

# THE POLITICS OF LEGALISM: RULES VERSUS STANDARDS IN NURSING-HOME REGULATION

JOHN BRAITHWAITE AND VALERIE BRAITHWAITE

---

*Australian National University*

**P**RECISE RULES give more explicit guidance than vague standards. It would seem to follow that regulators will enforce precise rules more reliably than vague standards. This article demonstrates empirically that this is not necessarily so. It also induces from data on nursing home regulation the beginnings of a theory of a reliability paradox. We show how pursuit of reliability for a part of the law can increase the unreliability of a whole body of law. Reliability is used here with the standard scientific meaning of the extent to which measures give consistent results, as contrasted with validity, which means the extent to which measures assess the 'true' position. A 12-inch ruler that in truth is 13 inches long gives reliable (consistent) but invalid (untrue) measurement.

Our aim is to contribute to the rules-versus-standards debate in law, to regulatory policy analysis and to debates about reliability in science. A rule is taken to be a legal norm of the form, in circumstance X, do Y or not-Y. A standard, in contrast, enjoins the pursuit or achievement of a value, a goal or outcome, without specifying the action(s) required to do so. The divide is not a neat one, the world being full of rules about standards and standards about rules. Nevertheless, some bodies of law are more dominated by standards and others more by rules.

### THE RELIABILITY PARADOX OF REGULATORY INSPECTION

The sociolegal literature on rules versus standards is itself paradoxical. On the one hand, there is a literature claiming to show that American law is more standards-oriented, British law more oriented to rules. The most important contribution to this literature, Atiyah and Summers' (1987) *Form and Substance in Anglo-American Law*, makes the more general claim that American law is more substantive, British law more formal (see also Krotoszynski, 1990). On the other hand, the comparative literature on regulatory enforcement concludes that enforcement in other nations is less formal than in the United States (see generally, Bardach and Kagan, 1982; Braithwaite, 1987), including Britain (Day and Klein, 1987; Hawkins, 1984; Vogel, 1986), Sweden (Kelman, 1981), Australia (Grabosky and Braithwaite, 1986) and Japan (Badaracco, 1985; Haley, 1988; Upham, 1987; Vogel, 1979). Data collected for the present study of nursing-home regulation fit both stories in terms of the British-American comparison, in that British nursing-home regulation operates with more precise and formal rules than the more substantive, standard-like American nursing-home regulations, while British regulatory practice is more discretionary, more oriented to securing improvement in the quality of care and less interested than in the USA with collecting evidence for litigation concerning non-compliance with rules. While American law is more substantive and less formal than British law, British nursing-home regulatory practice is less formal than is American practice.

It may be that these opposite tendencies are indeed both true and have their origins in the same historical variable: the greater distrust of the state in America, compared to Europe and Asia (see Kagan, 1991; Vogel, 1986). Distrust of the legislature by the courts and by litigants energized by the more vigorous American separation of powers (an institutionalization of distrust) is one factor that may have resulted in the creation of a more substantive jurisprudence by American courts. Distrust of the executive by interest groups (litigants) and the courts has led them to pursue the limitation of administrative discretion, driving regulators to more formal regulatory enforcement. In other words, courts and interest groups in America operationalize institutionalized distrust of the state by going substantive with statutory interpretation and by driving the executive to go formal with application of the law. Courts and interest groups play the separation of powers to seize power from the legislature by judicial discretion in the interpretation of statutes, just as they seize power from the executive by limiting its discretion.

While this is a way of reconciling two literatures on the greater formality of British jurisprudence and the greater formality of American regulatory enforcement, in this article we deal with a different comparison. We find that both Australian nursing-home law and Australian nursing-home regulatory practice are less formal than in the USA, which may express a higher degree of trust among the legislature, the executive, the industry and advocacy groups – tripartite consensus building. This has meant in this arena (since 1987 reforms) that private interests rarely used courts or lobbied parliamentarians for more formal laws to achieve their objectives (Braithwaite, 1994).

The cross-national comparison illuminates themes from the American Critical Legal Studies (CLS) literature on rules versus standards, particularly the work of Duncan Kennedy (1976) and Mark Kelman (1987). Kennedy and Kelman see a reliance on standards as premised 'on the hope of moral dialogue and ultimate consensus, since the standards will remain contentless unless such moral dialogue succeeds in overcoming the skeptical sense that one person's conception of what is just is nothing more than either a whimsical taste or a rhetorical cover for a self-serving program' (Kelman, 1987: 62; see also Michelman, 1986). Our case study shows this hope being more or less realized in Australia at least for one little domain of regulation for a very short period of history.<sup>1</sup> The point that Kelman and Kennedy make is that rules do not depend on communities seeking consensus; standards do.<sup>2</sup> Kelman and Kennedy contend that the rule mentality is therefore a liberal individualist one, a contention disputed by Pierre Schlag (1985; see also Radin, 1989: 806 and n. 88; Rose, 1987: 606–10; Sullivan, 1992).

From another ideological perspective, we would put it that a rule-orientation in law conduces to liberal individualism, whereas a standard-orientation can both conduce to and depend on a degree of republican community. Rather, if it does not depend on a degree of republican community among those concerned with a particular issue, like nursing-home regulation, a standards-orientation will conduce to unchecked domination. Meidinger (1987) reports the existence of 'regulatory communities' even in the USA.

In the rules-versus-standards debate, CLS scholars pick up themes from American Legal Realists, who were fond of showing how the presumed precision of rules enabled imprecision. One reason they pointed to was that courts usually confront a choice of which of a number of precise yet contradictory rules they choose to invoke in a particular situation: 'each rule was in fact radically undercut by its fratricidal twin' (Kelman, 1987: 48; see Schlag, 1985: 409). Moreover, both realists and critical scholars indicate that often when the nominal rule is formally clear and consistent, the real operative rule for invoking it is opaque and inconsistent. The speed limit is 60, but the operating rules are that anything up to 65 or so will be ignored or let off with a warning (unless contempt is shown to the police officer or the officer wants to detain a person to check out compliance with some other rule). Hence the law and society injunction: study the law in action rather than the law in books. If your concern is consistency, you are bound to be misled unless you study consistency of the law in action. Our present study seeks some clarification of the rules-versus-standards debate by examining law in action. It shows empirically what CLS scholars assert when they say 'the text cannot define its context' (Schlag, 1985: 410):

If we concede that the sector of the social world cordoned off by directive can be affected by the external world, then the only way in which a directive can be certain is if it is sufficiently flexible to accommodate the effects of the external world.

This refutes the view that 'the choice between formulating or interpreting a legal directive as a rule or as a standard is a choice between . . . certainty or flexibility, uniformity or individualization' (Schlag, 1985: 399). Rather, there are contexts

where standards are both more certain and more flexible, more uniform and more individualized than rules. Nursing home regulation we find here to be one such context. It will be shown to be a context where plural dialogue framed by few standards (rather than many rules) is better.

Consistency of compliance assessments has come to assume pre-eminent importance in policy-making within regulatory bureaucracies. This is especially true of political debates about nursing-home reform and regulation in Western democracies. It is our contention that this central regulatory preoccupation with consistency is misplaced and that reliability is not the most important desideratum of a set of standards for inspecting nursing homes or any other industry. We propose a paradox in which reliability is more likely to be achieved when reliability is not the central objective of public policy. When we make other objectives our central concern – such as designing standards which best foster a regulatory dialogue about how a nursing home can improve quality of life outcomes – an indirect effect may be that better reliability of ratings is achieved. We advance this paradox by considering recent reforms in Australia and then comparing them to the state of American nursing-home inspection.

#### NURSING-HOME REFORM IN AUSTRALIA

Following a series of nursing-home scandals, consumer activism and two parliamentary enquiries (McLeay Report, 1982; Giles Report, 1985), the Australian government launched a package of nursing-home reforms in 1987. The first element of the package was a new set of 31 outcome-oriented standards negotiated through the active collaboration of federal and state governments, industry and professional, union and consumer groups. The consensus standards that attracted the assent of all these groups were so broad that they were attacked as 'motherhood statements' by many who were not immediately engaged in their negotiation. We agreed with these critics at the time. Moreover, while the standards were marketed as 'Outcome Standards for Australian Nursing Homes', many of them (listed in Table 1) did not look like outcomes to us. Since then, we learned that degree of outcome-orientation is primarily a matter of regulatory process design rather than standard-wording (Braithwaite et al., 1990: 135–41), a point to which we return.

Another fundamental concern we had about the new Australian standards was that they just could not be reliable. They were broad, subjective, lacking in detailed protocols,<sup>3</sup> and the process by which they were to be rated was absolutely unattentive to sampling issues. Moreover, the process was resident-centred when we knew that a high proportion of residents would be confused and unreliable informants. Compounding these problems was a shift toward the inclusion of many 'soft' social and resident rights standards, in contrast to the previous exclusive focus on 'harder' structural or health-care inputs, that could be checked with a ruler, a thermometer or by confirming a doctor's signature. It was inconceivable to us that standards concerned with the resident's right to 'privacy and dignity' or a 'homelike environment' could be rated reliably. We

certainly thought these standards were a 'good thing' because they would stimulate a dialogue that was sorely needed. But we harboured the deepest of doubts that they could be legally enforceable, doubts shared at the time also by the government's lawyers, industry's legal advisers, and the consumer movement.

In 1987, when we started an international comparative study of nursing-home regulation in Australia, the USA, Britain and Japan, we were rather embarrassed by the Australian standards. We would say to 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.' They in turn would look aghast at how broad, vague, undefined and unenforceable these quaint Antipodean standards were. Since that time, our research findings provided quite strong grounds for believing that the broad, unrefined Australian standards are not just more reliable than US standards but more reliable by a wide margin. Furthermore, observations of 59 nursing-home inspections in Australia and 44 in the USA between 1988 and 1993 suggest that the reason Australian ratings are more reliable is precisely because they are more (a) broad, (b) subjective, (c) undefined with regard to protocols, (d) resident-centred and (e) devoid of random sampling.

#### THE AUSTRALIAN RELIABILITY STUDY

Inter-rater reliability studies are extremely rare in the literature on regulatory inspectorates. In fact, the nursing-home inspection data from the USA and Australia that we discuss in this article are almost the only reliability data of which we are aware on any type of regulatory inspectorate in any country.<sup>4</sup> There are two reasons for the rarity of such research. First, government agencies are normally fearful of studies which might show that their judgements about compliance with the law are arbitrary and capricious. Such data might be used by defence lawyers to destroy the legal foundations of the regulatory regime. Second, such studies are difficult to do – expensive, logistically a nightmare, and intrusive for the organizations being inspected. Funding agencies, governments and regulated organizations are all therefore resistant to regulatory reliability research.

Elsewhere (Braithwaite et al., 1991:12) we discuss how these sources of resistance were overcome in this study. An essential step toward overcoming industry resistance was the need to avoid compounding the disruption of a government inspection with a reliability inspection immediately before or after. We did this by placing a single reliability rater in the nursing home at the same time as the government inspection team. The government inspection team of two or three inspectors spends an average of 6.5 hours inspecting the facility. Australian facilities tend to be much smaller than US facilities, the median number of beds being 38. Because the reliability raters had to do the job alone, they generally arrived before the team and left after them, occasionally having to come back for a second day.<sup>5</sup> The single reliability raters could compensate for the need to do the work of two and sometimes three others in five additional ways, beyond spending more time in the facility: (1) they did not need to stop to

explain problems to facility staff and management; (2) they did not need to take extensive notes for use in an official report; (3) they did not need to collect the evidence to document a finding in the event of enforcement action; (4) they did not need to stop to compare notes with other team members; and (5) they also had the advantage of being more senior, experienced nurses. We were able, therefore, to sell the study to nursing homes on the grounds that the only extra disruption they would confront was having one additional inspector in their home for the duration of the normal standards monitoring visit, and perhaps a little longer.

Methodologically, there are both advantages and disadvantages to this approach. A disadvantage is that guidelines had to be enforced to prevent the team and the independent rater from communicating with each other in any way about standards during the visit. This we believe was successfully negotiated (Braithwaite et al., 1991). An advantage is avoiding the problem of sequential visits whereby the nursing home has the opportunity of rectifying problems identified by the visit of the first team before the second team arrives.

### *SAMPLE*

Details on the selection of the two independent raters, the 30 inspection teams from New South Wales and Victoria, and the sampling frame for the nursing homes are provided elsewhere (Braithwaite et al., 1991:13–15). Only one nursing home refused to cooperate in a quota sample of 50 nursing homes with quotas for (a) number of beds in the home, (b) non-profit-for-profit status, (c) state and (d) composition of the inspection team. The sampling for the reliability study had to be quota sampling because the inspectors at the time were in the midst of working through the stratified random sample of homes that we had selected for our wider evaluation (Braithwaite et al., 1990). 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 the 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 was 23, the average number of 'met' ratings for homes in the reliability study was 18.

### *RESULTS*

Agreement between the independent rater and the government inspection team was measured at three points in time. After the team completed its visit, the team met (usually 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 their (blind) initial ratings. The percent of agreement between these initial (totally independent) ratings are

provided in the first column of Table 1. The standards are rated into three categories – ‘met’, ‘action required’ and ‘urgent action required’.<sup>6</sup> On Standard 1.1, 84 percent agreement means that for 84 percent of the nursing homes, the team and the independent rater gave the home exactly the same rating on this 3-point scale. Once both sides had been apprised of each other’s ratings, they were asked to discuss why they had reached different conclusions on certain standards. During this discussion, one side would sometimes persuade the other that they were wrong. On occasion, the combining of their information caused both sides to conclude that they had been wrong. This generated the ‘after conferring’ ratings, the measure of agreement in the second column of Table 1. 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 that might rebut 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 of Table 1. At each stage, the reasons for disagreement and changes of heart were recorded. Data on the reasons for disagreement on different standards are elaborated elsewhere (Braithwaite et al., 1991: 18–30). Broadly, disagreements based on the collection of different information were equal in number to disagreements based on different interpretations of the standards.

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 made only a minor difference – on some standards increasing agreement slightly, on others reducing it slightly. There are some surprising results in Table 1. For example, the ‘homelike environment’ standard (4.1), which no one we know would have predicted to be reliable, was rated quite consistently – with 88 percent blind agreement rising to 94 percent after conferring. Similarly, it was assumed that reliability was implausible on the ‘soft’ social and residents’ rights standards. Yet soft standards such as 5.1, ‘The dignity of residents is respected by nursing home staff,’ were rated with impressive reliability.

The first suspicion one should harbour about the exceptionally high agreement in Table 1 stems from an assumption that in a 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. For example, if the probability of getting ‘a met’ is 0.9, the likelihood of getting two mets on purely statistical grounds from two independent assessors is going to be  $0.9 \times 0.9$ , that is, 0.81. Conversely, if the probability were lower for a met, that is 0.5, the likelihood of two mets is much lower (0.25). Two considerations render this explanation for high reliabilities implausible. First, with the Australian standards, 12 of the 31 had proportions of mets under 50 percent. In the US system, the distribution tends to be more extreme (met ratings often over 90 percent). Second, with the Australian data, independent raters are not more likely to agree with teams on ‘met’ ratings than they are on ‘action required’ and ‘urgent action required’ ratings. It is simply not true that our high reliabilities

TABLE I  
Overall Agreement between the Team and the Independent Rater<sup>a</sup> and the  
Kappa Coefficient<sup>b</sup> (n = 50)

<i>Standard</i>	<i>Initially</i>		<i>After Conferring</i>		<i>After Negotiation</i>	
	(%)		(%)		(%)	
1.1 Residents are enabled to receive appropriate medical care by a medical practitioner of their choice when needed.	84	(0.76)	90	(0.85)	92	(0.88)
1.2 Residents are enabled and encouraged to make informed choices about their individual care plans.	90	(0.81)	92	(0.85)	90	(0.81)
1.3 All residents are as free from pain as possible.	90	(0.61)	94	(0.75)	94	(0.70)
1.4 All residents are adequately nourished and adequately hydrated.	90	(0.86)	92	(0.87)	94	(0.90)
1.5 Residents are enabled to maintain continence.	78	(0.56)	86	(0.73)	86	(0.73)
1.6 Residents are enabled to maintain and, if possible, improve their mobility and dexterity.	90	(0.81)	96	(0.92)	96	(0.92)
1.7 Residents have clean healthy skin consistent with their age and general health.	98	(0.91)	98	(0.91)	98	(0.90)
1.8 Residents are enabled to maintain oral and dental health.	96	(0.91)	96	(0.91)	96	(0.91)
1.9 Sensory losses are identified and corrected so that residents are able to communicate effectively.	84	(0.66)	86	(0.71)	88	(0.74)
2.1 Residents are enabled and encouraged to have visitors of their choice and to maintain personal contacts.	90	(0.77)	94	(0.86)	96	(0.90)
2.2 Residents are enabled and encouraged to maintain control of their financial affairs.	94	(0.86)	94	(0.86)	96	(0.90)
2.3 Residents have maximum freedom of movement within and from the nursing home, restricted only for safety reasons.	94	(0.85)	98	(0.95)	94	(0.75)
2.4 Provision is made for residents with different religious, personal and cultural customs.	94	(0.84)	98	(0.94)	96	(0.88)
2.5 Residents are enabled and encouraged to maintain their responsibilities and obligations as citizens.	90	(0.62)	94	(0.77)	98	(0.90)
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	(0.75)	92	(0.84)	92	(0.83)
3.2 Residents and their representatives are enabled to comment or complain about conditions in the nursing home.	84	(0.71)	94	(0.89)	90	(0.81)
4.1 Management of the nursing home is attempting to create and maintain a homelike environment.	88	(0.77)	94	(0.89)	94	(0.89)
4.2 The nursing home has policies which enable residents to feel secure in their accommodation.	86	(0.73)	90	(0.81)	92	(0.84)
5.1 The dignity of residents is respected by nursing home staff.	92	(0.88)	98	(0.97)	98	(0.97)
5.2 Private property is not taken, lent or given to other people without the owner's permission.	96	(0.94)	98	(0.94)	98	(0.97)
5.3 Residents are enabled to undertake personal activities, including bathing, toileting and dressing in private.	88	(0.82)	94	(0.91)	94	(0.91)



<i>Standard</i>	<i>Initially</i>		<i>After Conferring</i>		<i>After Negotiation</i>	
	(%)		(%)		(%)	
5.4 The nursing home is free from undue noise.	94	(0.87)	96	(0.91)	92	(0.82)
5.5 Information about residents is treated confidentially.	90	(0.80)	96	(0.92)	96	(0.92)
5.6 Nursing-home practices support the residents' right to die with dignity.	96	(0.86)	98	(0.94)	98	(0.93)
6.1 Residents are enabled to participate in a wide range of activities appropriate to their interests and capacities.	92	(0.84)	94	(0.88)	94	(0.88)
7.1 The resident's right to participate in activities which may involve a degree of risk is respected.	94	(0.85)	96	(0.90)	94	(0.82)
7.2 Nursing-home design, equipment and practices contribute to a safe environment for residents, staff and visitors.	82	(0.71)	92	(0.87)	90	(0.84)
7.3 Residents, visitors and staff are protected from infection and infestation.	92	(0.87)	98	(0.97)	96	(0.94)
7.4 Residents and staff are protected from the hazards of fire and natural disasters.	94	(0.91)	94	(0.91)	96	(0.94)
7.5 The security of buildings, contents and people within the nursing home is safeguarded.	98	(0.96)	98	(0.96)	98	(0.96)
7.6 Physical and other forms of restraint are used correctly and appropriately.	92	(0.87)	92	(0.87)	90	(0.83)

<sup>a</sup> 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 percent 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.

<sup>b</sup> The kappa coefficients are in parentheses next to the percent figures of overall agreements.

reflect ease of agreement when the standard is met compared with difficulty in reaching agreement when it is not met (agreement for all standards by rating categories is presented in Braithwaite et al., 1991: 19–22).

Based on factor analytic work, we elsewhere argue that it is psychometrically defensible to add scores on the 31 standards to obtain a total compliance score (Braithwaite et al., 1990, 1992), something that would probably not be defensible on American Medicaid survey results. Overall, the inter-rater reliability coefficient for the blind ratings of the total compliance score is 0.93, increasing to 0.96 after conferring and remaining at 0.96 after negotiation with the nursing home. These reliability coefficients show no major variation by raters, state, size of home, level of disability of residents or ownership status. While these data show impressive intrastate reliability within the two largest Australian states, they do not demonstrate interstate reliability. That is, while two inspectors trained in the same state give the same ratings, they may both give different ratings from an inspection in a different state. Indeed, our qualitative fieldwork inclines us to agree with critics who say there are serious problems of interstate reliability.

The weakness of this reliability study is that the reliability rater is an individual rather than another team. Our expectation, however, was that reliance on an individual as the reliability rater would reduce rather than increase reliability. We expected that the reliability rater would fail to pick up information that the team picked up by virtue of having at least one extra set of eyes and ears. We tried to

compensate for this by allowing the reliability rater to stay in the nursing home longer than the team did and by using reliability raters who, in our opinion, were more experienced and acute observers than the average team member, but we never expected that this would fully allow the reliability rater to overcome the information-gathering deficit. This interpretation of why reliance on a single rater would reduce rather than increase reliability was borne out by the results of the study. At the conferring and negotiation stages, there were more cases of the reliability rater agreeing that she had made an erroneous rating by missing vital information than there were cases of teams agreeing that they had made erroneous ratings by failing to pick up some vital piece of information.

In addition to this reliability work, we also report some encouraging validation studies on the standards (Braithwaite et al., 1991; Braithwaite et al., 1992). Of particular interest here is that we asked 410 directors of nursing<sup>7</sup> to give their own nursing homes ratings on the 31 standards soon after an inspection team had visited their nursing home. The average agreement of directors of nursing with team ratings across the 31 standards was 92 percent, the lowest being 84 percent on homelike environment ratings (4.1). Directors of nursing naturally gave themselves higher ratings than did the teams, but the correlation between their total compliance scores and scores given by teams was 0.88. The consensus of industry understanding on the meaning of these standards that permits such a result is based on 93 percent of directors of nursing in Australia having attended a course on the standards. This process also led to an extraordinarily high level of belief in the 'clarity', 'desirability' and 'practicality' of the standards by directors of nursing and proprietors (Braithwaite et al., 1990, 1991, 1992). The worst result for any of the standards on these three criteria was on Standard 2.2, 'Residents are enabled and encouraged to maintain control of their financial affairs,' for which 24 percent of directors of nursing had doubts about the standard's practicality.

#### AMERICAN RELIABILITY STUDIES

How then do these Australian reliability results compare with the results of American reliability work? 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). Based on just 11 criteria, it was a much simpler measure than that used by Wisconsin state inspectors. Five two-person teams of nursing-home professionals visited nine nursing homes, giving 45 data points for the calculation of reliabilities. An average reliability coefficient on this simple measure of quality of care of 0.78 was obtained. This measure was also validated against a global 0-100 assessment of the quality of care of each home ( $r = 0.76$ ), the global assessment itself having been found to be reliable (Gustafson, 1977). This pilot therefore demonstrated that two-person teams could rate the quality of care in nursing homes with reasonable reliability. Unfortunately, when the two measures of quality of care were correlated with the number of deficiencies cited by the last government inspectors to visit the nursing home, the correlations

were not statistically significant at the 0.05 level. This early study thus gives us the first clear clue to the direction in which we are led by our results: 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). So we get the same general picture as in the pilot – reliable rating of a simple quality of care measure combined with a poor relationship of this reliable measure with compliance ratings from government inspections.

Another particularly discouraging finding of the Wisconsin project is summarized in the final report:

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 state inspectors had cited three standards as out of compliance for the 20 homes in the study, while the federal 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 percent of these between the state and federal validation teams. For 92 percent 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, 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.' In fact, deficiencies rose sharply between the first set of

inspections and the second. Moreover, 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 (two 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 percent of deficiencies cited were cited by both teams. An impressive 84 percent of deficiencies detected by state teams were also detected by independent teams, though there was 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 percent to be corrected and the independent teams 71 percent. 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 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.

Only 25 percent of the regulations cited by the official team were also cited by the validation 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. The second team, however, found twice as many deficiencies as the first.<sup>8</sup> These studies highlight the reliability paradox. How can the US system which involves many more inspector-hours in the nursing home, larger multidisciplinary teams and more sophisticated protocols on more precisely specified standards produce such low reliability, while the Australian system produces such high reliability?

While all the studies we are comparing from both sides of the Pacific have considerable deficiencies, the tenor of the findings are so diametrically opposed that it is hard to make sense of them in terms of method error.<sup>9</sup> This is especially so since we believe that the very extensive fieldwork we have undertaken since 1987 (see Braithwaite et al., 1993: Appendix A, 'Data and Methods'), including observing nursing-home inspectors doing their work during 103 inspections in the two countries, makes sense of our counterintuitive findings. To this interpretive work we now turn.

### WHY ARE AUSTRALIAN STANDARDS MORE RELIABLE?

We advance five answers to the question of why Australian nursing-home standards are more reliable than US standards: It is because they are (1) broad; (2) undefined with regard to protocols; (3) subjective; (4) resident-centred; and (5) devoid of random sampling. Put even more provocatively, we will show why the very desiderata revered within the received scientific wisdom of American gerontology are responsible for the unreliability of nursing-home inspection. The tendency of the gerontological consultants who advise US governments on nursing-home inspection policies is to cast blame at the competence of nursing-home inspectors. Our experience is that American nursing-home inspectors seem of better-than-average competence compared to business regulatory inspectors from many fields that we have observed in Australia and other parts of the world. What our investigation calls into question is the competence of the scientific analysis that has been offered of the American process.

#### *1. THE EFFECT OF BROADNESS OF STANDARDS ON RELIABILITY*

The Australian standards in the words of one consumer advocate are 'wishy-washy and blunt'. Their breadth and vagueness certainly makes them appear an implausible regulatory instrument. Consequently, there is pressure under the surface in Australia for standards that look more scientifically and legally respectable. Before we succumb to such pressures, it is well to contemplate how those pressures have been played out in the USA during the past 25 years.

Historically, what has happened in the United States is that key political players in the nursing-home regulatory game came to be 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, home X screams about inconsistency. It complains to its industry association about the vagueness of the standard 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 more specific, but for different reasons. They are concerned that vague standards are unenforceable. Legislators have been 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 federal government. They are frustrated at the failure of the states to deliver federal 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 which are not tightly specified must be narrowed. 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.

By 1986 the logical conclusion to this process was reached to the point where there were over 500 federal standards ('Tag numbers'). Outcome-oriented reforms in 1986 reduced these to 357, but this was short-lived respite with the 1987 Omnibus Budget Reconciliation Act (OBRA) reforms<sup>10</sup> adding a large number of new and tighter standards. Moreover, in some states, the number of state standards exceeds the federal standards by a factor of two or three. In most states, federal and state standards are surveyed simultaneously. The upshot is that most of the people who inspect US nursing homes are checking compliance with over a thousand regulations – a stark contrast with the 31 Australian standards. How do they cope with such a daunting task? The answer is that they do not. Some of the standards are completely forgotten, not suppressed by any malevolent or captured political motive, just plain forgotten. Such standards are never cited in the states where they are forgotten. Then there are those that become familiar by some accident of enforcement history that gave prominence to a particular standard in a particular state. Referring to state regulations, one midwestern inspector said: 'We use 10 percent of them repeatedly. You get into the habit of citing the same ones. Even though possibly you could use others [for the same offence]. Most are never used.'

The professional background of the inspection-team members is one important criterion that selects which standards will be attended to. Administrator: 'If you've got a nurse, it will be nursing deficiencies in the survey report; if a pharmacist, you'll get pharmacy deficiencies; a sanitarian, sanitary deficiencies; a lawyer, patient rights, etc.' We observed a Medicaid survey team to rate all dietary standards in a facility 'met' for the reason that the dietary problems looked so serious. How can that make sense? The team felt that the dietary problems were so bad that they could not check them all out properly and get all the other standards checked in the time-frame required by the Health Care Financing Administration (HCFA) to complete the survey; so they 'deferred' the dietary problem for two weeks until a dietician could be booked for a specialist survey. When we went out with the dietician two weeks later on this survey, sure enough, a great number of the dietary standards rated met in the survey report submitted to the federal government were rated 'not met'.

The point of these examples is that when surveyors have an impossible number of standards to check, arbitrary factors will cause particular standards to be checked in some homes but neglected in others, causing endemic unreliability. 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 physician's orders, resident choice, mobility and freedom of movement to all be out of compliance).<sup>11</sup> 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.

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 has 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 to 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. It would be difficult to do so since the team has neither debated compliance with them nor assured themselves that they have collected the data relevant to them. The crucial difference is 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 with which they must live makes this quite impossible. The end result of demands for more specific standards with more clearly defined protocols that cover all 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.

If American inspectors give up on keeping all these standards in their heads, what is their cognitive coping strategy? It seems to us that they have a gestalt of the prohibitions codified in the regulations – for example, that good infection

control is required; that privacy must be protected; that good nursing practice should be followed. It is likely that professional training informs these gestalts more than the law does. They then decide whether a citation ought to be written by deciding whether it offends against one of these gestalts. Then they search for the appropriate regulation under which to cite it. 'What will we call it? How about 1220 A? What about 1220 B? Why don't we use both of them?'

After explaining to a number of surveyors this interpretation, based on our observation of how they coped, they agreed that this was basically how they did it. When we pointed out that the most troubling implication of this process from the point of view of reliability was that depending on how hard they searched through the standards, they might find one or two or three deficiencies to write out, one of them said, tellingly: 'Or they might find none at all and have to muck it in.' Decisions about how hard to search for multiple citations for essentially the same problem are driven by a 'professional judgment' of 'how serious overall their problems have been' or 'how hard they've been trying.' 'You can write it out under [X] and create a repeat violation because they got a deficiency on [X] last time. Or you can write it out under [Y] so its just an element, which has no real consequences. Or you can put it out under both [X] and [Y], putting out a whole standard.'<sup>12</sup>

Hence, hand-in-hand with a paradox of reliability is a paradox of discretion. More and more specific standards are written by lawmakers in the misplaced belief that this narrows the discretion of inspectors. The opposite is the truth: the larger the smorgasbord of standards, the greater the discretion of regulators to pick and choose an enforcement cocktail tailored to meet their own objective. A proliferation of more specific laws is a resource to expand discretion, not a limitation upon it (Baldwin and Hawkins, 1984).

The beauty of a small number of broad standards is therefore that one can design a regulatory process to ensure that the ticking of a met rating means that a proper process of information-gathering and team deliberation has occurred on that standard. One accountability check in Australia is that whenever enforcement action is appealed, the team's worksheets listing all of the positives and negatives they found under each of the 31 standards must be tabled before the Standards Review Panel. Until 1990, teams were required to write a report for the nursing home with a statement in support of the rating for each of the 31 standards. This proved an impossible burden with reports often running over 50 typed pages. As a result, from 1990, the report gives a statement in support of the compliance ratings of standards grouped under seven objectives. This makes for 10-page reports which are more consumer-friendly.

In summary, the smaller the number of standards, the better the prospects of ensuring that (a) the most vital information for assessing the total quality of life and quality of care of residents is pursued; (b) lying behind each rating is a collective deliberative process on what that particular rating should be; (c) there is effective public accountability to audit that (a) and (b) actually occur; and (d) inspectors have the capacity to stand back to document the wider patterns in the problems they have identified, to see the wood for the trees. These indeed are just some of the ways that 'loose laws' can make for better public policy (Goodin, 1982: 59-72).



## 2. THE EFFECT OF PROTOCOLS ON RELIABILITY

The same argument against the proliferation of standards can be extended to the proliferation of protocols for rating standards. The misplaced faith of the legislator for narrowing broad discretion results in the enactment of more specific laws. The misplaced faith of the consultants from the scientific establishment of gerontology results in protocols to narrow discretion. When there are the number of protocols that are supposedly followed in the American process, all the inspector can do is fill out the forms mandated for certain protocols and essentially fudge the other protocols that cannot be checked. Realizing 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.

The commendable shift to resident interviews in the OBRA reforms to the survey process introduced in October 1990 has already fallen prey to the disease of the proliferation of protocols. Our observations during the early months of the new process illustrated the unintended consequences. Interviewers started at the beginning of the schedule for the resident interviews only to find that either they ran out of time or the resident became exhausted before they had got very far through the schedule. Thus, items placed early in the schedule were done according to protocol, and later items were fudged or ignored. For this and other reasons we will come to soon, we therefore agree with the exasperated surveyor who said to us: 'Our own questions are better than the nonsense on the OBRA forms.' When we raised this and other examples of OBRA protocols being selectively and partially followed, one state survey manager replied: 'We'll streamline it. In time we'll do it our way rather than follow the HCFA protocol.' The trouble is of course that every state has no choice but to streamline, and each state streamlines in its own way. Streamlining error is the inevitable consequence of overly ambitious pursuit of reliability through the proliferation of protocols.

Protocols can work well in the context of a social science evaluation, but fail in practice because in the evaluation study the protocol does not have to compete for limited time with 30 other protocols. There are, however, other reasons why a protocol that succeeds in the evaluation study fails in inspection practice. An evaluation might show that a protocol of putting a tick in a box for the name of every resident that participates in each activity can be done reliably. Moreover, scores from following the protocol are validated against more sophisticated detailed assessments of the effectiveness of activities programmes. Unfortunately, however, what was valid at the evaluation stage quickly becomes invalid at the implementation stage.

Administrators are quick learners in the business of getting good survey results. If ticks in activities boxes are what count, droves of sleeping residents will be wheeled into activities programmes to get the numbers up. Never mind that the quality of the activities programme will be compromised by the clutter of

sleeping residents; it's beating the protocol that counts. This is why nursing-home administrators love protocols: 'Give us the rules and we'll play the game.' Imprecision, undefined evidence-gathering procedures, make it harder for the efficient administrator to beat the system. In Australia, because there are no defined protocols for inspectors, nursing-home management has no choice but to focus on the outcomes for which the inspectors are searching. This makes their life more painful and uncertain. When protocols are defined, the administrators showed us how they create a documentation system, a paper trail that matches the protocols the inspectors follow: 'You can achieve paper compliance without real compliance. You can fool most inspectors on most standards with paper compliance.'

The source of unreliability then becomes the rare inspector who looks behind the paper trail to the quality of care that is actually being given. Validity then becomes the major source of unreliability!

American evaluators have been systematically blind to these possibilities. 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 were not legal standards but 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 when 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/his initiative to follow her/his suspicion by

digging deeper and to approve the nurse who reached the wrong conclusion because she/he followed the protocol. The orthodoxy of science is naïve here. It is naïve to believe that nurses, who are socialized to care about the patients they encounter, who are trained 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 that attempt to control it. We think that when a resident is being seriously neglected, two different nurses, with free rein to follow whatever evidentiary trail they pick up, are both more likely to 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. As one Australian inspector pointed out: 'There are a hundred different reasons for residents to be incontinent.' 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 on to an evidentiary trail that will lead to a not met rating the other team member may succeed in latching on to it, or one may discover the missing link in an evidentiary chain pieced together 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).

Inspectors boxed in by a proliferation of protocols cope in another way that makes it difficult for them to see the wood for the trees: *task-specialization*. One surveyor takes responsibility for filling out the forms required from record interviews and another completes the reviews of resident records. We observed very little reading by one surveyor of the protocols filled out by another. In the busy work of getting the huge number of survey forms completed, the process of following up problems identified in a resident interview by tracking down residents' records (and vice versa) is profoundly compromised. This is not to say such follow-through does not occur in the USA; it is just to say that it occurs more freely in the more free-wheeling Australian process. Protocols kill initiative under a pile of paper. With nursing-home staff and nursing-home inspectors alike, excessive demands for task-orientation distract attention from the outcomes that matter. The result is the creation of nursing-home bureaucracies and regulatory bureaucracies that miss the big picture.

This pathology of protocols is just a specific illustration of the more general problem of formalized regulation forgetting that 'policy problems can be solved only by taking account of numerous interdependent and highly variable factors which oblige decision-makers to manage a kind of cybernetic process involving tentative probe, feedback, adjustment, and reconciliation' (Schuck, 1979:29). The pursuit of precision, either by protocols or by the proliferation of ever-narrower rules, causes an unreliability that is a symptom of a deeper and many-sided malaise of regulatory failure. This is especially depressing since the

pursuit of precision usually fails in its own terms – it fails to deliver precision. There might be 30 or 40 US regulations for every one in Australia, but the American standards still seem vague. In the language game of regulation, the problem of one vague concept is solved by splitting into three vague concepts or by defining protocols with other vague concepts. An alternative we suggest to the perpetual struggle to get the words right is to concentrate more on getting the processes of dialogue right. Certainly there is merit in keeping the words simple. This is a necessary precondition to accomplishing processes of dialogue that will deliver reliable judgements on those simple words.

### *3. THE EFFECT OF SUBJECTIVITY OF STANDARDS ON RELIABILITY*

When we spoke to senior regulatory bureaucrats in the United States and to social scientists who had been involved in the development and evaluation of nursing home surveys, a common type of comment was: 'There are some things that the process cannot do reliably. So you don't do them. Examples are: "Are the staff pleasant? Is the room tastefully decorated?"' The thought occurred to us that if the Hyatt Hotel group adopted the view that decor and staff pleasantness were matters for which it could not set reliable standards (and therefore should not bother with), it would soon be bankrupt. In business, a head office effectively enforces all manner of 'soft' standards on franchisees by adopting a qualitative approach to evaluation of performance. In these cases, dialogue informs an evaluation that is made against the yardstick of 'What is it, subjectively, that consumers want?' Admittedly, some of these subjective assessments are easy and some are hard. You don't have to talk to many consumers to realize that they don't like vermin running around their hotel room or their nursing home. But to judge reactions as to how warm and non-institutional is the decor or what they think of their continental breakfast, you really need to work hard at talking to consumers. Surely one reason that American nursing homes are so cold, institutional, and unattentive to decor compared for example to English nursing homes is precisely the attitude that such things are so subjective as to be beyond control.

The reliability of the 'homelike environment' standard (4.1 in Table 1) in Australia shows that this American posture is in error. A properly subjective approach on a standard such as this involves talking to residents about whether they feel free to put up personal mementos in an area they define as their private space, whether there are spaces in the facility that they feel are inviting and homelike for chatting with friends, whether they feel there are inviting garden areas they can use. This subjectivity often comes under attack in Australia. For example, managers of nursing-home chains complain to us that they have provided exactly the same food to two nursing homes; the team in one home gives them a 'met' rating for the food and in the other home they get an 'action required' rating. There is absolutely no inconsistency here if the residents at the two homes have different subjective views about the food. Two teams will never agree on what is nice food, but we have found that they can agree, with high

reliability, on whether the residents in a nursing home generally like the food they are getting. Reliability is accomplished by rejecting objectivity in favour of subjectivity.

The impetus to reform subjectivity in standards through objective criteria and protocols is dangerous because quality of life, which is what nursing-home care should be about, is ultimately an irreducibly subjective matter. The paradox of objectivity is that its pursuit undercuts a desideratum on which the industry, politicians, consumer groups and gerontologists (if not the lawyers) generally agree. This is that the regulatory process should be more outcome-oriented. The trouble is that inputs (the temperature of the food as it leaves the serving line; the size of the room) are generally more 'objective' than outcomes (satisfaction of residents with the food and the comfort of the room). Objectivity disempowers residents and empowers nursing-home managements who know how to get objective inputs in a row for inspection day – reams of documentation of the temperatures on food lines. Subjectivity, in contrast, means that residents are empowered because it is no longer the documents under the control of management that matter; it is what they as residents think and want that counts.

Even the vision of outcomes which enjoys most support within the American gerontological establishment is an 'objective' conception – counting the number of residents with decubitus ulcers or the number of restrained residents (Institute of Medicine, 1986; Office of Health Systems Management, 1985; Phillips, 1987). Collecting such objective outcome information is something we applaud. However, it must be pointed out that it is a process that does little to shift power over the definition of regulatory problems out of the hands of management into the hands of consumers. Administrators can handle a regulatory process that counts decubitus ulcers or restraints. They can keep control of their own evaluation because they know what the score is objectively before the inspector walks through the door. Consequently, they are ready with a defensive documentary record to prove that the residents with the bedsores were all turned two-hourly, that there are physicians' orders, psychiatric assessments to justify the restraints they want to keep and the like (Wiener and Kayser-Jones, 1989). Hence, while the outcome of the number of restrained residents can be measured very reliably and while this is an extremely valuable thing to do, it does not solve the problem of reliably assessing a law that requires proper use of restraints (such as Standard 7.6 in Table 1). Reliable assessment of a legal standard requires investigative common sense, determination and the imagination to uncover leads and follow them. The protocol-following automatons lauded by the objective-outcomes movement will be incapable of doing this job reliably. Their protocols would not allow them reliably to find that over 90 percent of the American nursing homes we have seen fail to meet the Australian restraint Standard 7.6. Instead, they mostly conclude that American homes where half the residents are tied up or chemically restrained meet the US standards, and where they find non-compliance, they find it unreliably.

*4. THE EFFECT OF BEING RESIDENT-CENTRED ON RELIABILITY*

Being resident-centred means two things for us: first, it means relying on residents as a source of information for rating standards and, second, it means participation in a regulatory dialogue where quality of life outcomes for residents are the ultimate criteria of regulatory evaluation. Critics regard this as an orientation that is a prescription for unreliability because most residents are so sick or confused that what is a subjectively good outcome for them is unknowable in most particular cases. Moreover, for the same reason, they are incapable of being meaningfully interviewed. Our research team deals with both these objections elsewhere at greater length (Braithwaite and Makkai, 1993). Briefly, we should at least say here that most resident outcomes that are the focus of debate within any sensible regulatory system will be uncontroversially bad. We know that getting burnt in a fire, getting pills prescribed for someone else, or getting a decubitus ulcer are outcomes that residents are keen to avoid without having to ask them. Second, we use our fieldwork data to argue elsewhere that skilled inspectors know how to find those residents in a nursing home who will be outstanding informants on those issues that do require subjective feedback from residents and they also know how to get some useful information even from residents who spend most of their life extremely confused (Braithwaite and Makkai, 1993). The critics argue that it is harder, or even impossible, for a nursing home with many extremely high disability or demented residents to comply with standards under the more resident-centred process that we have in Australia. Our data do not show this to be the case (Braithwaite and Makkai, 1993).

In practical terms, the Australian nursing-home regulatory process is more resident-centred than any we know. Yet we have seen that this resident-centred process seems to have high reliability. It is true, as the critics point out, that Australian inspectors are often misled by confused residents. However, we also found it true that these errors are almost invariably corrected long before they affect final ratings. In our reliability study, inspectors being misled by confused residents did not even register as a reason for disagreements, though one side picking up useful information from residents that the other side missed was one of the more important reasons for disagreements (Braithwaite et al., 1991). Moreover, with our study of reasons for 889 disagreements between inspectors and directors of nursing on ratings, in only 3 percent of disagreements was one of the reasons given for disagreement that the director of nursing felt that the team had relied on misinformation from residents (Braithwaite et al., 1990: 73).

These data show that interpretive errors in a subjective, resident-centred process can be and are corrected through a process of dialogue. First, dialogue with residents and their carers is important. Second, with nursing-home staff and within the team, there is dialogue about whether the seven quality of life objectives of the Australian standards are being secured: health care, social independence, freedom of choice, homelike environment, privacy and dignity, variety of experience and safety. Consistency does not easily fall out of such processes of dialogue; it comes painfully and with a lot of backtracking and moving in circles as new inconsistencies are discovered along the way.

Ultimately, however, consistency will be greater to the extent that the *only* debate is about resident outcomes. When the debate is theoretically only about whether an input required in a rule is delivered, in practice the outcomes that motivated the rulemakers' specification of the input will unpredictably intrude into regulatory judgements. This is inevitable because sensible people do not like to enforce the law when its enforcement will defeat the very purposes for which the law was enacted. Because business regulatory laws (such as those that regulate health care) deal with such complex, changing and individually variable problems, mismatch between legally mandated input and desired outcome is exceedingly common.

Let us illustrate with a comparatively simple example. We observed a Chicago sanitarian point out during an exit conference following an inspection that it is against the regulations to have a male and a female in adjoining rooms sharing the same toilet. The sanitarian concedes that in this particular case neither resident is capable of using the toilet and that moving either of them would be upsetting to them. He says that he is going to turn a blind eye to the rule for the sake of the residents, but he warns management that someone else from the department could come along and cite them for this. In other words, he is pointing out that because there is such a mismatch between rule and outcome, he is giving an unreliable ruling. With Australian inspectors confronting such a predicament, there will be no such unreliability. Since what is the best outcome for the residents is clear and since inspectors are instructed only to be concerned about outcomes, dialogue should quickly lead to a reliable result.

Our claim is that dialogue about resident-centred outcomes conduces to more reliability than recourse to authoritative interpretations of the meaning of words in rules. A word like privacy is certainly a very slippery word, as is health or pain as a matter of fact. In a resident-centred process when the question arises 'But is this really an invasion of privacy?', the answer is discovered through a process of dialogue about what are the senses of privacy that are important to this particular resident. Dissension is more likely when the question is to be resolved by pitting one inspector's conception of what privacy means against another's; consensus is more likely when the professional responsibility of both is to focus on the practical sense of privacy that is subjectively important to that resident in that situation. There will always be inconsistency in trans-situational 'objective' judgements of whether privacy has been invaded. Resident-centred contextual dialogue about privacy outcomes, in contrast, can often reach reliable conclusions.

It follows that progress with increasing reliability is less likely to come from handing down more sharply defined authoritative interpretations of what privacy is, more likely to come from improving processes of dialogue. Dialogue occurs at many levels, all of which allow scope for improvement. Inspectors can improve their dialogue with residents by learning how to deal with resident intimidation, how to capture the moments of clarity of thought that normally confused residents experience, how to communicate non-verbally with residents when verbal communication is poor, how to use third parties (roommates, relatives) to draw out uncommunicative residents, how to mobilize translation

support with non-English speaking residents. Moreover, group discussions with residents (e.g. with Residents' Councils), something on which American inspectors are much more advanced than those in Australia, can draw out some people who will not be drawn one on one.

Inspectors can improve their dialogue with each other by scheduling interim discussions during the course of an inspection, learning how to be active listeners, learning how to break deadlocks by framing the sticking points on which more subjective information from residents is needed. They can also learn when it is wise to draw on the wider experience of a supervisor or to get the perspective of nursing-home staff on an issue. They can learn how to select crucial conflicts over consistency to be put on the agenda for regular meetings of all inspectors. Training courses can be improved by making them more genuinely dialogic – showing videotapes of real regulatory encounters and asking trainees to debate the appropriate compliance rating, for example. Attempts have been made to improve dialogue at most of these levels in both the USA and Australia, but with highly variable commitment and success.

Focusing reform energy on processes of dialogue rather than on rules recognizes something that the community of scholars who work on regulation and policing have begun to realize. This is that it is simply not true that police officers make decisions mostly by reference to rules (Shearing and Ericson, 1991). They do not, should not and could not do so. Police culture, Shearing and Ericson (1991) point out, is not a book of rules, but a storybook. Police learn how to handle difficult situations by hearing stories about how competent officers handled similar situations or by themselves experiencing and retelling such stories:

Stories constitute a consciousness, a sensibility, a way of being out of which action will flow without recourse to specific instructions. Unlike rules, stories do not address action directly but rather constitute a sensibility out of which action flows. (Shearing, personal communication, 1993)

Stories instruct the participants in a regulatory culture how to 'read', via a 'poetic apprehension', the layers of meaning in a situation. Shearing and Ericson (1991) show how this poetic apprehension is communicated through analogous reasoning – like advising young officers to avoid provocation in difficult situations by 'acting as if you were on holidays'. Nursing-home inspectors communicate a resident-centred sensibility, for example, with the analogous reason: 'Is this a home that you could be happy for your mother or grandmother to live in?' Reliable ratings will be maximally possible with a regulatory culture that accomplishes a common set of sensibilities through processes of dialogue.

Hence, a hotel chain can get staff pleasantness and decor to a state that appeals to consumers, but it will not accomplish this with a set of decor rules. Rather, it seeks to cultivate the right sort of sensibilities in its management and quality assurance staff with stories, concrete examples and analogies. Staff civility and pleasant decor then follow from these sensibilities.

The importance of legal standards is more in setting the framework and focus for storytelling, less as words that utter explicit guidance. To be good at



framework-setting and focusing dialogue, standards must be simple and few in number. Like good poetry, they must engage us by being replete with silences, leaving us to make of them what we can: 'For in leaving to us the talk of making sense of what is before us, this silence forces our continuous and attentive engagement with the poem itself' (White, 1984: 27).

##### 5. THE EFFECT OF RANDOM SAMPLING ON RELIABILITY

Yet another way in which the paradox of reliability came about in the past was 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 key players in the industry association 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 USA 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 their findings by questioning their competence in the statistical theory which would warrant the judgement 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/he was not randomly selected. On one occasion, an inspector from another part of the state asked the team to put a friend of hers who was suffering from a decubitus ulcer in the sample. On another occasion, a complainant was fudged into the sample to protect her – so that the problem would appear to have been discovered by the team. In another multistorey nursing home, where care seemed to vary by floor, the team decided to 'improve' on the standard sampling protocol by stratifying the sample by floor. These were all examples of principled cheating.

Examples of unprincipled cheating included the following. The team member met on the tour a resident who was a friend she enjoyed talking to. After a 20-minute chat with her over lunch, she realized that she had already collected half the information she needed from this resident. So she slipped her into the random sample. In another type of fudging repeatedly observed, the inspector would find 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 problem, this resident became 'a good one to do'. Slipping such a resident into the sample reduced the total number of residents investigated. We say these latter examples are of unprincipled cheating, but the teams did not view it this way. We have already made the point that systematic data collection to rate hundreds of regulations is impossible; the cheating, they contended, made an impossible job a little more possible.

Even when the cheating was clearly principled rather than designed to cut corners, surveyors were under no illusion that it was cheating that required concealment:

*Surveyor*: 'There are ways of bending these things [the sample]. That doesn't cause us any problem.'

*JB*: 'How do you mean?'

*Surveyor*: 'Well you can just number the list of patients where you are selecting every fifth one: 1,2,3,4,6,5.'

It is a sad commentary on the unreflexive empiricism of the behavioural sciences that so many books are written on the statistics of sampling, while no one does empirical studies of random sampling in practice. Our own observations are of a wide gulf between science in the books and science in action, even with surveyors with considerably more education and training in sampling protocols than the average opinion survey interviewer, for example.

The reasons for the gulf in this domain are multiple, but include: (a) laziness; (b) job survival; (c) the view that they have more serious professional obligations than to the numbers games of scientists; and (d) the view that they have a more sophisticated or rounded practitioner's view of randomness than the theoreticians. The last of these is the most interesting: surveyors who stratify by floor or who put into the sample someone they bump into on the tour have a social construction of randomness which they will defend as superior to the protocol they have been given. This then is just a special case of the naïvety of assuming that because protocols exist, they will be followed; because something works in a pilot, it will work in day-to-day practice. Behavioural scientists are empirically neglectful of behaviour in science. One lesson of observing nursing-home inspections is that trained professionals expect and extract working conditions where they exercise professional judgement: they simply refuse to succumb to demands that they follow instructions like machines. Inspection procedures should never be based on the hope that this will not happen; they should be designed on the expectation that it will.

Hence, random sampling in nursing-home inspection in the USA up to 1990 tended to fail in one of two ways. In some cases the team cheated by slipping bad cases into the sample, thereby defeating randomness. In other cases, they refrained from cheating when they saw bad cases; they settled for the randomly selected case and let the bad case slip by that might have been their best chance of getting to the deepest problems in the facility. As argued earlier, our own view is that inspectors are most likely to find problems of non-compliance reliably when their initiative in following evidentiary leads is cultivated instead of controlled.

The final reason why random sampling reduces reliability in regulatory inspections is that it is extraordinarily time-consuming. It distracts a great deal of time from the more important work of gathering evidence on all the standards (or in the US case, at least a larger fraction of the standards). We would observe one team member on the first day of a pre-1990 US inspection do little more than participate in the initial tour and gather all the information on residents and their categories of care in order to select the sample, selecting it with the correct number in each category (variably according to the number of qualified residents in the home), and recording the selected sample to prove that the sampling protocol had been followed. On one occasion, we observed four nurses debate for 37 minutes whether, for sampling purposes, group therapy counted in the 'physical therapy' category. A call to the supervisor was eventually needed to resolve the dispute. All this effort for the dubious statistical virtue of randomly selecting 16 residents from a population of 80! The USA made a sound scientific decision when it abandoned random sampling in October 1990.

### DISCUSSION

The first point we should make toward a conclusion is that our hypothesis that the US nursing-home inspection process is much less reliable than the Australian process may simply not be true. None of the quantitative studies has large samples, and all have design flaws. They do, nevertheless, amount to superior information on the reliability of nursing-home inspection than we have on any other area of business regulatory inspection we know. We would be surprised if our hypothesis were wrong, however, not only because of the dramatically different results of the quantitative reliability studies but because of the convergent conclusion from our extensive qualitative fieldwork.

It could be argued that even if our empirical claim about the comparative reliability of nursing-home inspection were right, this is simply a statistical artefact. When an American inspector finds a problem that should be cited, there are 499 different ways he/she 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/she can get it wrong (with 31 standards). Of course this is an overstatement because clearly there is little risk of fire-safety citations being written under a quality of food standard. Nevertheless, the basic point remains that more standards means more ways classification errors can occur.

To point this out, however, is not to erect a defence of the American standards. It is no comfort to nursing-home operators who feel they have been treated inconsistently to tell them that they unfortunately have been victims of a statistical artefact. 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. At the outset to this article, we said that reliability is not the only, or even the most important, criterion for evaluating a regulatory process. For one thing, reliability is not validity. A common assertion

about nursing homes from industry participants in all the countries we visited for this research is that competent people can assess validity rather easily: 'When I go into a home, I can look around and fairly quickly tell if it is a well-run nursing home or a nursing home with problems. If I can do it, they [surveyors] can because they go to different homes all the time' (Pennsylvanian administrator). This kind of assertion does not sit well with American researchers who find low reliability in the ratings of US inspection teams. 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 and life, and it is at this broad level that reliability can be achieved, as in this oft-repeated wisdom of the industry.

Once the team follows protocols that cause it to dig deeper, reliability may become problematic. 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, interteam reliability would have been shattered, but this team would be making a more valid rating. What we may have with Australian standards monitoring is a process that reliably reveals those sorts of problems that can be revealed by shallow digging.

Overall, we doubt that the American process digs deeper than the Australian process does, because Australian facilities that are identified as 'homes of concern' at the initial visit usually get many more follow-up visits than do American homes with comparable problems. However, there is no doubt that at the initial visit stage, there are several important ways in which the American process digs deeper. An example is the US requirement for a systematic survey of errors in the administration of medications. US inspections often uncover quite frightening 'med-pass error rates' of 10 percent and more. Australian standards monitoring visits reliably fail to uncover such problems. American inspectors observe treatments being given to a sample of residents. Mostly these are observations of the treatment of decubitus ulcers. This deeper digging in the American process uncovers many problems that remain reliably submerged in the Australian process – Class III ulcers that are documented and treated as Class II or poor infection-control practice in the treatment of bedsores. The biggest difference is with the deeper digging that American inspectors do in checking the documentation of resident care, though, unlike the other domains of deeper digging we have discussed, we doubt that this is reason for superiority in the US

process (Braithwaite, 1994). Moreover, we do not believe that the lower reliability of the US process is caused by its successes in collecting greater amounts of information than the initial Australian visit. On the contrary, the greater unreliability of the American process is caused by the fact that most of its hundreds of ratings are made on the basis of no information or at least no deliberation of any information.

In summary, we think it is fair to describe the initial Australian visit as reliably uncovering problems that can be exposed by shallow digging and reliably failing to uncover problems that can only be exposed by digging deeper. Moreover, it would improve the validity of the Australian process to adopt certain elements of the American process, such as the systematic observation of treatments. This is not to deny our more fundamental assertion that the best way to find the deepest problems is to follow leads rather than follow protocols. A police department that relies exclusively on the most sophisticated regimen of random patrol will not solve murders.

### CONCLUSION

The classic work on the optimal precision of regulatory standards has been that of Colin Diver (1980, 1989). Diver identifies three problems with regulatory standards – vagueness, overinclusiveness and complexity. Vague rules leave citizens to guess at their meaning in particular circumstances. Overinclusive rules command actions that are not beneficial (or are harmful). Overinclusive rules accomplish a poor fit between outcomes desired by policymakers and the requirements demanded by literal adherence to the rules. A complex set of rules is long, full of contradictions between one part of the rules and another, and dogged by verbal intricacy.

Diver (1989) advances three qualities of well-drafted legal rules – transparency, congruence and simplicity – to deal respectively with the three problems of vagueness, overinclusiveness and complexity. Clearly, the 31 Australian standards in Table 1 satisfy the criterion of simplicity. Congruence between desired outcomes and legal requirements, we have seen, is accomplished by shifting to a radical outcome-orientation wherein if resident outcomes cannot be shown to be put at risk by an action, then that action involves no violation. The problem of vagueness, however, is not dealt with by making the rules transparent. Different citizens cannot look at the Australian standards as if through a transparent window and see the same image. Transparency is accomplished at the next stage of the process. Vagueness about meaning is clarified by empowering the beneficiaries of the rules to define the meanings of the standard that are important to them in a particular situation. Because there is no way of solving the problem of vagueness at the level of the wording of rules without also rendering the rules overinclusive and complex,<sup>13</sup> the solution is to leave the words vague but to specify the interpretive evidence that is privileged and to require a regulatory dialogue about this evidence. Laws that fail the transparency test as disembodied law can be the starting point for a process that solves the vagueness problem

through mandating which subjectivity is privileged (and mandating dialogue focused on that subjectivity). The law fails the vagueness test, but the package of law–outcome–subjectivity–dialogue passes it. That is, two people just applying the law are likely to reach different conclusions about the same evidence; two people who use the law to guide outcome-oriented dialogue are likely to reach the same legal conclusion in a specific context. Dialogic accountability within the process, especially when this includes a requirement for consultation with residents and staff about the ratings given to their nursing home, prevents vague wording from being translated into unbridled discretion for inspectors, and therefore unreliability.

Like any solution, of course, there are limits to it. There will arise interpretations of 'freedom', subjectively defined by residents, that will be disallowed by inspectors as beyond the range of acceptable definitions of freedom or impossible to satisfy in a nursing home context. Freedom to assault other residents is a case in point, but this still allows wide scope for a disparate variety of constructions of freedom that are tolerated in a liberal democracy, even if they are abhorrent to the inspectors. For example, one Australian resident invoked the standards to insist on his freedom to use the services of a prostitute; this conception of freedom was respected.

Whatever its limitations, the Australian policy is an attempt to break radically with the past in an approach to averting overinclusiveness and complexity, while using dialogue and empowerment to render vague wording outcome-accountable. This puts an alternative to an American nursing-home industry that pleads for simplicity in the law at the same time as it drives the law to greater complexity through persistent complaints about inconsistency. It puts an alternative to consumer advocates who call for more outcome-oriented and resident-centred regulation, but who scream unaccountability and unenforceability as soon as vague wording appears in regulations. It puts an alternative to regulators who want consistency by some means other than back-breaking protocols that actually worsen reliability. It challenges the clarity of thinking in American governmental reports that pretend mutually contradictory critiques such as the following can be reconciled without radical regulatory transformation: 'There are too many regulations'; 'The regulations are too vague'; and 'The regulations are too picky, detailed' (Illinois Legislative Investigating Commission, 1984: 20–1).

The transformations we advocate involve: (a) opting for standards that are simple and few in number (in preference to many specific rules); (b) structuring the regulatory process to be resident-centred and outcome-oriented; and (c) trusting dialogue (not just top-down 'training') among people who have been persuaded to care about those outcomes. These are our conditions for regulatory context that renders text reliable.

These conclusions can be read as part of a more general questioning of the claim that precision and the elimination of discretion through detailed regulatory law is a path to either greater consistency or equity (Hawkins, 1992). Precision tends to 'permit by implication conduct that the rule was intended to forbid' (Posner, 1977: 425). Detailed laws can provide a set of signposts to navigate

around for those with the resources to employ a good legal navigator (Schoer, 1993). While our focus here has been restricted to the effects of precision on consistency, precision-driven inconsistency can be theorized as a strategic resource of the powerful, particularly of repeat players who have an interest in playing for rules as well as for outcomes (thereby opening up a pattern of loopholes that suit the big players). As Max Weber (1954) showed: '[T]he more formal and complex the body of law becomes, the more it will operate in favour of formal rational bureaucratic groups such as corporations' (Sutton and Wild, 1978: 195). Marching under the banner of consistency, business can co-opt lawyers, social scientists, legislators and consumer advocates to the delivery of strategically inconsistent regulation of limited potency.

### NOTES

This project has enjoyed the funding support of the Australian Department of Health, Housing and Community Services, the Australian Research Council, the American Bar Foundation and the Australian National University. The authors are indebted to the support of their colleagues on the Nursing Home Regulation in Action Project, David Ermann, Diane Gibson, Miriam Landau and Toni Makkai. Valerie Braithwaite benefited from a visiting appointment at the Buehler Center on Aging, Northwestern University, during the US field work. Thanks to Paul Finn and James Morauta for comments on this article.

1. At another point, Kelman (1987: 44) puts the hope that has been realized here another way: 'Rules are bad because they inevitably have gaps and conflicts and are thus less mechanically applicable than they might appear. . . . The open invocation of an apparently vague standard, though, may be reasonably predictable in practice because even relatively detailed tacit community norms so converge that application of vague policy sentiments to cases poses little danger of disagreement.' It would be overstating our results to say that we found little disagreement, though one might say these data show 'a surprisingly low level of disagreement'.
2. '[C]onversation and explanation of one's conduct are avoided . . . a rule's a rule, don't complain to me' (Kelman, 1987: 63).
3. There is a 20-page set of Standards Monitoring Guidelines (Department of Community Services and Health, 1988). But these instruct inspectors only in 'key issues' and some of the things to 'look for' under each standard.
4. A British study was conducted of the reliability of two standard protocols for the inspection of residential care homes (Gibbs and Sinclair, 1991). Residential care homes are for residents who require less extensive care than those in nursing homes, though in practice the extensiveness of care between the two types of institution overlaps considerably. Gibbs and Sinclair's (1991) results from 48 homes fall between the US and Australian results reported in this article. A global reliability coefficient of 0.38 increased to 0.67 when the instrument was culled to include only high reliability items. This was a test-retest reliability with four weeks between inspections.
5. Part of the study was to check whether unreliability was caused by the reliability rater, in part, being in the facility at different times from the official team. This turned out to be a very minor source of unreliability (Braithwaite et al., 1991).
6. The definitions of these three categories are:  
*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.

7. The Australian director of nursing is the chief executive of the nursing home, combining the functions of administrator and director of nurses in the American system.
8. 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 US process from 1986 to 1990, was slightly more resident-centred than the old process was (though nowhere near as resident-centred as the Australian process). This 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 percent more deficiencies than did the teams surveying the same standards under the less resident-centred process (Spector and Drugovich, 1989).
9. One might hypothesize, for example, that American nursing homes being larger make reliability more difficult. But within our Australian study, the reliabilities are much the same for large and small homes. Large Australian homes are rated much more reliably than those in the Wisconsin study that produced the best US result.
10. This was a major reform to nursing home regulatory law. The new survey process based on the reform came into effect in October 1990. Omnibus Budget Reconciliation Act of 1987, Pub. L. No. 100-203, Paras 4203-4213.
11. A limited amount of such cross-referencing, however, may be done by a 'quality assurance' officer back at head office.
12. This was a 1989 inspection which predated abolition of the distinction among 'elements', 'standards' and 'conditions of participation' in Medicaid.
13. Rorty would say there is no way of solving it at all: 'words take their meaning from other words rather than by virtue of their representative character, and the corollary that vocabularies acquire their privileges from the men who use them rather than from their transparency to the real' (Rorty, 1979:368; see also Wittgenstein, 1972:29-39).

#### REFERENCES

- Atiyah, P. S. and Robert S. Summers (1987) *Form and Substance in Anglo-American Law*. Oxford: Clarendon Press.
- Badaracco, Joseph L. (1985) *Loading the Dice: A Five Country Study of Vinyl Chloride Regulation*. Boston, MA: Harvard Business School Press.
- Baldwin, Robert and Keith Hawkins (1984) 'Discretionary Justice: Davis Reconsidered', *Public law*: 570-89.
- Bardach, Eugene and Robert A. Kagan (1982) *Going by the Book: The Problem of Regulatory Unreasonableness*. Philadelphia, PA: Temple University Press.
- Braithwaite, John (1987) 'Negotiation versus Litigation: Industry Regulation in Great Britain and the United States', *American Bar Foundation Research Journal* 2: 559-74.
- Braithwaite, John (1994) 'The Nursing Home Industry', pp. 11-54 in A. Reiss and M.



- Tonry (eds), *Organizational Crime: Crime and Justice: A Review of Research Volume V*. Chicago, IL: University of Chicago Press.
- Braithwaite, John, Valerie Braithwaite, Diane Gibson and Toni Makkai (1992) 'Progress in Assessing the Quality of Australian Nursing Home Care', *Australian Journal of Public Health* 16(1): 89-97.
- Braithwaite, John, Toni Makkai, Valerie Braithwaite and Diane Gibson (1993) *Raising the Standard: Resident-Centred Nursing Home Regulations in Australia*. Canberra: Australian Government Publishing Service.
- Braithwaite, John, Toni Makkai, Valerie Braithwaite, Diane Gibson and David Ermann (1990) *The Contribution of the Standards Monitoring Process to the Quality of Nursing Home Life*. Canberra: Department of Community Services and Health.
- Braithwaite, John and Toni Makkai (1993) 'Can Resident-Centred Inspection of Nursing Homes Work with Very Sick Residents?', *Health Policy* 24: 19-33.
- Braithwaite, Valerie, John Braithwaite, Diane Gibson, Miriam Landau and Toni Makkai (1991) *The Reliability and Validity of Nursing Home Standards*. Canberra: Department of Health, Housing and Community Services.
- Day, Patricia and Rudolf Klein (1987) 'The Regulation of Nursing Homes: A Comparative Perspective', *Milbank Quarterly* 65: 303-47.
- Department of Community Services and Health (1988) *Nursing Home Standards Monitoring Guidelines*. Canberra: Department of Community Services and Health.
- Diver, Colin S. (1980) 'A Theory of Regulatory Enforcement', *Public Policy* 28: 257-99.
- Diver, Colin S. (1989) 'Regulatory Precision', pp. 199-232 in K. Hawkins and J. M. Thomas (eds), *Making Regulatory Policy*. Pittsburgh, PA: University of Pittsburgh Press.
- Gibbs, Ian and Ian Sinclair (1991) 'A Checklist Approach to the Inspection of Old People's Homes', Report to the Department of Health. York Department of Social Policy and Social Work, University of York.
- 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.
- Goodin, Robert E. (1982) *Political Theory and Public Policy*. Chicago, IL: University of Chicago Press.
- Grabosky, Peter and John Braithwaite (1986) *Of Manners Gentle: Enforcement Strategies of Australian Business Regulatory Agencies*. Melbourne: Oxford University Press.
- Gustafson, David H. (Project Manager) (1977) *Monitoring Quality of Nursing Home Care: Final Report*. Madison, WI: Department of Health and Social Services.
- Gustafson, David H., Charles J. Fiss, Judy C. Fryback, Peggy A. Smelser and May E. Hiles (1980) 'Measuring the Quality of Care in Nursing Homes: A Pilot Study in Wisconsin', *Public Health Reports* 95(4): 336-43.
- Gustafson, David H., Robert Peterson, Sandra Casper, Ann Macco, Richard Van Koningsveld and Edward Kopetsky (1982) *The Impact of the Wisconsin Quality Assurance Project: A Field Evaluation*. Madison, WI: University of Wisconsin.
- Haley, John O. (ed.) (1988) *Law and Society in Contemporary Japan: American Perspectives*. Dubuque, IA: Kendall Hunt.
- Hawkins, Keith (1984) *Environment and Enforcement: Regulation and the Social Definition of Pollution*. Oxford: Clarendon Press.
- Hawkins, Keith (ed.) (1992) *The Uses of Discretion*. Oxford: Clarendon Press.
- Illinois Legislative Investigating Commission (1984) *Regulation and Funding of Illinois Nursing Homes*. Springfield, IL: Illinois General Assembly.
- Institute of Medicine (1986) *Improving the Quality of Care in Nursing Homes*. Washington, DC: National Academy Press.
- Kagan, R. A. (1991) 'Adversarial Legalism and American Government', *Journal of Policy Analysis and Management* 10(3): 369-406.

- Kelman, Mark (1987) *A Guide to Critical Legal Studies*. Cambridge, MA: Harvard University Press.
- Kelman, S. (1981) *Regulating America: Regulating Sweden: A Comparative Study of Occupational Safety and Health Policy*. Cambridge, MA: MIT Press.
- Kennedy, Duncan (1976) 'Form and Substance in Private Law Adjudication', *Harvard Law Review* 89: 1685–778.
- Krotoszynski, R. J. (1990) 'Autonomy, Community, and Traditions of Liberty – The Contrast of British and American Privacy Law', *Duke Law Journal* 6: 1398–454.
- McLeay Report: The House of Representatives Standing Committee on Expenditure (1982) *In a Home or at Home: Accommodation and Home Care for the Aged*. Canberra: Parliament of the Commonwealth of Australia, October.
- Meidinger, Errol (1987) 'Regulatory Culture: A Theoretical Outline', *Law and Policy* 9: 355.
- Michelman, Frank I. (1986) 'The Supreme Court, 1985 Term – Foreword: Traces of Self-Government', *Harvard Law Review* 100: 4–77.
- Office of Health Systems Management (1985) *Evaluation of the New York State Residential Health Care Facility Quality Assurance System*. Standards and Surveillance, Division of Health Care, Office of Health Systems Management. Albany, NY: New York State Department of Health.
- Phillips, Colleen (1987) *Developing Outcome Norms to Monitor Quality of Care in Rhode Island Nursing Homes*. Providence, RI: Rhode Island Department of Health.
- Posner, Richard (1977) *Economic Analysis of Law*. Boston, MA: Little, Brown.
- Radin, Margaret Jane (1989) 'Reconsidering the Rule of Law', *Boston University Law Review* 69: 781.
- Rorty, Richard (1979) *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.
- Rose, Carol M. (1987) 'Crystals and Mud in Property Law', *Stanford Law Review* 40: 577–610.
- Schlag, Pierre (1985) 'Rules and Standards', *UCLA Law Review* 33: 379–430.
- Schoer, Ray (1993) 'Self-Regulation and the Australian Stock Exchange', in P. Grabosky and J. Braithwaite (eds), *Business Regulation and Australia's Future*. Canberra: Australian Institute of Criminology.
- Schuck, Peter H. (1979) 'Litigation, Bargaining and Regulation', *Regulation* 3: 26–34.
- Shearing, Clifford D. and Richard V. Ericson (1991) 'Towards a Figurative Conception of Action', *British Journal of Sociology* 42: 481–506.
- Spector, William D. and Margaret L. Drugovich (1989) 'Reforming Nursing Home Quality Regulation: Impact on Cited Deficiencies and Nursing Home Outcomes', *Medical Care* 27: 789–801.
- Spector, William D., Adrianna H. Takada, Margaret L. Durgovich, Linda L. Laliberte and Richard Tucker (1987) *PaCS Evaluation: Final Report*. Baltimore, MD: Health Care Financing Bureau.
- Sutton, Adam and Ron Wild (1978) 'Corporate Crime and Social Structure', in P. Wilson and J. Braithwaite (eds), *Two Faces of Deviance*. Brisbane: University of Queensland Press.
- Sullivan, Kathleen, M. (1992) 'The Justices of Rules and Standards – Foreword', *Harvard Law Review* 106(1): 24–123.
- Upham, Frank K. (1987) *Law and Social Change in Postwar Japan*. Cambridge, MA: Harvard University Press.
- Vogel, David (1986) *National Styles of Regulation: Environment Policy in Great Britain and the United States*. Ithaca, NY: Cornell University Press.
- Vogel, Ezra F. (1979) *Japan as No. 1: Lessons for America*. Cambridge, MA: Harvard University Press.
- Weber, Max (1954) *On Law in Economics and Society*. New York: Clarion.

- White, James Boyd (1984) *When Words Lose Their Meaning: Constitutions and Reconstitutions of Language, Character and Community*. Chicago, IL: University of Chicago Press.
- Wiener, Carolyn L. and Jeanie Kayser-Jones (1989) 'Defensive Work in Nursing Homes – Accountability Gone Amok', *Social Science and Medicine* 28(1): 37–44.
- Wittgenstein, Ludwig (1972) *Philosophical Investigations*, trans. G. E. M. Anscombe. Oxford: Blackwell.
- Zimmerman, David, J. Robert Egan, David Gustafson, Charles Metcalf and Felicity Skidmore (1985) *Evaluation of the Three State Demonstrations in Nursing Home Quality Assurance Procedures*. Baltimore, MD: Health Care Financing Bureau, September.