

How Much of What Quality? A Comment on Conscientious Procedural Design

Jerry L. Mashaw

Follow this and additional works at: <http://scholarship.law.cornell.edu/clr>

 Part of the [Law Commons](#)

Recommended Citation

Jerry L. Mashaw, *How Much of What Quality? A Comment on Conscientious Procedural Design*, 65 Cornell L. Rev. 823 (1980)
Available at: <http://scholarship.law.cornell.edu/clr/vol65/iss5/8>

This Article is brought to you for free and open access by the Journals at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Review by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

HOW MUCH OF WHAT QUALITY? A COMMENT ON CONSCIENTIOUS PROCEDURAL DESIGN

Jerry L. Mashaw[†]

The OHA benefits adjudication process, as described by Chassman and Rolston,¹ seems badly awry. Evidence marshalled by them and by others² reveals that ALJ award rates vary widely, and that nearly one-half of all administrative denials appealed to federal district courts are reversed or remanded.³ Moreover, although ALJ's hear relatively few cases, grants by ALJ's represent a large and increasing percentage of the total awards made by the system.⁴ A neutral observer might interpret this as evidence of an adjudicatory system wildly out of control.

Yet OHA takes a quite modest approach to quality assurance. Quality assurance reviewers only note relatively clear errors. To avoid adjudicative delay, they postpone review (post-effectuation) until correction of erroneous awards is impossible. Errors are not remanded directly to the ALJ who made the mistake. OHA recognizes that these procedures limit the instructional value of quality assurance review and undermine the morale of the reviewers.⁵ Moreover, although studies show that devoting more resources to even this limited form of quality assurance review would be "cost-effective," budgetary concerns preclude all but marginal increases in the monitoring capacity of OHA. Why such a timid approach?

This is not to say that OHA's oversight stance is irrational; it is a function of the internal and external politics of the bureau. ALJ's are too powerful to be subjected to intensive review. Congress is unwilling to increase substantially funds for administra-

[†] Professor of Law, Yale University. B.A. 1962, LL.B. 1964, Tulane University; Ph.D. 1969, University of Edinburgh.

¹ P. 803 *supra*.

² See p. 807 *supra*; J. MASHAW, C. GEOTZ, F. GOODMAN, W. SCHWARTZ, P. VERKUIL & M. CARROW, SOCIAL SECURITY HEARINGS AND APPEALS 21-24 (1978) [hereinafter cited as HEARINGS AND APPEALS].

³ *Id.* at 128-30.

⁴ See Memorandum to the Secretary and Under-Secretary from D. Marcus, Project Manager, Disability Work Group, Department of Health, Education, and Welfare, Office of the Secretary, "Options Paper on Disability Initiative—Summary" (July 10, 1978) [hereinafter cited as Options Paper] (on file at the *Cornell Law Review*).

⁵ See p. 814, *supra*.

tion.⁶ The important question really is whether the current system can be justified in terms *other than* the lamentable state of political affairs. In other words, if one were to design the disability decision process—including its quality assurance component—with a free hand, would a different system be more attractive?

The payoff from a more effective quality assurance system would be error reduction and, presumably as a side-effect, increased consistency. Offsetting these benefits would be increased administrative expenses, loss of ALJ discretion, and increased delay. The policy question, then, is whether the benefits of a more aggressive quality assurance system outweigh its costs. This raises a corollary issue: how to conceptualize and quantify these costs and benefits in order to measure them. What exactly are the costs of error, inconsistency, or delay—and what are the benefits of reducing them?

I

ERROR REDUCTION

One method of measuring the benefit of error reduction is to ask, "How much is a disability claim worth?" The answer, on average, is about \$30,000.⁷ Is it worth \$30,000 per claim not to make an error? If so, enormous increases in the budget for administration would seem warranted since these \$30,000 claims are being decided, on average, for less than \$500 per claim.⁸ Thus, at the current cost/benefit ratio (1:60), a change in the decision process that increased accuracy by 1% would justify increasing administrative expenses by 60%. If this approach correctly determines the appropriate level of administrative expense in the dis-

⁶ See generally *Delays in Social Security Appeals, Hearings Before the Subcomm. on Social Security of the House Comm. on Ways and Means, 94th Cong., 1st Sess. (1975)* [hereinafter cited as *Hearings*].

⁷ See General Accounting Office, Report on Controls Over Medical Examinations Necessary for Better Determinations of Disability by the Social Security Administration, H.R.D. 79-119 (Oct. 9, 1979).

⁸ The total administrative expense allocable to the disability program for 1978 was \$327,000,000. H.R. REP. NO. 100, 96th Cong., 1st Sess. 19 (1979). During this same period, the program processed 1,300,000 cases. See Options Paper, *supra* note 4. Dividing the cost by the number of cases gives an average case processing cost of about \$300. In fact, since the administrative costs of the program include expenditures for items other than the processing of claims, this figure is somewhat high. However, this article assumes a more conservative \$500 average case processing cost.

ability program, it would provide substantial justification for costly error correction procedures.⁹

But this suggested approach is troublesome. Saving \$30,000 in program costs by avoiding an erroneous grant seems reasonably straightforward. However, does spending \$30,000 to avoid an erroneous denial also produce a \$30,000 savings? The problem here is not that reducing false negatives (denials) is less important than reducing false positives (grants), but that the units of measurement are wrong. Thirty thousand dollars is \$30,000 whoever is holding it—Social Security Administration (SSA) (for taxpayers) or recipients. When money is transferred from one pocket to another it must be done because it is more valuable in the hands of the transferee than the transferor. "More valuable," however, connotes some social welfare measurement that can not be expressed in dollars, such as greater "utility."

Furthermore, all disability transfers are not equally valuable. Some beneficiaries seem more "deserving" than others and some claims are denied because they lack relative, rather than absolute, merit. Functional ability or disability is, after all, a continuum; only the decisions are bi-polar. When the relative value of transfers is translated into the relative value of errors, a decision classifying Bruce Jenner as disabled or Karen Quinlan as not disabled is intuitively a costlier error than an erroneous assignment of the perennial marginal case (say a fifty-five year old, semi-skilled, white male with high blood pressure and chronic lower back syndrome) to either category. We need to develop an analysis of error reduction that recognizes that errors involve social costs of differing magnitudes.

Although not without serious difficulties, such an analysis can be made by charting disability claims on two related scales, scales that reflect the social benefits and social costs of disability decisions. First, the value of disability benefits is greatest to individuals who are severely handicapped and for whom work, were it possible at all, would only add to the pain and frustration that accompany severe physical or emotional illness. The social value of paying disability benefits to these claimants is obviously great; these payments reflect the core humanitarian instincts of the social insurance program. The social benefit of disability payments declines as handicaps become less severe and impose lower costs on work effort. Finally, disability payments made to an individual

⁹ Chassman and Rolston also discuss the cost-effectiveness of additional review. See p. 820 *supra*.

without any physical or emotional illness produce no social value at all.

Claims can also be placed on a social cost scale. These costs include lost productivity when people leave the work force to receive disability payments, and the demoralization that taxpayers experience when forced to contribute to the support of others. Claimants would occupy roughly complementary positions on the two scales. The most "deserving"—those with respect to whom we derive the most satisfaction from paying benefits—are also those from whom society misses least when they leave the work force and whose support at public expense imposes smaller psychic costs on taxpayers.

Figure I

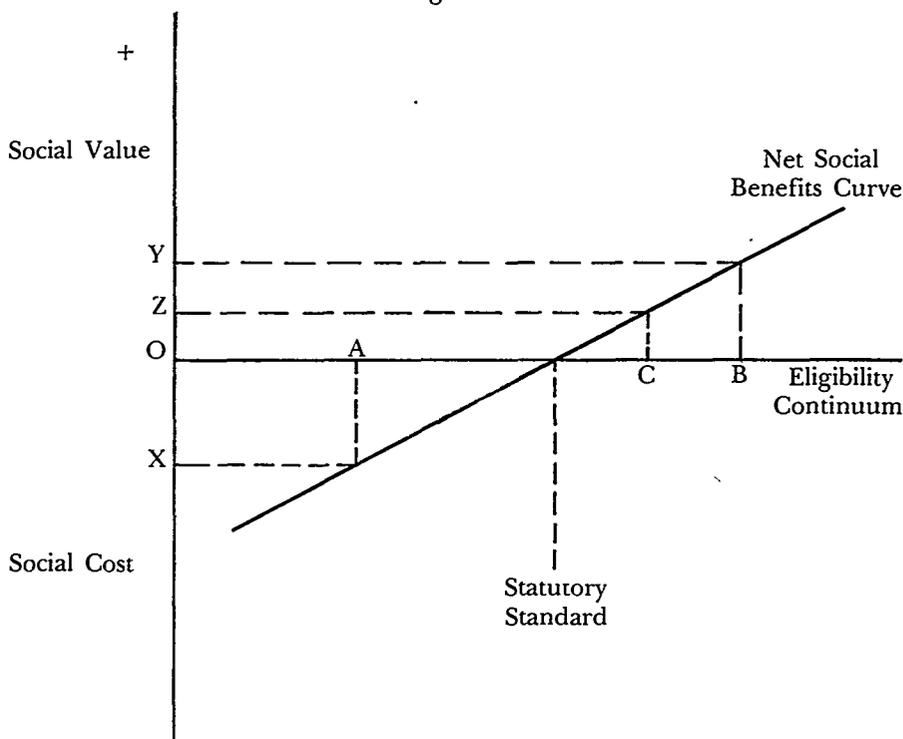


Figure I depicts these combined scales.¹⁰ The net social benefit curve rises from left to right as we consider persons who have increasingly severe disabilities. The intersection of the social benefits curve with the horizontal axis should represent the statutory

¹⁰ For a more detailed discussion of this graphic representation, see *HEARINGS AND APPEALS*, *supra* note 2, at 16-18.

definition of disability. To the right of the intersection, the social benefit of making the payment exceeds the social costs; the disability program defines individuals falling in this segment of the graph as "disabled" in order to capture these net social benefits. To the left of the intersection, social costs exceed social benefits and the individual is accordingly classified as "not disabled" to avoid these net social costs.

The implications of this analysis for allocating administrative costs are rather straightforward. If a claimant "really" at A is found to be at B, there is a net social loss (OX). Similarly, if a claimant "really" at B is found to be at A, the result is a net social loss (OY). The foregone benefits in the latter case are greater than those associated with a similar error concerning a claimant "really" at position C (OZ). Presumably, the more costly the potential error, the more society should be willing to pay to avoid it.

Although this analysis accounts for the political judgments that seem to underlie transfer programs, it substantially ignores operational content. Indeed, it has lapsed into the poetic language of social utility—a commodity that not only has no dollar prices attached to it, but that also assumes that individual preferences can be aggregated in some unspecified currency such as "utils" or "ergs".

Nevertheless the analysis has important implications. When Chassman and Rolston assert that ALJ decisions are highly "judgmental," they may simply be saying that these decisions involve close cases. That is, the applicant is very near the statutory standard: if disabled, not severely so; if not disabled, nevertheless seriously impaired. To the extent that errors occur in these marginal cases, they are relatively inexpensive in social welfare terms. Furthermore, error detection and correction are likely to be relatively more expensive in these cases. If most ALJ hearing decisions involve such claims, accuracy should not be too important since error reduction would be expensive and produce only marginal social benefits.

Nonetheless, that most cases are marginal does not imply that all are. OHA should still attempt to correct or avoid the higher cost errors that emerge from its caseload. This analysis supports OHA's decision to focus its quality assurance system on relatively clear—and thus presumably high cost—errors and to spend only a small percentage of total claim values or program costs on quality assurance.

The approach may also explain OHA's decision to use independent, and therefore politically troublesome ALJ's rather than a

less independent and more manageable corps of adjudicators. Since most cases heard by ALJ's are close ones, they can be considered "accurate" regardless of the outcome. The hearing process, then, functions primarily as a means of legitimizing dispute resolution, rather than to ensure "correct" adjudication. To accomplish this task of legitimation, a personalized decision process with a legally independent decisionmaker has much to recommend it.¹¹ The costs imposed on quality assurance by the independent ALJ system may be outweighed by the benefits of quasi-judicial symbolism in legitimizing dispute resolution.

II

INCONSISTENCY

Simply aggregating the costs of individual marginal errors may understate the true costs of ALJ decisionmaking. Arguably, a general "demoralization cost" should be added to that sum to reflect society's disappointment or anxiety when confronted with the system's apparent inability to make consistent determinations. That such costs exist is evident from the concern with inconsistency that Chassman and Rolston¹² and prior studies of the disability program express.¹³ If these costs are particularly associated with marginal cases, then the inconsistency costs in those cases may yet be sufficiently significant to warrant concern.

Concern about consistency does not, however, translate directly into a demand for increased quality assurance activity of the conventional error-assignment error-correction sort. If the cases are marginal ones in which errors are difficult (perhaps impossible) to identify reliably, the perception of inconsistency may result simply from interpretive differences between deciders and observers or participants or among the deciders themselves. "Inconsistency" and its associated costs can be addressed in this situation only by reducing the range of interpretive judgment in cases that are neither clear grants nor clear denials. Moreover, SSA must devise techniques for increasing consistency in its disability adjudication without incurring other disproportionate costs.

¹¹ See generally J. THIBAUT & L. WALKER, *PROCEDURAL JUSTICE* (1975).

¹² See p. 818 *supra*.

¹³ See, e.g., R. DIXON, *SOCIAL SECURITY DISABILITY AND MASS JUSTICE* 51-92 (1973); COMPTROLLER GENERAL OF THE UNITED STATES GENERAL ACCOUNTING OFFICE, *A PLAN FOR IMPROVING THE DISABILITY DETERMINATION PROCESS BY BRINGING IT UNDER COMPLETE FEDERAL MANAGEMENT SHOULD BE DEVELOPED* (1978).

For example, SSA could impose relatively objective standards for deciding disability cases. Claimants could be classified according to age (*e.g.*, 18-44, 45-60, 60 and above), residual functional capacity (*e.g.*, sedentary, light, medium, or heavy work as defined by the Dictionary of Occupational Titles), education (*e.g.*, illiterate, 1-8 grades, high school diploma or more) and work experience (*e.g.*, unskilled, semi-skilled, skilled as defined by the Dictionary of Occupational Titles). These categories could then be combined in a grid so that each set of the four categorical positions yields a determinate conclusion—disabled or not disabled.

The use of such a device would increase the consistency of determinations enormously. But its value cannot be assessed without considering its obvious costs. For one thing, many will charge that such a system is irrationally grounded on overgeneralizations, and invokes stereotypes that inevitably make large differences in outcomes turn on insignificant differences in relevant facts. Under this system, a claimant's grant or denial could depend on his being 45 years old rather than 44, or having a ninth rather than an eighth grade education.

Yet, it is not clear that these objections should be taken too seriously. If the generalizations are being made with respect to claimant situations that presently might be decided either way without error, it is not the case that the grid would "overgeneralize" in the sense of forcing erroneous classifications. The overgeneralization argument might, however, have another basis. The proposed categorical method contradicts the ideal of individualized determination. To the extent that this ideal legitimizes the decisionmaking process, a categorical methodology has substantial costs.

Similarly, the substantial disparity in the grant and denial rates of individual ALJ's¹⁴ could be eliminated or sharply constrained by a quota system. All ALJ's, for example, could be assigned a grant rate (*e.g.*, 35% \pm 5% per month). Awards would then be made on a comparative basis and award rate disparities would virtually disappear. Such systems are widely used to reduce grading disparities in educational institutions.

A quota system, however, contradicts the basic goals of the disability program. The purpose of social security is to provide insurance against certain potential or inevitable hazards. Although disability adjudicators may decide cases with regard to some im-

¹⁴ For a discussion of these disparities, see p. 807 *supra*.

plicit "curve" describing the relative deservedness of claimants, a procedure that explicitly utilizes this "curve" would conflict with the statutory policy of individual entitlement. It would interject an element of competition for benefits among participants. Such competition may be appropriate in other agency adjudicatory contexts, such as FCC licensing or the award of government contracts, but it has little place in the Social Security program.

The use of rotating panels of ALJ's instead of individual decisionmakers could also reduce the award rate disparities. Decisions by majority vote of an odd-numbered panel will produce more uniform results than if cases are decided by individual judges with diverse characteristics.¹⁵ This technique, however, would substantially increase administrative costs. Like a quota system, it may also limit actual disparity without significantly lessening the appearance of inconsistency when particular, close cases are compared.

Finally, consistency might be improved by skewing the decision calculus with a presumption to either award or deny benefits in all close cases. But, to justify this approach, SSA must have some reason to believe that positive errors are systematically either more or less costly than negative errors. Nothing in the statute, regulations or decisional structure of the program convincingly guides SSA to choose to incur either type of error in doubtful cases.¹⁶

Although this article earlier suggested that the individual error costs in marginal cases were relatively unimportant, society's perception of inconsistencies in the disability program is potentially costly. Perceptions of inconsistency may be unrelated to the existence of errors and thus beyond the reach of any quality assurance system; nonetheless, the appearance of inconsistency remains a source of tribulation for the disability system. Unfortunately, attempts to reduce inconsistencies inevitably increase administrative costs or create new costs, such as overgeneralization or abandoning important program goals. Given these difficulties, it is understandable that SSA has moved to make the disability standard¹⁷ more objective only after years of criticism. Nor is it surprising that the "grid" methodology adopted is less rigid than

¹⁵ See HEARINGS AND APPEALS, *supra* note 2, at 26-27.

¹⁶ Cf. R. DIXON, SOCIAL SECURITY DISABILITY AND MASS JUSTICE 59 (1973) (asserting presumption against disability claimant in borderline cases and congressional endorsement of this approach in 1967 disability insurance amendments).

¹⁷ See 20 C.F.R. §§ 404.1-1610, 416.101-2119 (1979).

the one previously hypothesized.¹⁸ Under the actual system, decisionmakers have numerous ways to escape the grid¹⁹ and individualize their judgments.

III

DELAY

The issue of delay implicates much of the prior analysis of error costs. Error costs are borne over a period of time. Any delay in deciding that an initial claim or making a redetermination of an existing grant may incur the costs of a false negative or a false positive over a certain time period. (These costs may ultimately be partially compensated for by a decision providing retroactive benefits or through recoupment.) Moreover, like the relative costs of decisional errors, delays affecting the extremes of the disability spectrum are more costly than delays affecting the marginal cases.²⁰ Those whom society deems "deserving" of disability funds deserve them immediately. Delay costs can also be considered as a form of compound interest—each additional unit of delay in a case is more costly than previous ones.

Some aspects of the delay problem are not subject to error cost analysis. Even when decisions are correctly made, a claimant must wait for the determination. Although the uncertainty pending this decision may impose psychic costs on the claimant, there seems to be no good way to measure that cost. These psychic costs depend on the economic and psychological circumstances of the claimant and may be unrelated to the severity of his disability.

Delay considerations affect the structure of the quality assurance system in two ways. First, OHA monitors ALJ timeliness and productivity. The difficult question is how far to push the demand for timeliness. Delay reduction has political²¹ and judicial support²² and is the system's most non-problematic quality measurement. Yet Congress' unwillingness to incur additional administrative costs suggests that reductions in delay will only come at the cost of increased error unless adjudicatory "technology" can somehow be improved.

¹⁸ See text accompanying notes 7-11, *supra*.

¹⁹ See, e.g., 20 C.F.R. § 404.1539, Subpart P, App. 2, § 200.00(e) (1979).

²⁰ See Table I, *supra*.

²¹ See *Hearings*, *supra* note 6, at 128-30.

²² See *White v. Mathews*, 559 F.2d 852, 858-60 (2d Cir. 1977) (holding delays averaging 211.8 days in Title II disability appeals unreasonable and approving prospective payments to claimants where appeal not timely adjudicated).

This trade-off highlights the second structural effect. OHA is unwilling to invoke a more effective, pre-effectuation, quality assurance system because this would increase delay costs. Is this justifiable? Perhaps, but I question whether a cost-benefit calculation will rationalize the decision. Assume that the OHA system decides 100,000 appeals or initial claims per year, that post-effectuation quality assurance review reduces errors by 1%, while pre-effectuation review reduces them by 2%, and that the average payment period for disability benefits is 60 months. The refusal to use pre-effectuation review costs 60,000 error-months yearly (the number of months benefits are erroneously paid). This figure is calculated from the product of the error reduction foregone (2%-1%), the number of cases per year (100,000), and the average payment period (60 months). If a 5% pre-effectuation review sample is delayed three months, and even if all the delayed cases are deemed to be grants, then 15,000 error-months are incurred (5% x 100,000 x 3). Thus, delay costs, such as the psychic costs of uncertainty, would have to be three times as great as the error-type delay costs to make pre-effectuation review a bad bargain. This multiplier may be even greater because the example ignores the fact that error-type delay costs are in part made up by recoupment and payment of retroactive benefits.

Of course, my hypothetical figures might be unrealistic; pre-effectuation leverage may be lower than I suggested. But I suspect that the better explanation for post-effectuation quality assurance review at OHA is that distributional concerns dominate maximizing total social welfare. Persons who have had to wait until the hearing stage for their disability benefits seem particularly inappropriate people to "tax" by further delaying their claims.

IV

CONCLUDING COMMENTS

This Comment has attempted to provide a basis for conscientious procedural design in the context of the problems and techniques discussed by Chassman and Rolston. One might conclude from the discussion that administrators designing control mechanisms in response to the political realities confronting them may nevertheless emerge with systems that are rational from a more detached perspective. But such happy outcomes do not always emerge from indirection. Administrators must confront the difficult cost-benefit questions raised by procedural design. They

should decide how important different types of errors really are; whether a perception of consistency can be purchased at any reasonable cost; whether values other than accuracy are at stake in the procedural system; and how the costs of reform should be distributed. Indeed, in elucidating and evaluating the relevant trade-offs, the political equation may be transformed.

A corollary to the justifications for OHA's post-effectuation quality-assurance system is that those justifications do not apply to the pre-hearing phases of the disability decision process. Pre-hearing decisionmaking is not limited to close cases, quality assurance does not threaten the dispute resolving legitimacy of independent deciders, and pre-effectuation review can be accomplished with dispatch. I would, therefore, expect that the payoff from quality assurance review would be much greater at the pre-hearing stage. In fact, SSA's current three-tier review system²³ is already an extensive system of quality assurance review of state agency decisionmaking. The system is also beginning to experiment with devices for targeting a more intensive review on error-prone portions of its caseload.²⁴

I am not overly disturbed by Chassman and Rolston's discovery that quality assurance may protect taxpayers proportionately more than disability beneficiaries. I agree with them that greater integrity in the decision process will ultimately benefit the truly eligible. Moreover, it is not clear that identifiable quality assurance errors are more frequent in awards than denials in all social welfare systems or, indeed, at every level of the disability system. For example, recent statistics from regional offices' review of state agency decisionmaking in SSI and concurrent SSI/disability cases revealed 20% more errors in state agencies' denials than in their awards.²⁵

More disturbing, however, is the potential for quality assurance systems to skew the incentives of social welfare adjudicators

²³ See Office of Program Operations, Social Security Administration, Disability Operating Manual, Part XV (June 1978).

²⁴ See, e.g., Memorandum from Office of the Regional Commissioner, Dallas Region, Social Security Administration to Associate Commissioner for Program Operations, Bureau of Disability Insurance, Report of Findings From Onsite Preadjudicative Review in Arkansas (Nov. 29, 1978) (on file at the *Cornell Law Review*).

²⁵ See generally Reports and Analysis Branch, Division of Quality Assurance, Office of Program Evaluation and Appraisal, Bureau of Disability Insurance, Social Security Administration, BDI Quarterly Report of CRS Review of Title XVI Initial Disability Determinations, February-March 1977, April-June 1977, July-September 1977, October-December 1977, January-March 1978, April-June 1978, July-September 1978, October-December 1978.

in inappropriate ways. Chassman and Rolston urge that OHA concentrate on awards in its quality system. If effective, this concentration will reduce awards more than denials. But such reduction is appropriate only after a finding of differential error rates based on a careful study such as the one the authors directed at OHA. In other quality assurance systems, for example, state public assistance quality assurance review²⁶ and SSA central office disability review,²⁷ a decision to concentrate on awards may be merely a reaction to the politics of the program or its administration. If that political instinct does not reflect a legislative judgment that erroneous awards are more costly or more likely than erroneous denials, using quality assurance to differentially "correct" adjudicators' awards induces an inappropriate skew in the decision process.

In my prior Article I thought that judicial review could remedy this concern. Although I still believe that quality assurance claims are justiciable and that the nonuse or misuse of such systems should be relevant to judges reviewing the procedural adequacy of public benefits adjudication, I am now less convinced that judicial review is the most appropriate vehicle for dealing with quality assurance issues. We may need to begin thinking about monitoring the monitors in a different way.

I do not now have a specific proposal. My rudimentary vision is of a "super-bureau" on benefits administration that would actively monitor the performance of line agencies, research administrative and managerial techniques, hear complaints about agency systems and provide "binding advice" on administrative reform. Such a review agency should be independent of the politics of particular programs and expert in the administration of public assistance and social welfare systems. It should focus on systemic complaints about the structure or operation of the adjudicatory process rather than on complaints concerning denial of benefits or individualized procedural irregularities.

²⁶ See Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy, Fairness, and Timeliness in the Adjudication of Social Welfare Claims*, 59 CORNELL L. REV. 772, 808-09 (1974).

²⁷ See generally Reports and Analysis Branch, Division of Disability Program Quality, Office of Assessment, Social Security Administration, Quarterly Report of Central Office Review of Title II Initial Disability Claims, April-June 1977, July-September 1977, October-December 1977, January-March 1978, April-June 1978, July-September 1978, October-December 1978.

But this is a story for another day. For now the best hope for quality assurance management is that agency professionals like Chassman and Rolston will continue to work at improving their capacity to ask and answer the right questions.