International Journal of Technology Assessment in Health Care, **17:1** (2001), 125–136.

USING DATA TO ENHANCE THE EXPERT PANEL PROCESS

Rating Indications of Alcohol-related Problems in Older Adults

Sabine M. Oishi
Health Services Research Center, UCLA

Sally C. Morton
RAND

Alison A. Moore
John C. Beck
Ron D. Hays
Karen L. Spritzer
Jennifer M. Partridge
Arlene Fink
UCLA

Abstract

Objective: To enhance the validity of a well-known expert panel process, we used data from patient surveys to identify and correct rating errors.
Methods: We used the two-round RAND/UCLA panel method to rate indications of harmful (presence of problems), hazardous (at risk for problems), and nonhazardous (no known risks) drinking in older adults. Results from the panel provided guidelines for classifying older individuals as harmful, hazardous, or nonhazardous drinkers, using a survey. The classifications yielded unexpectedly high numbers of harmful and hazardous drinkers. We hypothesized possible misclassifications of drinking risks and used the survey data to identify indications that may have led to invalid ratings. We modified problematic indications and asked three clinician panelists to evaluate the clinical usefulness of the modifications in a third panel round. We revised the indications based on panelist response and reexamined drinking classifications.
Results: Using the original indications, 48% of drinkers in the sample were classified as harmful, 31% as hazardous, and 21% as nonhazardous. A review of the indications revealed framing bias in the original rating task and vague definitions of certain symptoms and conditions. The modified indications resulted in classifications of 22% harmful, 47% hazardous, and 31% nonhazardous drinkers.
Conclusions: Analysis of survey data led to identification and correction of specific errors occurring during the panel-rating process. The validity of the RAND/UCLA method can be enhanced using data-driven modifications.

Keywords: Aged, Alcohol drinking, Geriatric medicine, Quality of health care

The authors thank Rebecca Shoai and Andrew Perry for assistance in preparing this manuscript. This project was funded by the U.S. National Institute on Alcohol Abuse and Alcoholism, contract no. NO1AA52009.

Expert panels are widely used to resolve issues of medical uncertainty, especially in situations for which rigorous studies, such as randomized controlled trials, are unavailable (9). In this approach, panelists make judgments about a question based on a systematic literature review, clinical experience, and discussion. One well-accepted panel method, the RAND/UCLA method, was developed to evaluate the appropriate use of medical procedures in population-based comparisons of practice patterns (9). In a two-round rating approach, panelists rate the appropriateness of a procedure on an ordinal scale (e.g., from very inappropriate to very appropriate) for different patient scenarios. The scenarios, or indications, combine symptoms, past medical history, and other relevant patient factors. For each indication, the distribution of panelists' ratings is categorized as meeting conditions for either agreement or disagreement. If conditions for agreement or disagreement are not met, the rating outcome is indeterminate.

The RAND/UCLA method has been used for purposes other than evaluating the appropriateness of procedures, such as determining priorities for examining quality of care for older patients (16). The reliability and construct validity of specific individual panels have been assessed favorably, and in recent years panel methods have been increasingly applied to clinical settings, as in guideline development (6;8;26;31). It has been shown, however, that similarly constructed panels rating the same indications do not arrive at the same conclusions often enough to guide individual clinical decision making (35). Certain barriers may limit the precision of the method. Panelists may interpret rating scales differently and may suffer from respondent burden in the face of long, complex rating tasks (25;32). Unintended biases in the wording of indications can also affect ratings, and cognitive biases of panelists may vary from panel to panel, depending on the backgrounds of selected experts (12;35). Problems with the precision of a panel's outcomes may occur most often when the method is modified for uses other than those for which it was originally designed.

We describe the use of the RAND/UCLA two-round method to develop indications of nonhazardous, hazardous, and harmful drinking in a general population of older adults, which is a new use of the method. We convened the panel[1] because scant data are available on the health effects of light to moderate alcohol consumption among the elderly (15). Existing definitions of problem drinking are often based on concepts of abuse and dependence and do not cover the spectrum of possible problems among older persons, for whom even small doses of alcohol can be dangerous (5;13). Older adults attain a higher blood alcohol level than do younger persons, and appear to have a greater central nervous system sensitivity to alcohol (19;23;36;37). They are also more likely to take medications that may interact with alcohol and to have chronic diseases or functional limitations that may be worsened by alcohol use (e.g., depression, hypertension, gastritis, and impairments in gait and cognition) (4;5;13;17). The literature identifies these and other patient factors likely to lead to poor health outcomes in combination with alcohol use; however, it provides minimal guidance for combining the risks of the multiple factors present for individual patients, especially at low to moderate consumption levels (1;2;14). Our panel assigned nonhazardous, hazardous, and harmful drinking designations to patient scenarios that combine relevant risk factors with varying amounts of alcohol use.

The assignment of a risk category as opposed to the traditional determination of the appropriateness of an intervention required modifications to the usual rating scale. In addition, the structure of the indications diverged from the usual format, because of the need to incorporate quantities of alcohol consumption along with the other more traditional health-related risk factors. We assessed the impact of these changes on panel ratings in the risk classification of elderly drinkers based on their responses to an elder-specific questionnaire, the Alcohol-Related Problems Survey (ARPS). This survey data allowed us to identify sources of possible measurement error (error in measuring the true opinions of panelists) in the panel results and to examine the clinical implications of the results empirically.

METHODS

The Expert Panel Method

The methods and results of our panel have been described (30). Briefly, based on a literature review of clinical indicators of alcohol-related risk and harm specific to the elderly, project geriatricians compiled a list of 846 indications to be rated by the panel. An indication consisted of a person's quantity and frequency of alcohol consumption in combination with medical conditions, symptoms, medications, physical functioning, emotional well-being, smoking, driving, and alcohol-related behaviors and consequences. An example of an indication is: "Persons report taking a total of 8 or more medications (including both prescription and non-prescription medications) and they report drinking at least 3 drinks four or more times a week or at least 4 drinks two or more times a week in the last 12 months" (Figure 1).

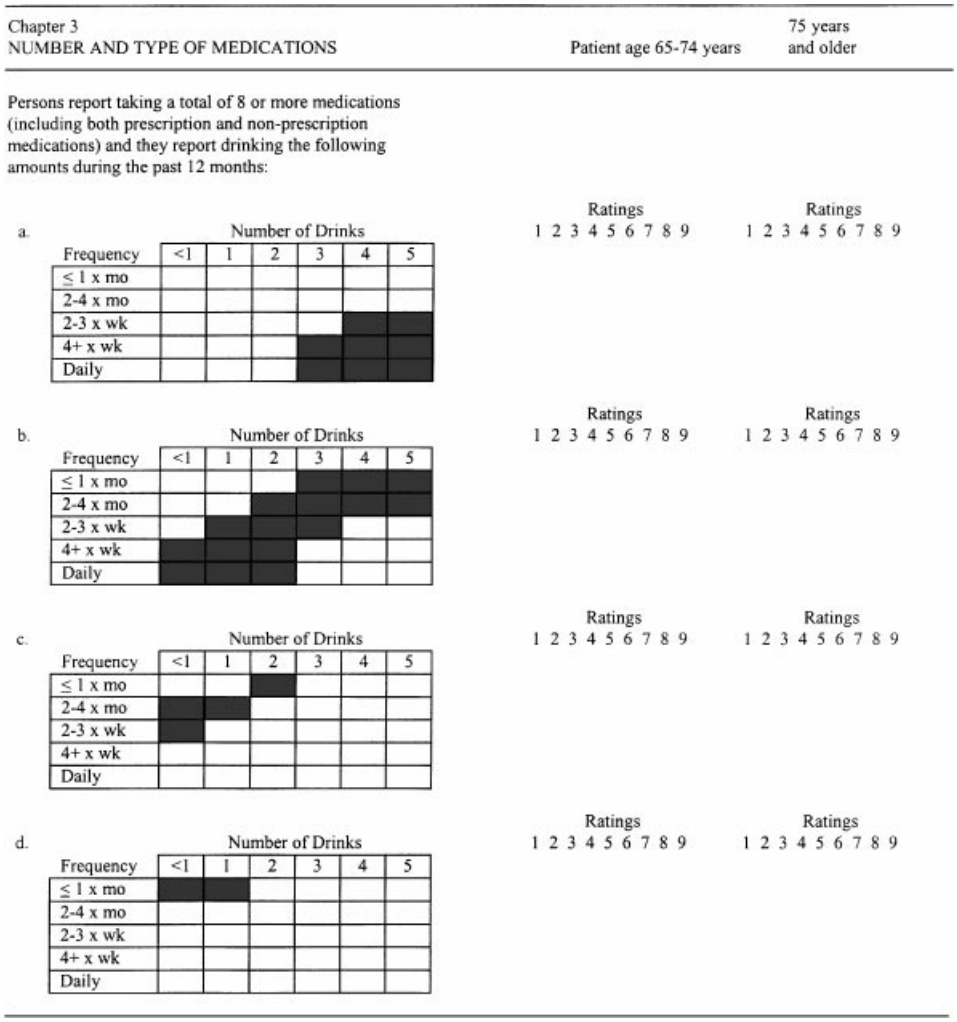

Figure 1. Example of a rating sheet. The shaded areas indicate different combinations of quantities and frequencies of drinking. For example, (a) indicates a person who drinks 4–5 drinks 2–3 times a week or 3–5 drinks 4 or more times a week or 3–5 drinks daily. Rating scale: Nonhazardous = 1–3, Hazardous = 4–6, Harmful = 7–9.

https://doi.org/10.1017/S0266462301104113 Published online by Cambridge University Press

The rating matrix was designed by the study geriatricians with involvement from one of the initial developers of the RAND/UCLA method (AF). Panelists did not participate in the development of the indications and the rating scheme. The panelists were nine experts in alcohol research, psychiatry, geriatrics, and internal medicine. Prior to round 1, we gave panelists a published literature review summarizing current evidence of the determinants and consequences of alcohol problems in older persons and copies of relevant papers, including available epidemiologic studies (15). Using the RAND/UCLA method, they initially produced two sets of ratings: the first by mail (round 1) and the second during a face-to-face meeting (round 2). For each indication, they assigned a rating of 1, 2, or 3 to represent use of alcohol without clear risk of physical or psychological complications (nonhazardous use); 4, 5, or 6 to represent use of alcohol associated with risk of future damage to physical or psychological health (hazardous use); and 7, 8, or 9 to mean a pattern of drinking already causing damage to health (harmful use, including but not limited to alcohol abuse and dependence) (7). We analyzed agreement, disagreement, and indeterminacy of panel ratings using traditional methods (30).

Implementation of the Panel Findings to Classify Elderly Patients into Drinking Categories

We used the panel results to categorize survey respondents into nonhazardous, hazardous, and harmful drinking groups. We wrote rules combining responses to the ARPS according to each panel indication and called the rules "algorithms." For example, an algorithm might combine the response to an alcohol consumption item with the response to an item asking about a diagnosis of hypertension. The algorithms were applied to the survey data to determine how many indications of harmful and hazardous drinking each respondent met. If at least one indication of harmful drinking was met, the classification was harmful; if no harmful indication was met and at least one hazardous drinking indication was met, the classification was hazardous. If no indications of harmful or hazardous drinking were met, the classification was nonhazardous.

Classification of a Test Sample. We used the panel findings to classify 161 older drinkers who had completed the ARPS during regular primary care clinic visits. This test sample was obtained from clinical practices, including the UCLA Internal Medicine Group Practice, Santa Barbara Medical Foundation Clinic, and King-Drew Medical Center in Los Angeles. Since the ARPS was designed to identify types of alcohol problems previously overlooked in older patients, we expected the classification rates indicating risky drinking to be higher than those published in the literature. However, we did not hypothesize what these rates might be. The logical evaluation standard was the clinical usefulness of the classifications, defined as face validity (intuitive acceptability to clinicians), content validity (scientific and medical correctness), clinical practicality (practicality for the intended setting), and clinical internal consistency (similar classifications for similar levels of compromise to health) (33).

Examination of the Test Sample Results. To begin an assessment of clinical usefulness, two study geriatricians examined the classifications for face validity: Did the proportions of subjects falling into nonhazardous, hazardous, and harmful categories appear intuitively reasonable? Would members of the medical community be likely to consider the rates meaningful based on their own clinical experience? The geriatricians determined that the harmful and hazardous rates among the test sample were higher than most clinicians would intuitively expect, particularly for respondents with certain conditions. While prevailing clinician belief is not a gold standard for judging the accuracy of panel outcomes, we used this face validity assessment to anticipate difficulties in acceptance of our findings. The combined harmful and hazardous rates were so high that study geriatricians questioned

whether they reflected the panel's intent and raised the possibility of misclassifications of subjects resulting from flawed indications.

Assessment of the Panel Indications. The geriatricians assessed the original indications and focused on answering the following questions:

- *Content validity*: Did the indications reflect appropriate types and levels of illness, functional impairment, medication use, and alcohol consumption based on the literature and their own clinical experience?
- *Clinical practicality*: Could the alcohol-related problems identified in the indications be addressed in a clinical setting?
- *Clinical internal consistency of classifications*: Did indications leading to the same classification include combinations of health factors and alcohol consumption intuitively representing comparable levels of risk? Did indications that shared factors with other indications lead to classifications consistent with those for the related indications?

Based on the review, we hypothesized that content validity, clinical practicality, and internal consistency standards had not been met, and that misclassifications could be due to flaws in the panel process. Guided by the survey data, study geriatricians reviewed the original panel indications and ratings qualitatively and identified specific sources of error possibly leading to misclassifications. They identified potentially flawed indications and proposed modifications. Using the proposed modifications, we prepared a formal round 3 exercise.

Revision of the Panel Indications. We chose a subset of the original panel (two internists and one geriatrician) to participate in round 3. By mail, we distributed a set of proposed modifications to the original panel indications, reasons for modifying the indications, a table showing classification of the 161 ARPS respondents based on the original and modified indications, and an honorarium. Some modifications applied to multiple indications because of the relationships between indications, and the proposed list consisted of 45 rating pages. The full set of round 2 indications was 70 pages long. We asked panelists to document any disagreements with our reasoning.

We tabulated panelist feedback on each indication according to whether the proposed change was agreed with, disagreed with, or modified further by each panelist. In cases where all three panelists did not arrive at the same conclusion, two study geriatricians weighed their responses implicitly and made final judgments by consensus. The ARPS classification algorithms were revised to reflect changes in the indications based on the round 3 panel findings.

Additional Data Collection. Panel-related tasks were funded as part of a larger effort to develop and test the ARPS as an alcohol screening tool. We designed and tested the ARPS in parallel with the three rounds of panel activities. We collected ARPS data from a total of 574 older drinkers, including the original 161 in the test sample, and called the larger sample the full sample. Figure 2 depicts the steps in our methodology.

Costs for the original two-round panel are estimated at US $50,000 (staff time for development of the indications and rating scale, preparation of reading materials and rating sheets, and conduct of the panel; panelist travel, lodging, and meals; honoraria; and ratings analysis). Costs for the survey portion, including development of the draft ARPS, writing of analysis algorithms, and test sample ($n = 161$) data collection and analysis were about US $20,000. The third-round costs (development of the indication modifications, panelist materials and honoraria, and analysis of third round results) were about US $7,500.

Table 1. Sample Composition[a]

	Test sample, % (n = 161)	Full sample, % (includes test sample) (n = 574)
Age 65–74	48	54
Age 75 and older	52	46
Male	53	55
White	85	87
Married	62	64
College educated or higher	47	51
Hypertension	51	49
Depression	16	16
Taking aspirin	47	51
Taking antihypertensives	41	42
Taking six or more medications daily	13	15
Drank three or more drinks at a sitting at least once in last year	40	40

[a] While denominators varied due to missing data, none of the values is based on fewer than 150 people in the test sample or 543 in the full sample.

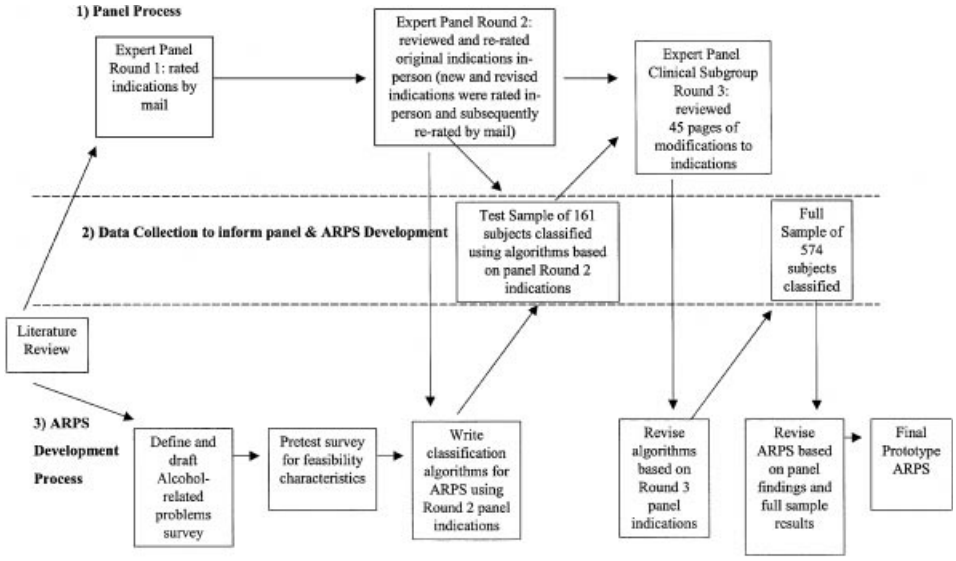

Figure 2. Panel steps and ARPS development steps.

RESULTS

Classification Results of the Test Sample

Characteristics of the test sample (n = 161) are summarized in Table 1. About half of the sample was 65 to 74 years of age (48%); the other half was 75 years or older (52%). Fifty-three percent were male. Respondents were predominantly white (85%), married (62%), and highly educated (47% were college-educated or higher). The first column of Table 2 shows the initial classification of the test sample into harmful (48%), hazardous (31%), and nonhazardous (21%) drinking categories. The combined harmful and hazardous rates were higher than expected.

https://doi.org/10.1017/S0266462301104113 Published online by Cambridge University Press

Table 2. ARPS Results Based on Initial (Panel Round 2) and Revised (Panel Round 3) Classification Systems

	Test sample, % (n = 161)		Full sample, % (includes algorithm sample) (n = 574)	
Drinking categories	Initial classifications (round 2 results)	Revised classifications (round 3 results)	Initial classifications (round 2 results)	Revised classifications (round 3 results)
Harmful	48	22	47	22
Hazardous	31	47	31	47
Nonhazardous	21	31	22	31

Assessment of the Panel Indications

Content Validity. Respondents reporting moderate amounts of alcohol consumption (e.g., two drinks per day), together with certain conditions such as hypertension, were classified as harmful drinkers. This conclusion seemed inconsistent with the literature (11;21). While two drinks per day might make medical management of hypertension more difficult, there was no obvious clinical justification for assuming damage to the respondent's health had certainly occurred, the definition of harmful drinking.

We concluded that such misclassifications stemmed from framing bias in the indication statements, resulting from the way factors were grouped. Since we could not expect panelists to rate each condition separately due to the overwhelming number of possible combinations of conditions and consumption amounts, single indications often contained multiple conditions, symptoms, or medications. Study geriatricians made grouping decisions on the basis of clinical judgment and available scientific evidence. For example, medications like flurazepam, thorazine, and narcotics were judged similar in the risks associated with similar amounts of alcohol.

In addition, alcohol consumption tables used in the rating process contained ranges of consumption patterns intended to reflect similar levels of total intake (low, moderate, or heavy). An average consumption of three drinks daily was considered heavy drinking, together with other heavy drinking patterns such as four drinks two to three times per week.

During the round 2 face-to-face panel meeting, we asked panelists to review and revise any possibly inappropriate groupings prior to providing their second set of ratings. However, the test sample results suggested that grouping errors remained. Certain conditions might have been rated as less serious for some levels of alcohol use, had they not been combined with other conditions that warranted a hazardous or harmful classification. That is, two conditions (e.g., hypertension and seizures) might be considered equally dangerous at high levels of alcohol intake (harmful drinking), but not be comparably risky at lower levels (moderate drinking might be hazardous with hypertension but still harmful with seizures). This resulted in panel ratings biased in favor of correctly classifying the most dangerous factors, leading to consistent misclassification of less dangerous symptoms, conditions, and medications in the groups. We dealt with this problem by suggesting modifications of groupings and/or alcohol consumption tables in round 3.

Clinical Practicality. Some indications contained symptoms and conditions that proved too vague to provide clinical guidance. Gastritis correctly diagnosed by a clinician is an important finding, and low to moderate alcohol consumption can exacerbate symptoms (20). Under the initial classification scheme, gastritis in combination with two drinks two to four times per month yielded a classification of harmful drinking. The test sample results revealed that despite the wording "In the past 12 months *has a doctor or other health*

professional told you that you have gastritis," this condition was reported with such high frequency that it appeared clinically useless. We modified the indication to include the presence of at least one of the following symptoms: stomach pains, nausea, vomiting, or heartburn.

To further increase clinical practicality, we omitted other vague symptoms, such as tiredness, from the indications. While fatigue can be caused or exacerbated by alcohol use, the symptom is so common that any relationship to alcohol becomes clinically impractical to assess.

Internal Consistency. Some indications were rated inconsistently with related ones. For example, if a certain amount of drinking was considered hazardous without additional factors, then such drinking in combination with any other risk factor should have a rating of at least hazardous. For our indications, this was not always true. Drinking five or more drinks once a month or less, for example, was agreed upon as representing hazardous drinking for individuals 65 years of age and older in the absence of any risk factors. The same amount of drinking for individuals who also felt depressed some of the time in the past four weeks was rated nonhazardous. This kind of inconsistency was infrequent and probably due to the complexity of the rating task and the large number of indications. We recommended modifications to classify indications with additional factors at least as severely as the same indication without those factors.

Another problem related to consistency resulted from the traditional way in which panel findings are interpreted. Historically, the panel process does not require consensus on all indications, and indications with indeterminate or disagreement ratings are omitted from use (8). In our study, omission of certain indications would have compromised clinical accuracy. For example, panelists might have agreed that drinking one drink a day is hazardous for a person taking four or more medications. If the panel could not agree on whether drinking three drinks a day with four or more medications was hazardous or harmful, the indication was labeled indeterminate. If the indication were omitted, the result would be that patients drinking more would not be classified.

This problem was identified before classification of the test sample, and we developed a convention of assigning the worst possible classification (hazardous if indeterminate between nonhazardous and hazardous; harmful if indeterminate between hazardous and harmful). The data indicated that this arbitrary rule might have contributed to unwarranted classifications of hazardous and harmful drinking. We examined the indeterminate indications individually and assigned the more serious classification only if both study clinicians agreed the classification was warranted based on content validity and internal consistency was maintained (e.g., if a patient scenario was classified as harmful with one amount of drinking, and the same scenario with a lesser amount was indeterminate between nonhazardous and hazardous, a hazardous classification would be consistent; if the greater drinking scenario were nonhazardous, a hazardous classification at the lesser drinking level would be inconsistent).

Revision of the Indications

In round 3, the panelists unanimously agreed with 17 of the 45 pages containing proposed modifications. There were two unanimous rejections, either by outright disagreement or by suggesting a different modification. For four pages, two of the three panelists disagreed with our reasoning. For 22 pages, at least one panelist suggested an additional change or was unclear about whether he or she agreed with the modification. In cases of disagreement or uncertainty, panelists made written comments that were tabulated and carefully reviewed by two study geriatricians. Final decisions were made by consensus of the study geriatricians. The indications and classification algorithms were revised to reflect the modifications.

Classification of Elderly Patients After Modification of the Indications

Additional Data Collection. Parallel with the panel process, we continued data collection in the field to evaluate the reliability and validity of the ARPS. Drinking classifications over two survey administrations in a subset of the full sample were reliable (kappa = 0.65) (unpublished data, 1999). In comparisons with two widely used diagnostic tests of alcohol abuse and dependency, the CAGE (cut down, annoyed by criticism, guilty about drinking, eye-opener drinks) and SMAST (Short Michigan Alcoholism Screening Test), the ARPS detected nearly all of the abusive and dependent drinkers identified by those instruments (ARPS for CAGE: 91%; ARPS for SMAST: 75%) (unpublished data, 1999) (24;34). The sensitivity and specificity of the ARPS compared with a criterion standard conducted by a study physician and a research assistant have been shown in a subsequent study to be 82% (29). The strength of the ARPS is its ability to detect harmful and hazardous drinkers not identified by the CAGE or SMAST. We applied the original and revised algorithms to both the original test sample of 161 and the total sample of 574 (which included the original 161). Characteristics of both samples are summarized in Table 1, and classifications are summarized in Table 2.

Classification of the Entire Sample. We applied the revised algorithms to the total sample upon completion of fieldwork and found the same proportions of classifications as for the test sample, 22% harmful (30% of males, 13% of females) and 47% hazardous (45% of males, 52% of females) classifications among elderly drinkers (Table 2). Nonhazardous classifications in both samples rose to 31%. It should be noted that about 50% of older people in community settings do not drink at all (3;27;28). Assuming an abstainer rate of 50%, the revised algorithms might be expected to identify 11% harmful and 24% hazardous drinkers in a given community outpatient population. That is, about 35% of elderly patients coming in for visits would be classified as being at risk for alcohol-related problems. A comparison of the results using round 2 and round 3 algorithms is given in Table 2.

DISCUSSION

The panel method is a valuable tool for eliciting and summarizing expert judgment and can offer important clinical guidance in the absence of rigorous studies, such as randomized controlled trials. Assessment of the implications of a panel exercise using a relevant data source can strengthen the validity of a panel's outcomes.

The case study we describe used data to identify problems created, in part, by a nontraditional use of the panel method. However, we also identified indications that were not clinically useful, and we found inconsistencies that may have been due to varying levels of medical-clinical orientation among panelists as well as panelist fatigue. We cannot presume to have improved the "truth" of the panel's findings in the sense of predictive validity; however, we believe the exercise improved the degree to which the outcomes reflected the panel's most informed clinical judgments. While the steps we propose remain primarily qualitative in nature, they can give panelists additional material upon which to base their final judgments.

Our data-driven review of the ARPS panel process led to the following useful outcomes:

- We evaluated the face validity of the proportions of nonhazardous, hazardous, and harmful drinkers using data from a test sample.
- We found unexpectedly high proportions of older drinkers in the test sample classified as hazardous and harmful drinkers and searched for possible errors in the panel method. The survey data guided us to specific flawed indications.

https://doi.org/10.1017/S0266462301104113 Published online by Cambridge University Press

- We involved panelists in a review of proposed indication modifications and determined how to deal with indications originally rated as indeterminate.
- We improved the face validity of ARPS classifications, with harmful classifications dropping from 48% to 22% and nonhazardous classifications increasing from 21% to 31%.
- We revised the indications and developed prototype algorithms for ARPS analysis.

Our method was not designed *a priori*, but rather evolved as we recognized possible problems in the validity of the proportions of harmful and hazardous drinking classifications. We recognize that it has shortcomings. We relied on the informed intuition of our study geriatricians in determining that the original percentages of high-risk classifications did not seem to have face validity. This step may be difficult to reproduce, especially for panels using the more standard appropriate/inappropriate rating scale. If data show a disproportionate number of patients to have been treated inappropriately for a given indication, isn't that exactly what the panel's judgments are meant to identify? However, it is also possible that the panel's ratings were flawed due to various sources of human error and bias previously described. It seems reasonable to take "unintuitive" findings, (e.g., unexpected numbers of inappropriate procedures in state-of-the-art facilities) and use them to make sure the indication they are based on is not identifiably flawed.

The original panel included some panelists who had a broad relevant knowledge base but were not necessarily medical clinicians. Since the round 3 review focused on clinical usefulness, only clinicians were asked to participate in this "mini-panel." In doing so, we may have traded one type of selection bias related to panel composition for another. We also relied on the expertise of our own study geriatricians to make final judgments, another source of potential bias. We attempted to remedy selection bias and framing bias in the indications by proposing modifications. These modification statements may in turn suffer from framing bias.

While the results of the round 3 panel exercise led to more clinically reasonable proportions of risk classifications, the rates may still appear high to some clinicians. The prevalence of current alcohol-related problems in older Americans has been estimated to range from 2% to 22%, with between 4% and 10% actively alcoholic (10;18;22). However, published estimates do not account for medical risks at low to moderate consumption levels. The NHANES I Epidemiologic Follow-up Study found that, of current drinkers aged 65 years or older, 77% are regular drinkers ($\geq$ one time per month) (28). When medical risks associated with low to moderate consumption levels are considered, higher percentages of risks may seem plausible. It should also be kept in mind that the risks apply only to the approximately 50% of the older American population who actually consume alcohol.

Our data-driven exercise did not solve all of the problems with our application of the panel method, but did focus thought on specific flaws in the indications. Though our methods were not designed *a priori*, we suggest that other investigators may benefit from advance planning to include more rigorous data-driven evaluations of panel results. Such evaluations might include the following steps:

- Examine face validity and clinical practicality of the implementation of panel findings using a relevant and representative clinical data source such as medical records or a survey.
- If possible, review rating outcomes for content validity using a literature review. When ratings seem inconsistent with the literature, identify affected indications and investigate the accuracy of the ratings.
- Review rating outcomes for internal consistency among indications.
- If a gold standard is available, even for a subset of indications, assess association of the panel findings with the gold standard.

- Involve the panel or a relevant subset of panelists in a third round of rating activities to formally rethink indications with identified problems and to examine the implications of unexpected clinical outcomes.

While such activities cannot assure complete accuracy, we suggest they can considerably enhance the reliability and usefulness of panel results.

POLICY IMPLICATIONS

Researchers who use the RAND/UCLA panel method, especially in nonstandard applications, should consider evaluating the impact of panel conclusions empirically in a third panel round. Journals may wish to consider empirical evaluation as a criterion for publication of the outcomes of nonstandard applications.

NOTE

[1] The full expert panel included the following panelists: Wendy L. Adams, MD, MPH, University of Nebraska, Omaha, NE; Roland Atkinson, MD, Oregon Health Sciences University, Portland, OR; Thomas Babor, PhD, MPH, The University of Connecticut Health Center, Farmington, CT; Thomas Beresford, MD, University of Colorado, Denver, CO; David Buchsbaum, MD, Aetna U.S. Health Care, Richmond, VA; Raul Caetano, MD, PhD, Alcohol Research Group, Berkeley, CA; Michael Jacobs, MD, Stanford University, Palo Alto, CA; Ernest Noble, MD, PhD, University of California, Los Angeles, CA; and Alfonso Paredes, MD, University of California, Los Angeles, CA. The panel was developed and conducted by John C. Beck, MD, chair, University of California, Los Angeles, CA; Alison Moore, MD, MPH, cochair, University of California, Los Angeles, CA; and Arlene Fink, PhD, principal investigator, Arlene Fink Associates and University of California, Los Angeles, CA. Panelists who participated in the third-round exercise to review proposed indication modifications were Wendy Adams, MD, MPH, David Buchsbaum, MD, and Michael Jacobs, MD.

REFERENCES

1. Adams WL. Interactions between alcohol and other drugs. *Int J Addict*. 1995;30:1903-1923.
2. Adams WL. Potential for adverse drug-alcohol interactions among retirement community residents. *J Am Geriatr Soc*. 1995;43:1021-1025.
3. Adams WL. Alcohol use in retirement communities. *J Am Geriatr Soc.* 1996;44:1082-1085.
4. Adams WL, Barry KL, Fleming MF. Screening for problem drinking in older primary care patients. *JAMA*. 1996;276:1964-1967.
5. Atkinson RM, Kofoed LI. Alcohol and drug abuse in older age: A clinical perspective. *Subst Alcohol Actions Misuse*. 1982;3:353-368.
6. Audet A, Greenfield S, Field M. Medical practice guidelines: Current activities and future directions. *Ann Intern Med*. 1990;113:709-714.
7. Babor TF, Grant M. From clinical research to secondary prevention: International collaboration in the development of the Alcohol Disorders Identification Test (AUDIT). *Alcohol Health and Research World*. 1989;13:371-374.
8. Brook RH. *The RAND/UCLA appropriateness method*. Rockville, MD: US Department of Health and Human Services; 1994:59-65.
9. Brook RH, Chassin MR, Fink A, et al. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care*. 1986;2:53-63.
10. Buchsbaum DG, Buchanan RG, Welsh J, Centor RM, Schnoll SH. Screening for drinking disorders in the elderly using the CAGE questionnaire. *J Am Geriatr Soc*. 1992;40:662-663.
11. Camargo CA, Rimm EB Jr. Epidemiological research on moderate alcohol consumption and blood pressure. In: Zakhari S, Wassef M (eds). *Alcohol and the cardiovascular system*. Bethesda, MD: National Institutes for Health; 1996:25-62.
12. Coulter I, Adams A, Shekelle P. Impact of varying panel membership on ratings of appropriateness in consensus panels: A comparison of a multi- and single disciplinary panel. *Health Serv Res*. 1995;30:577-591.

13. Dufour MC, Archer L, Gordis E. Alcohol and the elderly. *Clin Geriatr Med*. 1992;8:127-141.
14. Dufour MC, Fuller RK. Alcohol in the elderly. *Annu Rev Med*. 1995;46:123-132.
15. Fink A, Hays RD, Moore AA, Beck JC. Alcohol-related problems in older persons: Determinants, consequences, and screening. *Arch Intern Med*. 1996;156:1150-1156.
16. Fink A, Siu A, Brook RH, Park RE, Solomon DH. An expert panel's priorities for assuring the quality of health care given to older persons. *JAMA*. 1987;258:1905-1908.
17. Gurnack AM, ed. Special issue on drugs and the elderly: Use and misuse of drugs, medicine, alcohol and tobacco. *Int J Addict*. 1995;30.
18. Heltzer JE, Burnham A, McEvoy LT. Alcohol abuse and dependence. In: Robins LN, Regier DA, eds. Psychiatric disorders in America: The epidemiologic catchment area study. New York: Macmillan; 1991:81-115.
19. Jones A, Neri A. Age-related differences in blood ethanol parameters and subjective feelings of intoxication in healthy men. *Alcohol*. 1985;20:45-52.
20. Kelly JP, Kaufman DW, Koff RS, et al. Alcohol consumption and the risk of major upper gastrointestinal bleeding. *Am J Gastroenterol*. 1995;90:1058-1064.
21. Klatsky AL. Cardiovascular effects of alcohol. *Sci Am*. 1995;March/April:28-37.
22. Liberto JG, Oslin DW, Ruskin PE. Alcoholism in older persons: A review of the literature. *Hosp Community Psychiatry*. 1992;43:975-983.
23. Linnoila M, Erwin CW, Ramm D, Cleveland WP. Effects of age and alcohol on psychomotorperformance in men. *J Stud Alcohol*. 1980;41:488-495.
24. Mayfield D, McLeod G, Hall P. The CAGE questionnaire: Validation of a new alcoholism screening instrument. *Am J Psychiatr*. 1974;131:238-246.
25. McClellan M, Brook RH. Appropriateness of care. A comparison of global and outcome methods to set standards. *Med Care*. 1992;30:565-566.
26. Merrick NJ, Fink A, Park RE, et al. Derivation of clinical indications for carotid endarterectomy by an expert panel. *Am J Public Health*. 1987;77:187-190.
27. Molgaard CA, Nakamura CM, Stanford EP, Peddecord KM, Morton DJ. Prevalence of alcohol consumption among older people. *J Commun Health*. 1990;15:239-251.
28. Moore AA, Hays RD, Greendale GA, Damesyn M, Reuben DB. Drinking habits among older persons: Findings from the NHANES I epidemiologic follow-up study (1982–84). *J Am Geriatr Soc*. 1999;47:412-416.
29. Moore AA, Hays RD, Reuben DB, Beck JC. Using a criterion standard to validate a screening measure to identify harmful and hazardous drinking in older persons. *Aging Clin Exp Res*. 2000;12:221-227.
30. Moore AA, Morton SC, Beck JC, et al. A new paradigm for alcohol use in older persons. *Med Care*. 1999;37:165-179.
31. Naylor CD. What is appropriate care? *N Engl J Med*. 1998;338:1918-1920.
32. Oddone EZ, Samsa G, Matchar DB. Global judgments versus decision-model-facilitated judgments: Are experts internally consistent? *Med Decis Making*. 1994;14:19-26.
33. Selker HP. Criteria for adoption in practice of medical practice guidelines. *Am J Cardiol*. 1993; 71:339-341.
34. Selzer ML, Vinokur A, van Rooijen MA. A self-administered Short Michigan Alcoholism Screening Test (SMAST). *J Stud Alcohol*. 1975;36:117-126.
35. Shekelle PG, Kahan JP, Bernstein SJ, et al. The reproducibility of a method to identify the overuse and underuse of medical procedures. *N Engl J Med*. 1998;338:1888-1904.
36. Vestal RE, McGuire EA, Tobin JD, et al. Aging and ethanol metabolism in man. *Clin Pharm and Ther*. 1977;21:343-345.
37. Vogel-Sprott M, Barrett P. Age, drinking habits and the effects of alcohol. *J Stud Alcohol*. 1984;45:517-521.