explain the nursing home's system for ensuring that the standard is met. Documentary evidence of these systems will then be sighted. In addition, the team will ask residents and visitors if they are experiencing any problems in maintaining resident control of their financial affairs. What we have found is that different raters who only dig this deep in their investigations of this matter will come up with the same ratings. However, if one of the teams were to dig deeper and conduct a full-scale financial audit of all of the residents accounts, it might find instances of residents being deceived and defrauded by the nursing home. By one team digging deeper, inter-team reliability would have been shattered, but this team would be making a more valid rating. What we have with Australian standards monitoring is a process that reliably reveals those sorts of problems that can be revealed by shallow digging.

There are many ways that the standards monitoring process could be made more valid by more comprehensive data collection. Unfortunately, however, as in the example of systematic financial auditing, they would be expensive to implement. Another example is the American practice of conducting a systematic survey of errors in the administration of medications. United States inspections quite often uncover frightening "med-pass error rates" of 10 per cent and more. Australian standards monitoring visits reliably fail to uncover such problems. A further difference is that American inspectors observe treatments being given by nurses to a sample of residents. Mostly these are observations of the treatment of decubitus ulcers (bedsores). This deeper digging in the American process uncovers many problems which remain reliably submerged in the Australian process – Class III ulcers that are documented and treated as Class II, poor infection control practice in the treatment of bedsores and inadequate respect for the privacy or dignity of the resident during this most undignified aspect of nursing home life.

In conclusion, while the standards are comprehensive in the sense of covering well all the outcomes that are of substantial concern, for each standard the process for uncovering these adverse outcomes is less than comprehensive. The Australian process reliably uncovers most of the problems that can be exposed by shallow digging and reliably fails to uncover the problems that can only be exposed by deeper digging.

Clarity of the Standards

The data source for this section comes from interviews with the directors of nursing of 410 homes conducted over a twenty-three month period from May 1988 to March 1990. These homes were located around the major capital cities of the states of Queensland, New South Wales, Victoria and South Australia where two-thirds of all Australian nursing homes are located. Two hundred and forty-two of these homes represent a proportionate stratified random sample. The remaining 168 homes are a supplementary sample of all other homes visited in three of the sampling regions during the period of the study.³

The homes were initially contacted by letter asking them to participate in the study once the negotiations with the standards monitoring team had been completed. If they agreed to participate a time and date was arranged for an interview. The interviews were conducted by three highly experienced interviewers. Qualitative fieldwork which had commenced in 1987 informed the construction of the schedule. It was initially piloted on seven homes resulting in word changes on the questionnaire and feedback to the interviewers on their performance.

Directors of nursing were asked "Are there any of these standards which you think are unclear? That is, you are unsure what the standard means?" Table 3.1, first column, indicates a surprisingly high level of clarity and certainty in the minds of directors of nursing as to what the standards mean. For 29 of the 31 standards, over 95 per cent of directors of nursing thought the standards were clear. This is an exceptionally good result for both the standards, and the training that has been made available to directors of nursing on the standards.

³ For a more detailed account of the sample see Braithwaite *et al.* (1990).

Table 3.1: Evaluating the face and content validity of the standards (Percentages)

		Unclear/Difficult to rate ^a	Doubt Desirability	Doubt Practicality
1.1	Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed			
	Director of nursing	2.7	1.7	7.1
	Standards monitor	10.9	0.6	14.5
1.2	Residents are enabled and encouraged to make informed choices about their individual care plans			
	Director of nursing	3.2	3.7	20.0
	Standards monitor	32.1	2.4	35.2
1.3	All residents are as free from pain as possible			
	Director of nursing	1.2	0.0	1.2
	Standards monitor	10.9	0.6	3.0
1.4	All residents are adequately nourished and adequately hydrated			
	Director of nursing	1.0	2.0	1.7
	Standards monitor	3.6	0.6	0.0
1.5	Residents are enabled to maintain continence			
	Director of nursing	3.9	0.2	11.7
	Standards monitor	24.2	1.2	10.3
1.6	Residents are enabled to maintain, and if possible improve, their mobility and dexterity			
	Director of nursing	0.7	0.0	4.4
	Standards monitor	10.3	0.6	4.2
1.7	Residents have clean healthy skin consistent with their age and general health			
	Director of nursing	0.5	0.0	1.2
	Standards monitor	1.2	0.6	0.6
1.8	Residents are enabled to maintain oral and dental health			
	Director of nursing	0.5	0.2	3.4
	Standards monitor	2.4	1.2	0.6
1.9	Sensory losses are identified and corrected so that residents are able to communicate effectively.			
	Director of nursing	1.0	0.2	5.9
	Standards monitor	10.0	0.6	3.6
2.1	Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts.			
	Director of nursing	1.2	0.0	2.9
	Standards monitor	6.7	0.6	4.2
2.2	Residents are enabled and encouraged to maintain control of their financial affairs.			
	Director of nursing	1.2	4.6	24.1
	Standards monitor	33.9	1.8	28.5
2.3	Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.			
	Director of nursing	3.2	0.2	6.6
	Standards monitor	22.4	1.2	7.9

Continued

		Unclear/Difficult to rate ^a	Doubt Desirability	Doubt Practicality
2.4	Provision is made for residents with different religious, personal and cultural customs.			
	Director of nursing	1.0	0.0	2.7
	Standards monitor	19.4	0.6	7.9
2.5	Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.			
	Director of nursing	2.0	0.7	8.0
	Standards monitor	8.5	1.2	8.5
3.1	The nursing home has policies which have been developed in consultation with residents and which: regarding their daily activities; – provide an appropriate balance between residents rights and effective management of the nursing home; – and are interpreted flexibly taking into account individual resident needs.			
	Director of nursing	4.1	2.4	22.2
	Standards monitor	33.3	3.6	24.8
3.2	Residents and their representatives are enabled to comment or complain about conditions in the nursing home.			
	Director of nursing	0.7	0.2	2.9
	Standards monitor	15.2	1.2	6.7
4.1	Management of the nursing home is attempting to create and maintain a homelike environment.			
	Director of nursing	3.7	0.5	12.2
	Standards monitor	23.0	0.0	9.1
4.2	The nursing home has policies which enable residents to feel secure in their accommodation.			
	Director of nursing	2.7	1.0	3.7
	Standards monitor	16.4	0.0	3.0
5.1	The dignity of residents is respected by nursing home staff.			
	Director of nursing	1.0	0.0	1.2
	Standards monitor	4.8	0.6	1.8
5.2	Private property is not taken, lent or given to other people without the owners permission.			
	Director of nursing	1.2	0.7	5.1
	Standards monitor	5.5	0.0	3.0
5.3	Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private.			
	Director of nursing	1.2	0.7	5.4
	Standards monitor	3.0	0.0	3.6
5.4	The nursing home is free from undue noise.			
	Director of nursing	3.2	0.5	4.9
	Standards monitor	6.1	0.0	7.3
5.5	Information about residents is treated confidentially.			
	Director of nursing	1.0	0.5	1.5
	Standards monitor	6.1	0.0	0.6
5.6	Nursing home practices support the residents right to die with dignity.			
	Director of nursing	1.2	0.0	1.7
	Standards monitor	21.2	2.4	8.5
				Continued

		Unclear/Difficult to rate ^a	Doubt Desirability	Doubt Practicality
6.1	Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.			
	Director of nursing	1.2	0.2	5.1
	Standards monitor	10.9	0.6	5.5
7.1	The residents right to participate in activities which many involve a degree of risk is respected.			
	Director of nursing	8.3	5.4	13.7
	Standards monitor	35.8	3.0	18.8
7.2	Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors.			
	Director of nursing	1.7	0.0	3.4
	Standards monitor	10.9	1.2	9.1
7.3	Residents, visitors and staff are protected from infection and infestation.			
	Director of nursing	2.0	0.2	1.2
	Standards monitor	6.7	1.2	3.0
7.4	Residents and staff are protected from the hazards of fire and natural disasters.			
	Director of nursing	2.4	0.0	0.7
	Standards monitor	3.6	0.6	2.4
7.5	The security of buildings, contents and people within the nursing home is safeguarded.			
	Director of nursing	1.5	0.0	0.7
	Standards monitor	5.5	0.0	2.4
7.6	Physical and other forms of restraint are used correctly and appropriately.			
	Director of nursing	3.4	1.5	3.7
	Standards monitor	15.2	0.6	2.4

^a Directors of nursing were asked which standards were unclear. Teams were asked which standards they had difficulty rating consistently.

The only standards with significant clarity problems were 3.1, concerning policies developed in consultation with residents (4 per cent unclear), 1.5, "Residents are enabled to maintain continence" (4 per cent unclear); 4.1, concerning a homelike environment (4 per cent unclear), and 7.1, "The resident's right to participate in activities which may involve a degree of risk is respected." (8 per cent unclear). Standard 3.1 was perceived as a problem by directors of nursing who wanted clear guidance on where to draw the balance between policies that guarantee resident rights, and policies which assure effective management of the nursing home. In this regard, the use of the term "appropriate balance" was said to be vague. Only one of these standards with minor problems of clarity was among the standards with minor problems on inter-rater reliability – 1.5.

Standard 1.5 was interpreted as vague by directors of nursing who wished to be told what "maintaining continence" meant. A number of directors of nursing pointed out that maintaining continence is impossible for some residents. In this regard, it seems to the

consultants that the "look fors" in the *Nursing Home Standards Monitoring Guidelines* (Department of Community Services and Health, 1987: 6) provide helpful clarification. In the case of standard 1.5 the guidelines provide four "look fors".

- 1) Practices which identify residents experiencing problems maintaining continence and/or residents with urinary tract and related infections.
- 2) Evidence that individual continence management programs have been developed for residents who require them, and that programs are regularly reviewed.
- 3) Toilets which are accessible to residents and aids and equipment to assist this access.
- 4) Appropriate aids to assist incontinent residents.'

Perhaps the "look fors" with standard 1.5 should be given particular emphasis at pre-visit seminars and other training courses.

The notion of a homelike environment in standard 4.1 was viewed as subjective and variable in light of personal preference, ethnicity, and so forth. The critics said that what is homelike to a resident, the team may see as sterile. One director of nursing made the interesting comment that the process focuses too much on physical aspects of the environment:

"The atmosphere [is important] – the feeling of belonging and caring. Perhaps the team cannot measure this as they can inputs like plants and pictures."

Standard 7.1, on the resident's right to take risks, was viewed by some as unclear in its relationship to the legal duty of care which nurses owe to their patients. Many directors of nursing had difficulty in assessing what was an acceptable risk. Moreover, they pointed out it is not just a matter of striking a balance between a resident's right and a nurse's duty; there were also relatives to consider: "Relatives don't like to see their elderly parents with broken limbs". This standard thus calls into question some broader societal perceptions concerning "protection" of the elderly; some directors of nursing reported cases where children of residents were insisting on physical restraint following a fall; restraint which was totally unnecessary in their opinion. The issue of the relationship between outcome standards and common community perceptions may thus require some further scrutiny.

Toward the end of the first wave of the study, an important report on this matter was released, *Commonwealth Nursing Home Outcome Standards: A Practical Guide to the Duty of Care* (McDonald and Bates, 1989). This report provides invaluable guidance on what the law requires in resolving the extremely difficult dilemmas that arise under standard

7.1. Hopefully, as this guidance percolates into industry training courses, directors of nursing will become less anxious about this standard. Among other things, McDonald and Bates (1989:30) advise:

It is not a breach of the common law obligation of reasonable care for a nursing home or its staff to respect the lifestyle choices of residents of sound mind, even if these choices are foolish or dangerous, provided that a reasonable effort is made to inform the resident prior to the activity about the risks involved in what they are choosing to do. If the resident persists and is injured, then the nursing home would not be liable provided they had made reasonable efforts to counsel the resident. All of this should be documented in order to protect the nursing home and staff.

Difficulty in Rating the Standards Consistently

The data for this section comes from 165 standards monitoring team members who completed a mail questionnaire. All team members, past and present, who participated in standards monitoring since its introduction in 1987 who could be contacted were asked to complete a mail questionnaire. Names of 222 team members were provided by the Department of Community Services and Health. These people were mailed a questionnaire on 1 May 1990. Refusals and returns to sender were not contacted further. Those who did not reply were sent two mail reminders and a telephone follow-up. There were 8 refusals, 29 returns to sender, 20 non-returns with 165 returned usable questionnaires.

Team members were asked: "Do you find that there are some standards which you have difficulty rating consistently? Please circle "1" for those standards which you have difficulty rating consistently." Table 3.1 shows that the same standard which received most unclear ratings from directors of nursing also received most difficulty in rating responses from team members. This was standard 7.1, "The resident's right to participate in activities which may involve a degree of risk is respected." Thirty-six per cent of team members said they had difficulty rating this standard consistently. Yet we know that the raters in the inter-rater reliability study in practice rated this standard very consistently. Our fieldwork observations suggest that when this standard is called into question quite intense dialogue and rigorous resident-centred investigation occurs. This intensity and rigour reflects the understanding of team members that this is a sensitive standard for many in the industry. But it would seem that the upshot of this intensive debate and investigation is ratings of impressive reliability.

Other standards which teams felt they had difficulty rating were: 1.2, residents enabled to make choices about care plans (32 per cent); 2.2, residents enabled to maintain control of their financial affairs (34 per cent); and 3.1, nursing home policies developed in consultation with residents (33 per cent).

Desirability and Practicality of the Standards

In addition to asking directors of nursing whether they saw the standards as unclear, they were also asked whether they saw any of the standards as impractical or undesirable. Table 3.1, column 2 shows the proportion of directors of nursing rating the standards as undesirable. For 28 of the 31 standards, over 97 per cent of respondents had no doubts about their desirability. Standard 1.1 "Residents enabled appropriate medical care by a medical practitioner of their choice" is a standard one might have expected to arouse controversy in light of the recent opposition of the Australian Catholic Bishops Conference to residents having a contractual right to their own doctor (on grounds that Catholic homes should be able to reject a doctor who practises euthanasia). However, only 7 out of 410 directors of nursing doubted the desirability of standard 1.1. There are 106 church homes in the sample and 2 of these had directors of nursing who had doubts about this particular standard, in these cases euthanasia was not mentioned as the reason.

The only three standards causing a ripple of doubt concerning their desirability were 1.2 "Residents are enabled and encouraged to make informed choices about their individual care plans" (4 per cent with doubts); 2.2 "Residents are enabled and encouraged to maintain control of their financial affairs" (5 per cent with doubts); and 7.1 "The resident's right to participate in activities which may involve a degree of risk is respected" (5 per cent with doubts). Note that these were all standards that the teams said they had difficulty rating. Again, the interpretation is suggested that the teams agonize most over the standards with face validity problems for a minority of the industry.

Four types of reasons were given against the desirability of standard 1.2. First, it was said that confused residents are not capable of making informed choices about their care plans. Second, it was said that residents did not want to be involved in care plans: "You don't keep a dog and bark yourself". Third, it would be too costly to have residents active in preparing all individual care plans: "Individual care plans are impractical. The nursing home does not have enough resources to cope with every choice. There need to be rules." Fourth, handing authority over care plans to residents was seen by some as an abdication of professional responsibility:

"We as professional nurses should make the decision."

"I think we should be the spokesperson for the residents."

Standard 2.2 was subjected to criticism because of the view that confused or forgetful residents could not, and should not, be encouraged to maintain control of their own financial affairs. Sometimes, it was pointed out that it was important for the nursing home to exert some control to protect residents from unscrupulous relatives who would take their money.

Concerns about the desirability of standard 7.1 were strongly related to the same reasons that made it unclear. Particularly salient was the fear that relatives would perceive care as inadequate if the nursing home allowed residents to take risks. Another primary concern was that it would be irresponsible to allow, in particular, demented residents to take risks. And finally, the standard was seen by a few as a threat to professional responsibility:

"The buck stops with me [the director of nursing]".

"The nursing home should have the sole right – in consultation with doctors – to determine whether patients undertake certain activities."

Generally, it is fair to say that while doubts about the desirability of the standards were unusual, when they did occur, they were motivated by a belief in the need to be protective of residents, as in the case of the privacy standard, 5.3. This standard was questioned by a director of nursing who doubted "the desirability of allowing elderly couples to have intercourse."

Table 3.1 shows that the standards monitoring team members were even less likely to harbour doubts about the desirability of the standards than directors of nursing. For 27 of the standards, over 98 per cent of team members regarded them as desirable. Four of the standards for which at least 3 standards monitors expressed doubts about their desirability – 1.2, 2.2, 3.1, 7.1 – were also standards which stirred some dissent from directors of nursing. No directors of nursing, but 3.6 per cent of team members had doubts about standard 5.6; nursing home practices support the residents right to die with dignity. Team member opposition here relates to the difficulty and intrusiveness of getting evidence of compliance with this standard.

When we asked directors of nursing and team members whether, even if desirable, some of the standards were not practical, the level of endorsement fell somewhat. Even so, Table 3.1 shows that for all standards at least 65 per cent of both directors of nursing and

team members had no doubts about their practicality and, for most standards, more than 90 per cent thought them practical. Problems again were with the same standards for both groups and standards that emerged as the greatest problems on the desirability ratings:

- 1.2, resident participation in care planning (20 per cent of directors and 35 per cent of monitors doubt practicality);
- 2.2, resident control of financial affairs (24 per cent of directors and 29 per cent monitors doubt practicality); and
- 7.1, resident's right to participate in activities with a degree of risk (14 per cent of directors and 19 per cent of monitors doubt practicality).

These were joined by 3.1, policies developed in consultation with residents (22 per cent of directors and 25 per cent of monitors doubt practicality), one of the standards with significant problems of clarity. Standard 3.1 was viewed as impractical by many directors of nursing who said that policies to allow residents a choice in matters like meal and shower times were difficult in an institutional setting:

"Does the nursing home become a short-order cook. The balance of residents' rights has gone too far in favour of residents."

Across all standards, common themes were that the standards would only become practical if the Commonwealth provided more money⁴, or proprietors provided more staff to the nursing home, and/or if the level of disability of residents were to fall to the level existing in hostels rather than nursing homes.

It is noteworthy that the results from our less structured interviews with proprietors were consistent with the results from the director of nursing interviews and the standards monitors questionnaires. Three quarters of proprietors saw no problems with any of the standards. We asked them, "Are there any standards you disagree with?" As with the directors of nursing on the issue of practicality, 2.2, resident control of financial affairs, was the standard most mentioned by proprietors as a problem. Next was 2.3, then 3.1, 4.1, and 1.2. The higher prominence of 2.3, freedom of movement within and from the nursing home, for the proprietor was the major difference between them and the other two groups.

⁴ At another point in the questionnaire, directors of nursing were asked to agree or disagree with the statement: "It is impossible for nursing homes like mine to meet the standards unless the level of Commonwealth funding is increased." Forty-three per cent agreed with this statement, 35 per cent disagreed, and 22 per cent neither agreed nor disagreed. For the statement: "It is quite possible for my home to make ends meet while complying with the standards", 49 per cent agreed, 30 per cent disagreed, and 21 per cent neither agreed nor disagreed.

Summary

1. The 31 standards would seem to be comprehensive in covering nursing home quality of life outcomes of concern.
2. The process for gathering information relevant to this comprehensive set of outcomes is far from comprehensive. The process may reliably detect those problems that can be exposed by shallow investigation, yet reliability fails to detect other types of problems that require deeper investigation.
3. The standards are overwhelmingly rated clear, desirable and practical by directors of nursing. More qualitative data from proprietors of the same sample of nursing homes also supports this conclusion.
4. Standards monitoring team members do not report difficulties on rating most standards consistently. There are even fewer team members who have doubts about the desirability of standards than directors of nursing. However, there are more team members than directors of nursing who have doubts about the practicality of some of the standards.
5. The only standards with significant clarity problems for directors of nursing were: 1.5, "Residents are enabled to maintain continence"; 3.1, policies developed in consultation with residents; 4.1, homelike environment; and 7.1, resident's right to participate in activities which involve risk.
6. The standards which a significant minority of team members said they had difficulty rating consistently were: 1.2, residents enabled to make choices about care plans; 2.2, residents enabled to maintain control of their financial affairs; 3.1, policies developed in consultation with residents; and 7.1, resident's right to participate in activities which involve risk.
7. Standards for which some small numbers of both directors of nursing and team members harboured doubts about desirability were: 1.2, residents enabled to make choices about care plans; 2.2, residents enabled to maintain control of their financial affairs; 3.1, policies developed in consultation with residents; and 7.1, resident's right to participate in activities which involve risk. A standard for which no directors of nursing had doubts about its desirability, but 3.6 per cent of team members had doubts about was 5.6, "Nursing home practices support the resident's right to die with dignity."
8. The standards for which significant minorities of both directors of nursing and team members harboured doubts about practicality were: 1.2, resident participation in care planning; 2.2, resident control of financial affairs; 3.1, policies developed in consultation with residents; and 7.1, resident's right to participate in activities which involve risk. Two of these – 1.2 and 3.1 – were also among the standards most often mentioned adversely in our unstructured interviews with proprietors. In addition, a minority of proprietors also had concerns about the practicality of standard 2.3, freedom of movement within and from the nursing home and standard 4.1, homelike environment.
9. The methodology we have followed with our interviews and questionnaires is designed to give the different players of the regulatory game an opportunity to indicate their doubts about the face and content validity of each of the standards. Statistically, surprisingly few doubts were raised by this process. The consultants see some of the doubts raised as primarily going to the face validity of the standards in the eyes of a small minority rather than raising fundamental concerns about the content validity of the standards. Fundamental concern is raised, however, by our observations of the non-comprehensiveness of the process for gathering information to rate the standards.

4. CONCURRENT VALIDITY

Concurrent validity is concerned with whether a rating correlates positively with other measures of the same or a similar concept assessed at the same point in time. In this chapter, we explore two approaches to concurrent validation – comparing the director of nursing's rating of the home with the team's rating and comparing team ratings with a global assessment of the quality of care in the home by the team. We will start with the latter and then go on to assess agreement between the director of nursing and the team ratings of the home.

Validation Against a Global Assessment of Quality

Each team which assessed one of the 410 nursing homes in the first wave of the quantitative study, were asked to fill out a short questionnaire on the nursing home. Completed questionnaires were returned by 405 teams. Most were filled out by a single member of the team and those cases where there was doubt about other members of the team's views, they were consulted and a consensus reached.

The team was asked "How would you rate the quality of care in this nursing home compared with all the nursing homes members of the Team have experienced? Would you say this home is well above average (in the top twenty per cent), above average, average, below average or well below average (in the bottom twenty per cent)?" They were supplied with a five-point scale marked with these labels on which they could indicate their answer. Some previous research suggests that such global ratings of nursing homes by experts can achieve reasonable inter-rater reliability (Gustafson *et al.*, 1977)

As this single item is a global measure of the perceived quality of care in the nursing home, if it does not correlate positively with our total compliance measure we should be very concerned about the validity of the latter. Of course, we might also be concerned if there were a perfect correlation between the two. That might suggest that there is little point in the business of detailed assessment of each of the 31 standards as a global "seat of the pants" judgment gives exactly the same result.

In fact, the correlation between the two is .64, providing satisfactory validation. If it is true, as concluded in Chapter 2, that each of the standards is measuring something different (so that none should be combined) but that each is also measuring something that is common to all the standards, then we would expect each of the standards to have a positive

correlation with the global compliance measure. We would certainly want to look carefully at any that does not.

Table 4.1 lists these correlations. All correlate in the expected way with the global validation measure. The standards with the lowest validation are: 1.3, freedom from pain; 2.2, resident control of financial affairs; 2.5, residents enabled to maintain responsibilities as citizens; 5.6, nursing home practices support the right to die with dignity; 7.4, protection from fire and natural disasters; and 7.5, security of buildings and contents.

We can make very good sense of these low validation results. The last two standards 7.4 and 7.5 may correlate lowly with an assessment of overall "quality of care" because this assessment is made rather independently of the structural problems of the buildings where that care must be provided.

What is particularly notable about three of the standards with low correlations with the global validation measure is that they are the three standards which attract the highest proportions of met ratings – 1.3, 2.5 and 5.6. Our qualitative fieldwork suggests that these three standards almost always get met ratings because standards monitoring teams do not have very effective means of gathering information on them. Put another way, these are important standards which would be strongly correlated with global quality of care if the problems which existed under them were being detected, but these problems are not being detected.

On pain management (1.3), the American process is much more successful than the Australian process in detecting problems because of its more thorough cross-checking of documentation. This involves checking to see that corresponding to every recorded instance of pain there is evidence of care to manage the pain: American inspectors also directly observe pain management in the giving of treatments. They put more emphasis in resident and staff interviews on enquiring about pain and because their superior drug regimen review also frequently uncovers pain problems.

Standard 2.5 is practically interpreted by many teams as little more than a right to vote in elections. Since elections are rarely underway at the time of standards monitoring visits, evidence is a problem. Many teams are satisfied to give a met rating if they see some newspapers around the nursing home and if the director of nursing tells them that provision

Table 4.1: Correlation of standards monitors ratings with a rating of how the quality of care in the nursing home compares with other nursing homes^a

	Correlation
Objective 1: Health care	
1.1 Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed.	.28
1.2 Residents are enabled and encouraged to make informed choices about their individual care plans	.41
1.3 All residents are as free from pain as possible.	.21
1.4 All residents are adequately nourished and adequately hydrated.	.44
1.5 Residents are enabled to maintain continence	.45
1.6 Residents are enabled to maintain, and if possible improve, their mobility and dexterity.	.38
1.7 Residents have clean healthy skin consistent with their age and general health.	.30
1.8 Residents are enabled to maintain oral and dental health.	.23
1.9 Sensory losses are identified and corrected so that residents are able to communicate effectively.	.37
Objective 2: Social independence	
2.1 Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts	.37
2.2 Residents are enabled and encouraged to maintain control of their financial affairs.	.22
2.3 Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.	.37
2.4 Provision is made for residents with different religious, personal and cultural customs	.31
2.5 Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.	.22
Objective 3: Freedom of choice	
3.1 The nursing home has policies which have been developed in consultation with residents and which:	
- enable residents to make decisions and exercise choices regarding their daily activities	
- provide an appropriate balance between residents' rights and effective management of the nursing home	
- and are interpreted flexibly taking into account individual resident needs.	.46
3.2 Residents and their representatives are enabled to comment or complain about conditions in the nursing home	.40
Objective 4: Homelike environment	
4.1 Management of the nursing home is attempting to create and maintain a homelike environment	.44
4.2 The nursing home has policies which enable residents to feel secure in their accommodation	.28
Objective 5: Privacy and dignity	
5.1 The dignity of residents is respected by nursing home staff.	.40
5.2 Private property is not taken, lent or given to other people without the owner's permission.	.40
5.3 Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private	.42
5.4 The nursing home is free from undue noise	.27
5.5 Information about residents is treated confidentially.	.27
5.6 Nursing home practices support the resident's right to die with dignity.	.25
Objective 6: Variety of experience	
6.1 Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.	.43
Objective 7: Safety	
7.1 The resident's right to participate in activities which may involve a degree of risk is respected.	.36
7.2 Nursing home design, equipment and practices contribute to a safe environment for residents, staff and visitors	.45
7.3 Residents, visitors and staff are protected from infection and infestation.	.40
7.4 Residents and staff are protected from the hazards of fire and natural disasters.	.24
7.5 The security of buildings, contents and people within the nursing home is safeguarded.	.26
7.6 Physical and other forms of restraint are used correctly and appropriately.	.33

^aQuality of care is a 5-point scale ranging from 5 (well above average) to 1 (well below average). The standards are rated on a three point scale - met, met in part, and not met.

is made for residents who wish to vote at election time. Recall is a problem with interviewing residents on this matter if the last election was a long time ago. There is a need for a wider view to be taken of this standard and for more serious attention to be devoted to it during the standards monitoring process.

Standard 5.6 puts the team in a difficult situation. It is intrusive and inappropriate to gather evidence by interviewing residents about whether their wishes concerning terminal care have been identified, as it is to rely on observational evidence at a death. Nor can teams demand the input of say systematic recording on admission of residents' terminal care wishes. When there is no documentary evidence of terminal care preferences, teams enquire of directors of nursing as to how they know and discover such wishes. If the director of nursing demonstrates some sort of informal, but systematic and sensitive attention to the subject, then the team is satisfied. This standard then may be one that is inherently unsatisfactory at an evidentiary level. A piloting of new approaches to come to grips with the standard might be considered in consultation with a bereavement expert. Perhaps retrospective asking of residents if the wishes of recently deceased room-mates had been respected is one approach. Another would be to raise this issue in a group discussion with the residents' committee, where residents who preferred not to discuss the issue could remain silent.

It is notable that four of the standards which featured in the last chapter as standards about which some in the industry had expressed doubts were among the standards which received the strongest validation here. These are 1.2, resident choice about care plans; 1.5, maintaining continence; 3.1, policies developed in consultation with residents to ensure freedom of choice; and 4.1, homelike environment. Note also that these four are all standards which we found in Table 2.1 to have particularly strong item-total correlations. It may be then that some of the standards for which the anxieties of a minority question face validity are among the very best standards for getting to the heart of quality of care.

Director of Nursing Agreement With Team Ratings

For all standards, at least 84 per cent of the time, directors of nursing gave themselves the same rating as the team gave them (see Figure 4.1). The average level of agreement across the 31 standards was 92 per cent. To determine the extent to which directors of nursing and the teams were in agreement, the directors of nursing were asked whether they agreed with the ratings given by the teams. From this information a measure of overall agreement was calculated. Overall agreement means that if the team rated their home met, the director of

nursing must think it was met; or if the team rating was met in part, the director of nursing thought met in part; or if the team rating was not met, the director nursing thought not met. The standards with the lowest levels of agreement were 1.1, appropriate medical care from a medical practitioner of the resident's choice (86 per cent agreement) and 4.1, homelike environment (84 per cent). The correlation between the team's total compliance score and the director of nursing's self-rated compliance score was .88.

Perhaps surprisingly, sometimes the director of nursing gave tougher ratings than the team (for example a not met when the team gave it a met). This happened in no fewer than 55 instances. For example, one of our interviewers paraphrased the following reasons a director of nursing gave for why not met was the right rating on 7.5 when the team gave the nursing home a met:

Over the period of the last year, a man has made phone calls and exposed himself. Only when staff called the union was anything done. The man could not be identified because lights were broken and management had not bothered to fix them. "We didn't tell the team because they didn't ask."

Naturally, however, the reverse sort of disagreement (with the team giving the tougher rating) was much more common. Standard 1.1 attracted a lot of disagreement with team ratings from directors of nursing who viewed it as unfair that they were held responsible for the failures of doctors, particularly in the area of filling out treatment sheets: "I can ask the doctor. I cannot get hold of his hand and write." The department has attempted to address this problem with a guideline that requires the nursing home to raise with the doctor the problem of non-compliance. If the doctor then refuses to comply, the nursing home may consult the resident. If the resident says she wishes to stick with this doctor, notwithstanding his non-compliance, then the nursing home can be rated met, in spite of the doctor-driven non-compliance.

Directors of nursing disagreed with ratings on 4.1, homelike environment, for three main reasons. First, they disagreed with what they saw as a subjective interpretation of homelike by teams. Second, they felt that what the teams saw as homelike sometimes was at odds with what the residents wanted. Third, they felt they should not be marked down because they had the impossible task of making a building constructed as a hospital look like a home.

The results in Figure 4.1 are very good ones for the acceptance of the standards monitoring program. However, we should look behind them because of the possibility that

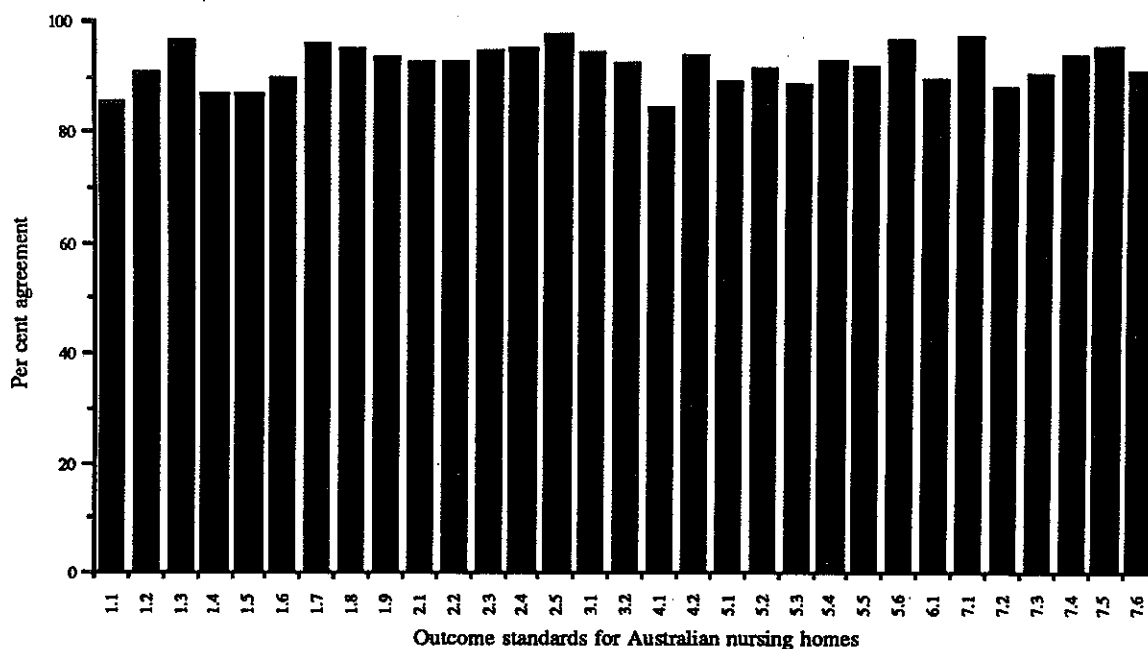


Figure 4.1: Per cent of overall agreement of directors of nursing with the rating given them by the standards monitoring team (n = 410)

most of the agreement is accounted for by teams and directors of nursing agreeing on met ratings. Figures 4.2 and 4.3 show that the level of agreement when teams issue not met and met in part ratings indeed is much lower. Agreement of directors of nursing when the team gives them a not met ranges from a low of 39 per cent on standard 5.4 “The nursing home is free from undue noise” to a high of 88 per cent on standard 7.1 “The resident’s right to participate in activities which may involve a degree of risk is respected”, a surprising result in light of the concerns expressed about this standard in the last chapter. With some variation between standards, approximately half of the director of nursing disagreements with not met ratings were cases where the director of nursing agreed they did not fully meet the standard, but where they thought met in part should have been the fair rating.

Other standards with low levels of agreement from directors of nursing for not met ratings were 1.4 “All residents are adequately nourished and adequately hydrated” and 5.6 “Nursing home practices support the resident’s right to die with dignity”. The latter is not surprising given the evidentiary difficulties discussed earlier with this standard: “This standard cannot be judged unless the team is there at the time. Dying is a long process – counseling and guidance and so forth are involved.” Comments from directors of nursing suggest that noise and food may be simply matters where there is a lot of room for different opinions on what is acceptable and what is not. Disagreement with noise ratings often arose from situations where the director of nursing felt the team encountered an atypically noisy situation in the nursing home (for example, because of renovations).

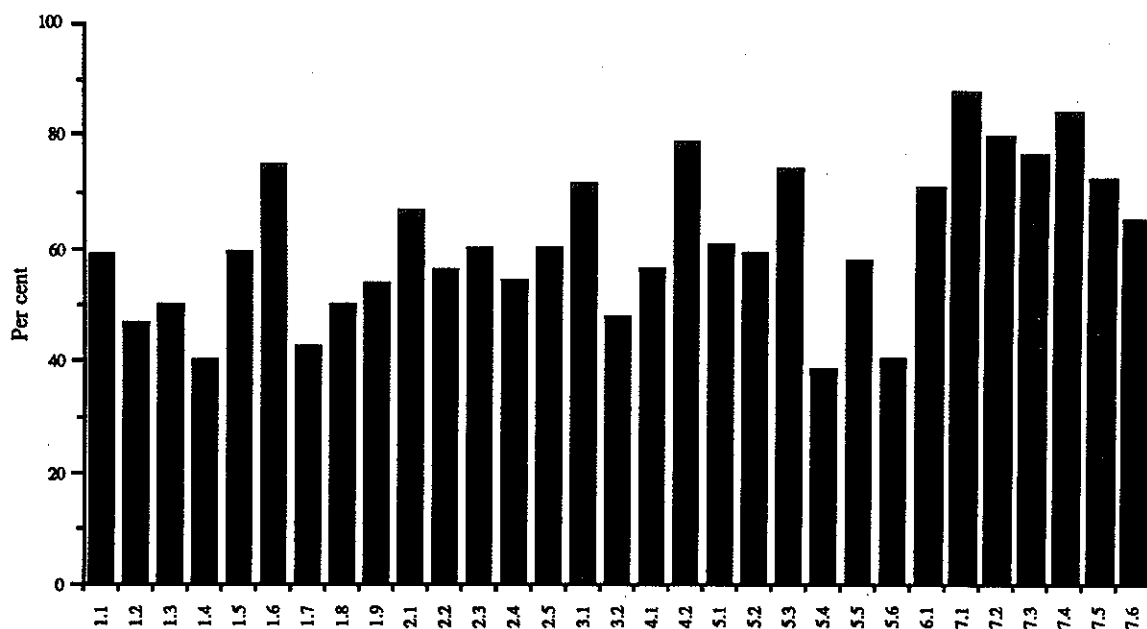


Figure 4.2: Per cent of director of nursing agreement with the standard monitoring teams' rating of 'not met'
(Standard (n=); 1.1 (61); 1.2(30); 1.3(6); 1.4(25); 1.5(47); 1.6(24); 1.7(14); 1.8(10); 1.9(13);
2.1(6); 2.2(23); 2.3(15); 2.4(11); 2.5(5); 3.1(28); 3.2(23); 4.1(32); 4.2(23); 5.1(46);
5.2(22); 5.3(85); 5.4(13); 5.5(26); 5.6(5); 6.1(34); 7.1(16); 7.2(96); 7.3(55); 7.4(110);
7.5(25); 7.6(51))

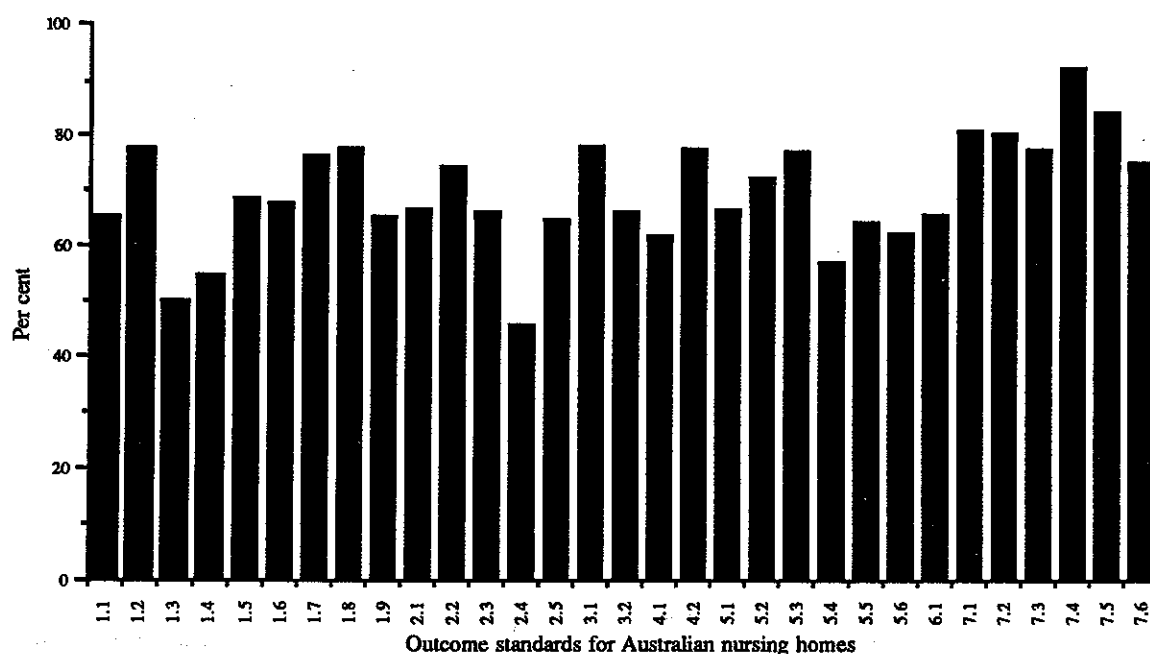


Figure 4.3: Per cent of director of nursing agreement with the standard monitoring teams' rating of 'met in part'
(Standard (n=); 1.1(95); 1.2(68); 1.3(18); 1.4(84); 1.5(93); 1.6(97); 1.7(25); 1.8(58);
1.9(52); 2.1(79); 2.2(51); 2.3(44); 2.4(24); 2.5(17); 3.1(69); 3.2(56); 4.1(130); 4.2(86);
5.1(70); 5.2(92); 5.3(92); 5.4(47); 5.5(56); 5.6(24); 6.1(93); 7.1(42); 7.2(138); 7.3(116);
7.4(96); 7.5(59); 7.6(69).)

The lowest levels of agreement with met in part ratings by the team were on 1.3, "Residents are as free from pain as possible", and 2.4, "Provision is made for residents with different religious, personal and cultural customs" (see Figure 4.3). On pain

management, disagreements arose over whether there was a sufficient pattern of poor pain outcomes to justify other than a met rating: "Only one resident in 144 was in pain". Similarly, with providing for residents of different religious or cultural backgrounds, there was criticism of teams for giving too much weight to a single case of, for example, "one Greek lady who is demented."

When we look at the 889 cases where directors of nursing gave us detailed reasons as to why they disagreed with the rating the team gave them⁵, in 26 per cent of cases the director of nursing simply disagreed with the team's interpretation of the standard. The next most common reason for disagreement with their rating (16 per cent of cases) was that they believed there was nothing they could do about the problem because it was the fault of someone other than nursing home management (for example, doctors, residents, renovation workers). Following this (12 per cent of cases) was the complaint that the evidence against the nursing home did not constitute a pattern of harm – the evidence was of a one-off incident, minor or nit picking.

Proprietor: "If the attitude is if one resident or relative is not happy, you may as well throw the whole thing out if you are going to be rated met in part or not met. There is no way known to man you are going to please all of the people all of the time. That is an impossibility, and if that is their attitude, you are doomed to failure."

The next most common reason given for disagreement related to the teams examining inputs or processes (mostly documentation inputs) rather than outcomes (9 per cent of cases), and the view that the team's expectations were impossible due to resident disability (9 per cent). Then came the view that the team got it wrong through erroneous observation (7 per cent).⁶ Sometimes directors of nursing had clear cut reasons for rejecting the team's observations: "[The team said] the resident only had one slipper. What the team member did not realise was that the resident only had one leg". Next in frequency as a reason for rejecting team ratings was that the rating was seen as inconsistent with the ratings other teams were known to have given this or other nursing homes (6 per cent). Then there were directors of nursing who disagreed with ratings because they believed the structure of their building made compliance impossible (6 per cent). The last of the common sources of disagreement was that residents preferred things the way they are (5 per cent). In a sense, the latter is related to the issue of teams focusing on outcomes. In addition to these more

⁵ There were also 925 cases where directors of nursing gave us reasons as to why particular standards were undesirable, impractical, or generally a problem; and 250 cases where directors of nursing gave us reasons as to why they saw particular standards as unclear.

⁶ In addition, for 12 per cent of cases, it was simply said that the team got it wrong because "this is the way it is". That is, it was suggested that the team got its facts wrong without specifying whether this

common reasons for disagreement, there were a myriad of more specific reasons which applied in smaller numbers of cases.

Summary

1. Concurrent validity was supported by: (a) a correlation of .64 between the total compliance score and a standards monitoring team assessment of how the nursing home compared with other nursing homes in terms of quality of care; and (b) a correlation of .88 between the officially measured total compliance score and total compliance as self-assessed by directors of nursing.
2. Individual standards with the weakest validation against the global assessment of quality of care were: 1.3, freedom from pain; 2.2, resident control of financial affairs; 2.5, residents enabled to maintain responsibilities as citizens; 5.6, nursing home practices support the right to die with dignity; 7.4, protection from fire and natural disasters; and 7.5, security of buildings and contents. Concern was expressed that the reason three of these standards were only weakly validated (1.3, 2.5, 5.6) may have been that they are standards for which the standards monitoring process does not dig deep enough to find non-compliance when non-compliance exists.
3. Strong validation against the global assessment of quality of care was found for four of the standards which featured in Chapter 3 as standards about which some in the industry had doubts. These were 1.2, resident choice about care plans; 1.5, maintaining continence; 3.1 policies developed in consultation with residents to develop freedom of choice; and 4.1 homelike environment.
4. The standards with the lowest levels of agreement of directors of nursing with the ratings given them by teams were 1.1, appropriate medical care from a practitioner of the resident's choice and 4.1, homelike environment.
5. When not met ratings are issued, the level of agreement from directors of nursing varies enormously between standards – from a low of 39 per cent agreement on standard 5.4 (undue noise) to a high of 88 per cent agreement on standard 7.1 (resident's right to take risks). For about half the cases of director of nursing disagreement with not met ratings, the director of nursing did not feel that the nursing home had met the standard; rather they disagreed because they felt met in part would have been a fairer rating.
6. In order of importance, the major reasons for disagreements of directors of nursing with team ratings were: (a) rejection of the team's interpretation of the standard; (b) belief that there was nothing nursing home management could do about the problem because it was someone else's fault (e.g. doctors, residents, renovation workers); (c) belief that the evidence against the nursing home was a one-off incident rather than a pattern of harm; (d) belief that the rating was based on inputs or processes rather than outcomes; (e) the view that the team's expectations were unreasonable due to resident disability; (f) the claim that the team got it wrong through erroneous observation; (g) belief that the rating was inconsistent with the ratings given by other teams or on other occasions; (h) concern that the structure of the nursing home building made compliance impossible; and (i) residents preferred things the way they are. It is noteworthy, if unsurprising, that some of these reasons for disagreement were not at all important as reasons for disagreement between the independent rater and the team in the inter-rater reliability study. Reasons for

was because of erroneous observation, misreading of documentation, or misinformation supplied by a staff member, resident or visitor.

disagreement from the above list which were rarely if ever important to the independent raters were: (b) belief that there was nothing nursing home management could do about the problem because it was someone else's fault; (e) the view that the team's expectations were unreasonable due to resident disability; and (h) concern that the structure of the nursing home building made compliance impossible.

5. COMPARISON WITH AMERICAN STUDIES

While this is the most systematic study of the reliability and validity of the compliance judgments of government regulators that we know of, it could have been done better. The sample in the inter-rater reliability study is certainly considerably larger than in all the American nursing home inspection reliability studies, but it is still too small to enable confidence about the reliability of some of the individual standards.

The most serious deficiency is the want of a predictive validation study. Knowing that a home has an outstanding compliance score should enable one to predict that certain adverse outcomes will be less common for residents in the year after the standards monitoring visit. Most homes with excellent compliance with quality of care standards should have no decubitus ulcers in their nursing home, but when they do have them, we should observe reductions in their number and severity over time. We would expect most homes with excellent compliance to have little or no physical or chemical restraint, but where they do have a number of restrained residents, we should see this number fall. In nursing homes with poor standards, we would expect to see a variety of indicators of well-being deteriorate (such as Activities of Daily Living scores) at a faster rate than in the excellent homes. While in the excellent homes we would expect to see some residents deteriorate, we would expect to see improvements in others in mobility and dexterity, communication, continence, participation in activities and other outcomes as a result of restorative nursing and other rehabilitative programs. Unfortunately, in Australia we do not at this time have the data to make such validation work possible. It is, however, something we must aspire to do in the future.

The United States is the country that in the past has undertaken most of the work on the reliability and validity of the ratings of nursing home inspectors. In fact, it is to our knowledge the only country where systematic reliability studies have been undertaken. This reflects the fact that nursing home regulation has been a far higher priority for United States governments for a longer time than anywhere else in the world. Nursing home inspection agencies are much better resourced than elsewhere; the inspection protocols they follow are seemingly much more sophisticated and certainly cover a larger number of issues (currently over 500 federal regulations).

Our assumption in approaching this research was that the Commonwealth government was new to comprehensive assessment of the quality of nursing home life and that we would have much to learn from the more sophisticated approach that the United States had

built up over decades of experience and research. Certainly, when we first went to the United States with Australia's brand new 31 outcome standards in 1987, we were a little embarrassed by them. We would say to the American regulators: "We know they are only a start and we have a lot of work to do to flesh out the kinds of guidelines and protocols that you have built up over the years." And they in turn would look aghast at how broad, vague, undefined and seemingly unenforceable our standards were. So they would agree that indeed we had a long way to go.

To our surprise, we now find, however, that the broad and unrefined Australian standards appear much more reliable than the American standards. So we must now confront the paradox that all those years of making the American standards more tightly defined, of breaking them down into more specific standards, of responding to refusals of the courts to enforce vague standards, may have been counter-productive. But first let us review this American evidence.

The first major empirical study of nursing home inspection was the Wisconsin Quality Assurance Project. That project piloted its own quality of care measure which was independent of the state regulatory process (Gustafson *et al.*, 1980). It was a much simpler measure, based on just 11 criteria, than that used by Wisconsin state inspectors. Five two-person teams of nursing home professionals visited 9 nursing homes, giving 45 data points for the calculation of reliabilities. An average reliability coefficient on this simple measure of quality of care of .78 was obtained. This measure was also validated against a global 0-100 assessment of the quality of care of each home ($r=.76$), the global assessment itself having been found to be reliable (Gustafson *et al.*, 1977). This pilot therefore demonstrated that 2-person teams could rate the quality of care in nursing homes with reasonable reliability. Unfortunately, when the two measures of quality of care correlated with the number of deficiencies cited by the last government inspectors to visit the nursing home, the correlations of $-.53$ and $-.11$ were not statistically significant at the .05 level. This early study thus gives us the first clear clue to the direction in which we are led by our results. This is that reliable ratings of the quality of care in nursing homes are possible when professional raters use a limited number of criteria; but when raters use the large number of specific American regulations as their criteria, reliability is lost.

This was the pilot study on the Wisconsin quality of care instrument. In the final study 12 homes were visited by 3 teams (36 data points). The reliability and validation of the simple quality of care instrument improved slightly compared to the pilot, but a very low

association of this simple instrument with the number of citations issued by government inspectors of -0.12 was found on another sample of 65 homes (Gustafson *et al.*, 1982).

Another particularly discouraging finding of the Wisconsin project was summarized in the final report as follows:

The final important result in problem identification came from a comparison of problems identified by federal validation teams versus the QAP and traditional processes. There were no differences in methods in terms of the number of conditions found out of compliance. However, there were substantial differences between state and federal teams at the standard and element level (Gustafson *et al.*, 1982: 5).

This sounds discouraging but it was actually something of an understatement. It is true that "There were no differences ... in the number of conditions found out of compliance". Neither the state inspectors nor the federal validation inspectors found any conditions out of compliance, so there was perfect agreement! Condition-level was the most serious level of non-compliance rating, and was rarely given. A freedom of information request to the funding agency, the United States Department of Health and Human Services, by Dr. John Gardiner revealed that at the next level of seriousness of citation – the standard level – the federal validation team had cited 3 standards as out of compliance for the 20 homes in the study, while the state validation team cited 28. Total deficiencies cited at the standard and element levels (the latter being the lowest level of seriousness) for the 20 nursing homes were 437. There was agreement on only 8 per cent of these between the state and federal validation teams. For 92 per cent of these deficiencies, one team was citing something that the other team had not cited.

One reason for these disturbing differences could be that the federal validation surveys (inspections) were done on average 30.5 days after the state survey. However, as the internal memorandum secured by Dr. Gardiner's freedom of information request pointed out: "While changes at the facility between surveys do cause differences in findings, most facility changes are corrective actions which would reduce the deficiencies between the first (State) and second (Federal) surveys." And in fact deficiencies rose sharply between the first set of inspections and the second. Moreover, as the memorandum pointed out, "The majority of deficiencies found on the second (Federal) survey existed, and should have been identified during the first (State) survey."

Another major University of Wisconsin study was published in 1985 on data from New York, Massachusetts and Wisconsin (Zimmerman *et al.*, 1985). While the number of data points for the reliability coefficients was only 13 (2 teams visiting 13 homes in New

York and Massachusetts), the results were much better. This time, with the teams in the nursing homes at the same time and in different states, 58 per cent of deficiencies cited were cited by both teams. An impressive 84 per cent of deficiencies detected by state teams were also detected by independent teams, though there were a much larger number of deficiencies detected by independent teams which were not detected by state teams.

Both teams returned to these nursing homes four months later to assess whether the deficiencies on which the two teams agreed had been corrected. The state teams judged 96 per cent to be corrected and the independent teams 71 per cent. In this study multiple regressions using the Wisconsin quality of care indicator found only weak validation of the number of deficiencies detected by inspectors ($p < 0.1$), but much stronger validation of the total severity of deficiencies detected ($p < 0.05$).

The third American reliability study of the ratings of compliance with nursing home regulations was based on double inspections of 21 Tennessee homes (Spector *et al.*, 1987). There was a one-day interval between visits to the nursing home. Both teams were inspectors from the Tennessee Department of Health and Environment. Both teams were unusually large (averaging 8.7 for the (official) first team and 5.5 for the second validation team) and both had unusual breadth of disciplinary coverage – always including nurses, a generalist, a social worker, physiotherapist and pharmacist. The official team also included a dietician, sanitarian and fire inspector. Moreover, they came from one of the most competent state inspection agencies in the United States.

Only 25 per cent of the regulations cited by the official team were also cited by the validation team, and this when the validation team cited twice as many regulations as the official team (Spector *et al.*, 1987: 119-23). Again, one might have expected that the second team would have found less because the nursing home would have acted to correct the deficiencies detected by the first team. However, the second team, in fact, found twice as many deficiencies as the first.⁷

⁷ Another aspect of the same project was to compare the effect of this new survey process (PaCS) with the then existing process on deficiencies cited. Both processes were rating identical standards. However, the new PaCS process, which became the official U.S. process from 1986 to 1990, was slightly more resident-centred than the old process (though nowhere near as resident-centred as the Australian process). This quasi-reliability test was conducted on a larger sample of 51 nursing homes, with the second team entering the home within a day or two of the first departing. PaCS teams cited 50 per cent more deficiencies than team surveying the same standards under the less resident-centred process (Spector and Drugovich, 1989).

This last study is particularly discouraging for the American process as it is both the most recent study and the study on the largest sample of nursing homes. How can we explain the paradox that in spite of the fact that in the Tennessee reliability study the inspections involved five times as many inspector-hours in the nursing home compared to Australian independent raters, in spite of the fact that they were using multidisciplinary teams with more sophisticated protocols on more precisely specified standards, the Americans get appalling reliability, while the Australians get outstanding reliability? Even if we take the American study with the best results – the 13-home Wisconsin study – we can see that with a larger sample than any of the American studies, the Australian reliability result is considerably better than the best American result. How can this result, which was the exact opposite of our expectations, make sense?

Our qualitative fieldwork in the United States, during which we have observed 40 nursing home inspections in 24 states, as well as interviewing most of the key players, suggests some answers to this puzzle. The American case illustrates what we call the reliability paradox: the result of persistent political pleas to increase the consistency of regulatory inspectors is to increase their inconsistency. Historically, what has happened in the United States is that all of the key political players in the nursing home regulatory game have been critical of broad, vaguely defined standards. The industry has been at the forefront of this criticism; when nursing home X gets a not met rating on a broad standard on which nursing home Y in similar circumstances gets a met rating, nursing home X screams about inconsistency. It complains to its industry association about the vagueness of the standard inevitably leading to “subjective” and “unfair” judgements by inspectors. The industry association, representing these member grievances, pleads for the standard to be “tightened up”. Consumer groups also agree that the standards should be made tighter and more specific, but for different reasons. They are concerned that vague standards are unenforceable. Legislators are responsive to these pleas because they feel frustrated that inspectors are not cleaning up the industry the way they had hoped; their analysis fits nicely with that of the industry and consumer groups. These standards, the legislators conclude, are so vague that they give the inspectors too much discretion to subvert the legislative mandate. This indeed is also the analysis of many top regulatory bureaucrats in the United States federal government. They are frustrated at the failure of the states to deliver their hopes. Part of the blame they lay at the door of standards so vague as to allow wide discretion for inaction. Finally, the technocrats – the behavioural and medical scientists and the lawyers – despair at vague standards. The scientists believe in tight protocols to ensure that the same things are being assessed in exactly the same way using precisely defined criteria. The lawyers believe, like the consumer advocates, that vague standards are difficult

to enforce in the courts, and like the industry associations, they believe that vague standards result in abuse of discretion.

Hence, if there is one thing that all of the influential players of the American regulatory game have agreed upon it is that broad standards that are not tightly specified are a bad thing. The consequence has been an historical process of all these constituencies succeeding in having one broad standard broken down into two narrower standards; then later each of those two standards being subdivided into three standards. Not only does the number of standards grow, their content changes also. The impetus to reform "subjectivity" in standards fuels a search for "objective" criteria and protocols for rating standards. Ironically, this undercuts one of the other things the industry, politicians, consumer groups and scientists (if not the lawyers) agree on – that the regulatory process should be more outcome-oriented and more resident-centred. The trouble is that inputs (the temperature of the food as it leaves the serving line) are generally more "objective" than outcomes (satisfaction of residents with the temperature of the food). Documentation-centred standards are generally more objective than resident-centred standards. For example, it is believed that to rate access to activities programs, it is less objective to ask residents if they are satisfied with their access to activities programs, more objective to count ticks beside residents' names in an activities book. When there is a protocol requiring inspectors to count those ticks, the belief is that two different inspectors will come up with the same answer. When the process allows inspectors to wander around the nursing home simply chatting with residents about activities, the variable and subjective judgements which result are bound to be unreliable, or so it is believed. Even though the key players want the regulatory process in the abstract to become more outcome-oriented and resident-centred, there is exactly the opposite effect from the aggregation of their demands about what should be done to make particular standards less subjective. Input- and documentation-centred standards appear more reliable.

Our claim is that this appearance is misleading. This is a counter-intuitive claim. It is not radical to argue that input- or documentation-centred standards will not be valid. The employee can put ticks beside the names of residents who never attend the activities program; or they can put ticks beside the names of sleeping residents who are wheeled into the room to make up the numbers. But we are making not only the claim that this approach is less valid, but also that it is less reliable. It is less reliable because of the context in which it occurs. That context is one where standards have multiplied to staggering numbers and where for many of these standards there exist detailed protocols for counting inputs. In the United States, there are at present over 500 federal regulations ("tag numbers"), and in

some states there are even more state standards than this which are assessed by the same inspectors at the same time as they rate compliance with the federal standards. The problem is that no human being can reasonably keep all these standards and all these protocols in their heads as they move around the nursing home looking for non-compliance.

How then do American inspectors cope? They cope by giving up on keeping the standards in their heads. At its best, the American process works in the following way. The inspectors meet together at the conclusion to their information gathering, as they do at certain intermediate points during the inspection, to share the problems they have found. When a number of negative findings are judged to constitute a pattern of non-compliance of a particular type, a search begins for a "Tag number" which can be written up as not met. Once all the problems have been agreed and Tag numbers found to write not mets for them, the team leader ticks met for all the remaining standards. As she does so, she does not read them or think about them and she certainly does not check with her colleagues that someone has collected the information necessary to reach that met rating. Usually, she will not discuss with her colleagues the possibility that the same pattern of conduct that caused one standard to be rated not met should also cause several other standards to be out of compliance (for example, an inappropriate use of restraint may cause standards concerned with restraint, following doctors orders, resident choice, mobility and freedom of movement to all be out of compliance). In other words, she makes one valid not met rating and several invalid met ratings as a result of this strategy. This we said was the American process at its best. At its worst, the team partitions responsibility for the standards, each writing up their own standards with little input from the other team members.

One cannot be critical of American inspection teams for this. There is no other way they can cope with so many standards. There is no way they can cope with so many protocols under those standards than to do what is necessary to fill out the forms mandated for certain protocols and essentially fudge the other protocols that cannot be checked. Realising that this is the way the game is played, advocates of protocols for certain standards that they regard as especially important lobby the federal government to mandate auditable protocol forms that the state inspectors must fill out. While this improves the attention given to the lobbyist's cherished regulation, it further worsens the structural malaise of the process.

What is the relevant contrast with the Australian process in this regard? It is not easy for Australian teams to keep 31 standards in their heads even though none of them have mandated protocols. Yet they can make a fist of it. More critically, after their visit the team

can (and generally does) sit down to discuss, standard by standard, the evidence collected by all team members relevant to each one. This dialogue is formalized by the team agreeing on a list of positives and negatives to be written beside each standard. Sometimes they will find that they have not collected the data necessary to reach a reliable rating on the standard. They must then take steps to collect the extra information. There is no escape from this because the team is required to sit down with the management of the nursing home, to summarize the positives and negatives on each standard and give reasons for their final ratings. Again the American "exit conference" is different in that it reports only exceptions. Nothing is said about standards that have been ticked met even though the team has neither debated compliance with them nor assured themselves that they have collected the data relevant to them. The crucial difference is then that Australian teams actually do deliberate on all their standards and collect the evidence that they judge sufficient to support that deliberation.

It is not the fault of American inspectors that they do not do this – the number of standards and protocols that they must live with make this quite impossible. The end result of demands for more specific standards with more clearly defined protocols that cover all of the things judged to be important for nursing homes is an inspection package that is structurally unreliable. The pursuit of the reliability of parts causes the unreliability of the whole.

Even at the level of the parts, when we look at the long list of United States standards many, even most, of them look mightily "subjective". Efforts to conquer one symptom of subjectivity simply seem to introduce new ones in the new standards. This is hardly a surprising finding when the domains which are our ultimate concern – quality of care and quality of life – would seem to be rather fundamentally subjective. Yet as a result of this futile war against subjectivity, we end up with a Byzantine set of regulations. The upshot is certainly individual regulations which flagrantly flout Diver's (1989) desiderata for regulatory standards of transparency, congruence and simplicity.

There is an additional sense in which the pursuit of reliability for parts undermines the reliability of the whole. We have said that the American team cannot avoid following some protocols which are carefully designed with reliability in mind because their performance is evaluated by ensuring that certain federally mandated forms (relating to these protocols) are filled out. This does not mean that the regulations that require these data will be reliably rated. The emphasis on monitoring team compliance with protocols susceptible to paper evaluation mean that teams lose their ability to see the wood for the trees. The busy work of

documenting evidence in relation to the types of trees that will be checked mean that all other types of trees are barely noticed and reflection on the state of the wood as a whole is neglected.

Yet another way in which the paradox of reliability has come about in the past has been on the question of the random sampling of residents. The behavioural and medical scientists who were influential in shaping the American process as it emerged in the 1980s believed that randomness was important to valid and reliable ratings. The old fashioned inspection practice of allowing inspectors to concentrate their evidence-gathering on residents of their choice was viewed as unscientific. Many of the key players in the industry associations were also vigorous advocates of random sampling, but for different and perhaps more sophisticated reasons. Some regulators alleged that these industry players supported random sampling because it would inhibit inspectors from following their noses to the residents who were getting the worst deal out of the nursing home. The lawyers had a hand in this shift as well. Up until October 1990, when the United States abandoned random sampling of residents for nursing home inspections, standard training practice would confront the American inspector with the scenario of a company lawyer challenging her findings by questioning her competence in the statistical theory which would warrant the judgment that a pattern of non-compliance existed.

Our observations of the random sampling process revealed endemic cheating by inspectors. They would cheat for both principled and unprincipled reasons. When on the initial tour of the nursing home, the inspector met a resident who complained of mistreatment or who manifested signs of neglectful care, the inspector would sometimes cheat by putting that resident into the random sample even though she was not randomly selected. This was principled cheating. Examples of unprincipled cheating are the following. The team member meets on the tour a resident who is a friend she enjoys talking to. After a twenty minute chat with her over lunch, she realises that she has already collected half the information she needs from this resident. So she slips her into the random sample. In another type of fudging repeatedly observed, the inspector finds a resident with multiple problems – restraint, catheter, decubitus ulcer and others. Because the sampling protocol demands a number of residents with each of these types of special problems, by slipping into the sample one resident who has all of the problems, the total number of residents who must be investigated is reduced. We say these latter examples are of unprincipled cheating, but the teams do not view it this way. We have already made the point that systematic data collection to rate over 500 regulations is impossible; the cheating, they contended, made an impossible job a little more possible.

We can see then how the pursuit of reliability and validity by random sampling failed. Either the team cheated by slipping bad cases into the sample, thereby defeating randomness, or they refrained from cheating, thereby neglecting the trees in the wood that were about to fall in order to get their documentation right on the randomly selected trees. Moreover, the process of random sampling itself was extremely time consuming, often occupying many person-hours of collecting a list of residents with all data relevant to the sampling strata, selection to ensure the correct number in each category (variably according to the number of qualified residents in the nursing home), and recording of the selected sample. The sampling process therefore absorbed a significant proportion of the very limited time that inspectors had for gathering data on all the regulations.

We can see that in a number of ways the pursuit of more tightly specified standards and protocols can undermine reliability when we are dealing with a process which must be completed in a finite period of time. The results for the Australian process in comparison encourage some confidence that satisfactory reliability and validity can be achieved by keeping standards broad and small in number, allowing standards monitors the broadest discretion to follow whatever evidence they consider most strategic, and emphasising in training the need for team dialogue to focus on the wood as well as the trees.

American evaluators have been systematically blind to this possibility. When they fail to find reliability after innovations to "tighten up" the standards and protocols, they call for more of the same. They conclude that the tightening and refinement did not go far enough. Consider, for example, the evaluation of the state of New York's methodologically sophisticated and pace-setting Sentinel Health Events (Office of Health Systems Management, 1985). The Sentinel Health Events are not legal standards, but they are outcome measures designed to be at the heart of the innovative New York regulatory system. When this study obtained poor reliabilities for nursing home ratings using the Sentinel Health Events, the evaluators concluded:

It is important to note that although the Stage I and Stage II reliabilities were disappointing, it is expected that the old system in New York State of PaCS (the system to be implemented nationally in April 1986) would have even less reliability. This is because the new system in New York has far more structure than either the old system or PaCS (Office of Health Systems Management, 1985: 105).

The assumption that more structure is better was particularly obstinate in light of the reasons for unreliability that were diagnosed in the New York study. The first and "extremely prevalent" reason found was that "some surveyors (incorrectly) extended protocol requirements by noting a quality issue when no such quality issue is defined in the

protocol" (Office of Health Systems Management, 1985: 39). An illustration of an "incorrect" deviation from protocol arose where one surveyor who was supposed to assess improvement of a decubitus ulcer on the basis of "chart review" found inadequate care and deterioration by observing care being given:

The protocol states that only a chart review is necessary for this protocol, so the first cause for difference of opinion was a result of one surveyor doing more than he/she was instructed to do (Office of Health Systems Management, 1985: 36).

This clearly illustrates the pathology of punishing surveyors for looking beyond the trees specified in their protocols to the wood. The orthodoxy of science is to disapprove of the nurse who used her initiative to follow her suspicion by digging deeper, and to approve the nurse who reached the wrong conclusion because she followed the protocol. The orthodoxy of science is naive here. It is naive to believe that nurses, who are socialized to really care about the patients they encounter, who are socialized to use their initiative as professionals to get to the bottom of problems, can be turned into uncaring, mindless automatons who simply stick to the protocols.

Given that many nurses will be caring enough and have the initiative to follow the evidentiary trail toward conclusions of poor quality care, we think it best to design inspection systems which both assume and encourage this, rather than systems which attempt to control it. We think that when a resident is being seriously neglected in some way two different nurses, with free reign to follow whatever evidentiary trail they pick up, are more likely to both detect the neglect than are two nurses who we ask to be automatons by following a standard protocol. This is particularly so with the many idiosyncratic types of neglect that the designers of the protocol never foresaw. The advantage of wide procedural discretion over tight definition of protocols in generating valid ratings seem to us especially profound when we are considering team inspections. This is because when one team member fails to latch onto an evidentiary trail that will lead to a not met rating, the other team member may succeed in latching onto it. Or one may discover the missing link in an evidentiary trail opened up by the other. With our own reliability data, after all, the main source of disagreement was the single independent rater failing to pick up information that the team had detected (and vice versa to a lesser extent).

One way of defending the American standards is to say that the comparison with Australia is the result of a statistical artifact. When an American inspector finds a problem which should be cited, there are 499 ways he can cite it under the wrong standard (if there are 500 standards). When an Australian decides to give a not met rating, there are only 30

ways he can get it wrong (with 31 standards). This is an overstatement because clearly there is little risk of fire safety breeches being written out under a quality of food standard. Nevertheless, the basic point is right – more standards means more ways classification errors can occur.

To point this out, however, is not to erect a defence of the American standards. If inspectors give the wrong ratings because of the many standards under which they might write non-compliance, then this is a bad feature of the design of a system with too many standards. It is the design features of this system that cause the unreliability. American inspectors are indeed repeatedly confused about whether they should write out a particular violation under this standard or that. Moreover, the discretion that comes with a large number of standards means that inspectors who want to get tough with a nursing home can work hard at finding a number of standards that can be put out of compliance by each violation of concern. Inspectors who want to go soft with a nursing home can search for the least serious single standard under which the violation can be written. A profusion of standards therefore also increases unreliability by giving inspectors more degrees of freedom to work with in making choices between getting tough and being soft.

So we have the paradox of an Australian set of standards designed to satisfy the “common sense” of the industry and its professions being more reliable than a process driven by the demands for reliability of a United States industry obsessed with consistency, a consumer movement concerned with enforceability, centralist politicians and bureaucrats seeking to bring decentralized discretion under control, scientists concerned with procedural rigour and lawyers concerned with the doctrine of vagueness. This multi-pronged pursuit of regulatory reliability in the United States has delivered unreliability. The unsophisticated pursuit of face validity in Australia has, in contrast, delivered an impressive degree of reliability, at least within the two most populous states – New South Wales and Victoria. Moreover, we have not found any reasons why this impressive reliability cannot be duplicated within and between all states and we have found encouraging evidence of content and concurrent validity as well.

BIBLIOGRAPHY

- Aiken, Lewis R. (1985) 'Psychological Testing and Assessment', United States of America: Allyn and Bacon
- Anastasi, Anne (1982) 'Psychological Testing', New York: Collier Macmillan Publishers
- Braithwaite, John; Makkai, Toni; Braithwaite, Valerie; Gibson, Diane and Ermann, David (1990) 'The Contribution of the Standards Monitoring Process to the Quality of Nursing Home Life: A Preliminary Report', Canberra: Department of Community Services and Health
- Commonwealth/State Working Party on Nursing Home Standards (1987) *Living in a Nursing Home: Outcome Standards for Australian Nursing Homes*. Canberra: Australian Government Publishing Service
- Department of Community Services and Health (1987) *Nursing Home Standards Monitoring Guidelines*. Canberra: Department of Community Services and Health
- Diver, Colin S. (1989) 'Regulatory Precision', *Making Regulatory Policy*. Pittsburgh, USA: University of Pittsburgh Press. pp199-229
- Giles Report; Senate Select Committee on Private Hospitals and Nursing Homes (1985) *Private Nursing Homes in Australia: Their Conduct, Administration and Ownership*. Canberra: Australian Government Publishing Service
- Gustafson, David H. (Project Manager) (1977) *Monitoring Quality of Nursing Home Care: Final Report*, Madison, Wisconsin: Department of Health and Social Services
- Gustafson, David H.; Fiss, Charles J.; Fryback, Judy C.; Smelser, Peggy A.; Hiles and May E. (1990) 'Measuring the Quality of care in Nursing Homes: a Pilot Study in Wisconsin', *Public Health Reports*. 95 (4) pp336-343
- Gustafson, David H.; Peterson, Robert; Casper, Sandra; Macco, Ann; Van Koningsveld, Richard and Kopetsky, Edward (1982) *The Impact of the Wisconsin Quality Assurance Project: A Field Evaluation*, Madison: University of Wisconsin
- McDonald, Tracey T. A. and Bates, Philip W.; Department of Community Service and Health (1989) *Commonwealth Nursing Home Outcome Standards – A practical guide to duty of care*. Canberra: Australian Government Publishing Service

McLeay Report, the House of Representatives Standing Committee on Expenditure (October 1982) *In a Home or At Home: Accommodation and Home Care for the Aged*. Parliament of the Commonwealth of Australia

Office of Health Systems Management (1985) *Evaluation of the New York State Residential Health Care Facility Quality Assurance System*. New York: Office of Health Systems Management, New York State Department of Health

Spector, William D.; Takada, Adrianna H.; Durgovich, Margaret L.; Laliberte, Linda L. and Tucker, Richard (1987) 'PaCS Evaluation: Final Report', Baltimore: Health Care Financing Bureau

Zimmerman, David; Egan, J. Robert; Gustafson, David; Metcalf, Charles and Skidmore, Felicity (Sep, 1985) 'Evaluation of the three State Demonstrations in Nursing Home Quality Assurance Procedures', Baltimore: Health Care Financing Bureau