Pergamon

Journal of Safety Research 34 (2003) 389–397

www.elsevier.com/locate/jsr

www.nsc.org

MaryPODS revisited: Updated crash analysis and implications for screening program implementation

Loren Staplin*, Kenneth W. Gish, Esther K. Wagner

TransAnalytics, LLC, Kulpsville, PA, USA

Abstract

Problem: Due to the relative scarcity of crashes, there has consistently been a problem with analyses that use crashes as a criterion measure in their analyses. *Method*: Previous analyses of the relationships between functional capacity measures and at-fault crash involvement for older drivers as reported in the NHTSA *Model Driver Screening and Evaluation Program* Final Technical Report have been updated to include one additional year of driving experience. Eighteen new at-fault crashes involving drivers who previously had no crash involvement were recorded for the Maryland Motor Vehicle Administration (MVA) test sample during this interval. The method of odds ratio (OR) calculation was used to examine the relationships between functional status predictors and the most salient among the safety outcome measures identified in the Maryland research. Peak valid OR values for the prior and current analyses were contrasted, and the stability of candidate pass–fail cut-points for each predictor relative to values identified in the Final Technical Report was examined. *Results*: Results indicate that the predictive value of functional tests appears to decrease over time, particularly for the perceptual–cognitive measures. *Impact on Industry*: The impact of these findings on programs and policies is to underscore a need for periodic reevaluation, spaced at the shortest practical intervals but not more than 2 years apart, in order for functional capacity screening to be applied effectively by licensing authorities, health care professionals, and others to reduce personal risk and enhance public safety.
© 2003 National Safety Council and Elsevier Science Ltd. All rights reserved.

Keywords: Driver age; Driving impairment; Crash analysis; Functional screening; Safety

1. Background and problem statement

Between November 1998 and October 1999, older drivers in Maryland performed a battery of tests to determine mental and physical functional abilities that, based on extensive literature review and expert opinion, offered the highest construct validity for discriminating at-risk drivers (Staplin, Lococo, Gish, & Decina, 2003). In a effort to seek empirical validations, 10 measures of functional performance were obtained and used as predictors in odds ratio (OR) analyses conducted for key safety criterion measures, including three categories each of moving violations and of crashes. While the strongest relationships between functional decline and safety were obtained for at-fault crash involvements, confirming a central hypothesis in the prior research, this category also offered the fewest events for analysis. Access to Maryland State Highway Administration data documenting one additional year of prospective driving experience for individuals in the Maryland Pilot Older Driver Study (MaryPODS) sample spurred the updated analyses reported in this article.

A principal result of the research in Maryland has been the broad set of recommendations developed for (nonvisual) screening and evaluation program activities implemented by licensing jurisdictions in North America. Specifically, the following domains of functional ability should be addressed: directed visual search, working memory, information processing speed under divided attention conditions, the visualization of missing information, lower limb strength and mobility, and flexibility of the head, neck, and upper torso (Staplin & Lococo, 2003). A subset of statistically significant predictor measures were identified in the MaryPODS, each representing a viable (although not necessarily exclusive) measurement technique for detecting a loss of capacity in the essential domains of functional ability listed above. These measures include, respectively, the Trail-Making test, Part B; the Delayed Recall test; the Useful Field of View

* Corresponding author. 1743 Valley Forge Road, Allentown, PA 18104, USA. Tel.: +1-215-855-5380; fax: +1-215-855-5381.

E-mail address: lstaplin@transanalytics.com (L. Staplin).

0022-4375/$ - see front matter © 2003 National Safety Council and Elsevier Science Ltd. All rights reserved.
doi:10.1016/j.jsr.2003.09.002

test, subtest 2; the Motor-Free Visual Perception test, Visual Closure subtest; the Rapid Pace Walk test; and the Head–Neck Rotation test.

The analyses upon which these research findings were determined were conducted during the fall of 2001, using driving history data for the MaryPODS test sample bracketing each individual's test date by 1 year retrospectively, and, on average, slightly under 2 years prospectively. The retrospective observation interval, which was constant across drivers, reflected a medical opinion[1] about how far in the past a person's functional ability could be reliably gauged based upon test performance. In contrast, prospective observation intervals varied: While test dates for individual drivers ranged across an 11-month period, driving history data were extracted at a single, identical point in time for the entire sample. This approach raised questions about the potential for bias, vis-à-vis the amount of driving history data available for analysis for drivers with, versus without, crash involvement.

To address this concern, the distributions of observation intervals for these two groups were compared. The mean number of (full) months after test date for which driving experience data were available for drivers involved in crashes was 20.2, with a standard deviation of 2.6; for drivers who were *not* involved in crashes, the mean number of months after test date for which driving experience data were available was 19.9, with a standard deviation of 2.9. In other words, the observation interval was 9 days longer, on average, for drivers in the crash-involved group than for drivers in the non-crash-involved group. A t test between these means was not significant ($p < .27$).

Odds ratios were calculated to assess relationships between functional status predictors and crash outcomes for the MaryPODS sample. The "baseline" as well as "updated" OR analyses reported in this article were designed to answer this question: How much more likely is it that drivers will be involved in a crash if they fail a test than if they pass the test?

The peak values for valid OR calculations in the baseline analyses were interpreted as (approximations of) the best performance possible for a given predictor, in detecting drivers at higher risk of crashing. These values can also serve to inform researchers and program administrators about candidate cut-points for pass/fail decisions for a given test, *if* viewed in the context of the overall trend in the relative proportions of crash- and non-crash-involved drivers at different levels of functional ability. The updated analyses reported here maintain the same primary focus, while offering insights about the stability of predictor–criterion relationships and highlighting cautions that must be observed whenever ORs or related techniques are applied to analyze crash risk.

[1] Dr. Robert Raleigh, Medical Advisory Board, Maryland MVA (personal communication, November 11, 2001).

2. Method

The MaryPODS database used in the baseline analyses drew upon crash and violation records for events occurring through January 5, 2001, for a sample of 1,876 licensed drivers age 55 or older. This sample was recruited and tested at offices of the Maryland Motor Vehicle Administration (MVA). Study participation was voluntary; randomly selected candidates were asked to complete the functional test battery after completing license renewals or other transactions at the MVA, with an assurance that the data obtained would be used for research purposes only and that study participation would have no impact on license status.

The generalizability of results obtained from a "volunteer" sample was a primary concern. Accordingly, the baseline analyses began with a comparison of the age characteristics and violation and crash experience of those who accepted versus those who declined participation in the study. Table 1 shows the mean, median, and standard deviation ages of these respective groups. As indicated, these values are nearly identical, and a t test confirmed that there were no significant differences ($t = 1.24$, $p < .22$).

Various categories of crash and violation experience were contrasted for the groups of drivers who accepted versus declined study participation; the most important are presented in Table 2. As shown in this table, those who accepted functional screening, although fewer in number, actually demonstrated higher event counts as a group, for the crash measures. This difference was statistically significant for the "all crash" category ($\chi^2 = 4.79$, $p < .03$). No observed differences in violation experience between groups were reliable. These comparisons provided evidence that there was no basis upon which to infer a "volunteer bias" (such that more at-risk individuals were avoiding screening) and support the position that the findings obtained for the 1,876-member study sample are representative of the older (≥ 55) driver population statewide, at a minimum.

The functional performance data included in the baseline analyses were collected by MVA personnel who received 2 days of training, followed by on-site observation to ensure that procedures were administered in a consistent and reliable fashion. All test procedures were conducted in accordance with published protocols for their administration in clinical settings, where available, subject to constraints of

Table 1
Age comparison for groups of drivers who accepted and declined to be screened

Statistic	Driver group	
	Declined screening	Accepted screening
Number of cases	2098	1876
Minimum age	55.00	55.00
Maximum age	90.00	96.00
Median age	68.00	68.00
Mean age	68.59	68.28
Standard deviation	7.95	7.92

Table 2
Event counts for groups of drivers who accepted and declined to be screened

Event type	Driver group	
	Declined screening	Accepted screening
All crashes (except alcohol related)	93	111
At-fault crashes (except alcohol related)	39	43
All moving violations	197	196

the MVA offices where data collection took place. Each MVA facility used for this purpose provided for a private, internal (no windows) testing room 3.6 × 3.6 m (12 × 12 ft) or larger in size and illuminated by overhead fluorescent fixtures. Two complete sets of all equipment and materials needed for testing were present in each room; it was therefore possible to screen two drivers concurrently, on opposite sides of the room. In addition, MVA personnel with business requiring access to the testing rooms were allowed through these spaces as needed; direct interactions with test administrators or drivers engaged in the study were not permitted, however. No data were recorded documenting the number or percentage of cases where potential distractions during screening occurred as the result of concurrent test administration or the entrance of another person into the room.

For a detailed discussion of data collection methods and distributions of performance on the included functional tests, refer to the Final Technical Report for this study (Staplin et al., 2003). Variables of the analysis procedures central to the present article are reported below, including a critical discussion of OR calculations and interpretation.

An OR is derived from four values, depending—in the present application—on whether a driver has passed or failed a test and whether he or she has or has not been involved in a crash. The values are frequency counts or numbers of observations in each of the four cells in a contingency table defined by these possibilities: (a) fail test/crash, (b) fail test/no crash, (c) pass test/crash, and (d) pass test/no crash. As the cut-point for a pass–fail decision changes, so do the values in each cell (although the row and column totals are unchanged). An OR value is calculated by dividing a/b by c/d.

The value of a predictor—in this case, failing a test of functional ability—in gauging crash risk is indicated by the magnitude of the OR, calculated as described above. A high OR can result when the number of drivers who fail a test and crash (cell a) or who pass a test and do not crash (cell d) increase relative to, respectively, the number of drivers in cells b and c. Either of these occurrences increases the numerator in the OR calculation. In traffic safety research, where the number of drivers involved in actual crashes will always (hopefully) be comparatively small, OR calculations are relatively insensitive to changes in the numbers of drivers who are *not* crash involved since the observations in the "no crash" column already comprise such a large proportion of the overall data set. Changes in the absolute number of crash-involved drivers included in an analysis and in how they are distributed between cells a and c (i.e., failing vs. passing a test) can have a very substantial impact on the OR calculation, however.

The updated analyses included records of events occurring through December 26, 2001. The additional data included a total of 103 new crashes. Of this number, 75 crashes were experienced by drivers previously included in the "no crash" group, and 28 crashes represented multiple involvements by drivers who already had at least one crash on their records. Because the criterion for the present analysis contrasts the sample proportions that are crash- versus non-crash-involved (i.e., not single- vs. multiple-crash-involved), attention was focused upon the 75 crashes experienced by the MaryPODS drivers who were previously crash-free. These events, upon closer inspection, included 1 alcohol-related crash, 38 not-at-fault crashes, 18 crashes where fault status was unknown or unassigned by the investigating officer, and 18 at-fault crashes.

A logical hypothesis in this area of research is that, to the extent loss of functional ability can be identified as the proximate cause of a motor vehicle crash, the strongest relationships between functional decline and safety should be obvious in analyses of at-fault crashes, specifically. Staplin et al. (2003) found clear evidence that this is the case. The scope of the updated analysis was therefore limited to the subset of 18 at-fault crashes occurring between January 5, 2001, and December 26, 2001, that were added to the MaryPODS database.

The 18 new at-fault crashes included in this updated analysis represent an increase of over 40% compared to the baseline MaryPODS analysis. This alone provided a strong rationale for revisiting the predictor–criterion relationships indicated in the Final Technical Report. But it is through examination of the distributions of these crashes, between drivers who would pass versus fail each test at a given cut-point, where the recalculation of OR values is most instructive. These changes are reported in the following section.

The updated analyses relating at-fault crash involvement to functional performance were conducted using SPSS SYSTAT (v. 9.01). Microsoft Access tables containing crash records through December 26, 2001, for all drivers in the test sample were received from the Maryland MVA; these were imported into SYSTAT using the Open Database Connectivity Database Capture feature. Once the data were successfully imported, SYSTAT was used as described below:

- *Filter out extraneous data.* Only drivers 55 or older were included in the present analyses. Crashes coded as "not at fault" or "fault unknown" were removed.

- *Remove outliers.* Walk times greater than 15 s were set to null. Useful Field of View, subtest 2, scores greater than 500 ms were set to 500 ms, and analyzed in this data bin. Trail-Making, Part B, scores greater than 360 s were set to null.
- *Transformations.* For consistency among measures, Delayed Recall scores were converted from number correct to number incorrect.
- *Set pass/fail cut-points for OR calculations and criteria for chi-square tests.* Cut-point pass–fail scores to use in the OR calculations were created. Criterion scores for performing chi-square tests were selected based on the highest OR obtained where *at least five* drivers failed the test and had an at-fault crash.
- *Convert safety outcomes to binary data.* At-fault crash counts, within 1 year bins, were set to binary values indicating that a crash occurred or that no crash occurred, for each driver.
- *Binning of continuous variables.* Continuous variables (Useful Field of View, subtest 2; Trail-Making, Part B; Rapid Pace Walk) were ''binned'' (segmented) to facilitate comparison among OR plots. The optimal number of bins for all measures—where observations within a bin lie within a range of scores—was found to be about 10. Any more bins than this would have resulted in too few observations per bin which, in turn, would have resulted in spuriously high or low OR values at various cutoffs (increasing the likelihood of Type I errors). Fewer bins would have diminished the ability to represent systematic changes in the OR plots with changes in cutoff and would not have faithfully represented the shape of the underlying continuous distribution.
- *Observation interval calculation.* An observation interval was calculated for each driver in the test sample as the difference (in months) between the cutoff date for the crash table (December 31, 2001, for all drivers) and the date he or she was tested.

Initially, the relative distributions of the lengths of observation intervals for crash- versus non-crash-involved drivers were plotted, showing the number of months of driving history data after test date that could be accessed for specific proportions of each group. These distributions overlapped closely. The mean number of (full) months after test date for which driving experience data were available for drivers involved in crashes was 32.0, with a standard deviation of 2.58; for drivers who were *not* involved in crashes, the mean number of months after test date for which driving experience data were available was 31.8, with a standard deviation of 2.83. A t test between these means was not significant ($p < .32$). These steps provided assurance that the amount of time—but not necessarily the actual exposure (miles driven)—during which the respective groups had the opportunity to experience a crash was not significantly different.

Due consideration at this stage of analysis was given to adjusting raw crash frequencies to reflect differences in exposure between drivers, based on self-reports of driving habits obtained as follow-up to the functional status tests at the MVA. Unfortunately, an examination of the subjective reports of weekly versus annual miles driven by the test sample revealed a substantial discrepancy—when the same drivers were asked in two different ways to report how many miles they drove, their answers differed by 50% for over 40% of the test sample. Without any means of obtaining reliable, objective measures of exposure differences between drivers in the test sample, no correction or adjustment to permit analysis of crash rates instead of raw crash frequencies could be justified.

Odds ratio calculations proceeded under the conditions and limitations noted above. All valid test data for each of the 1,876 drivers in the License Renewal sample in the MaryPODS database with a valid identification (Soundex) number were included. Because multiple measures of functional ability were collected for each driver, however, there are missing data that lead to differences in the sample size and/or number of crashes analyzed in each OR calculation.

Four measures of perceptual–cognitive ability and two measures of physical ability were of interest in the updated analyses (i.e., those that reached statistical significance in the baseline analyses). These included the Motor-Free

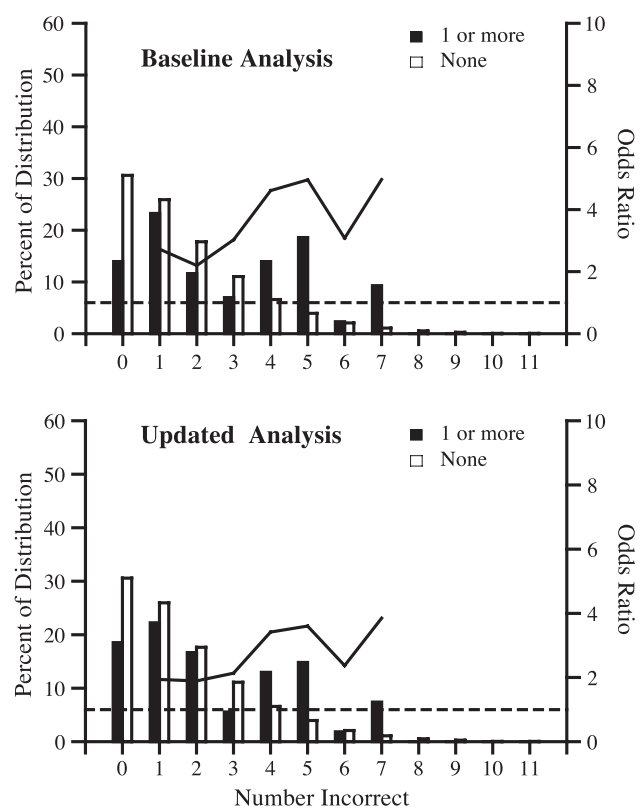

Fig. 1. Distributions of at-fault-crash and non-crash-involved drivers at indicated performance levels on Motor-Free Visual Perception test, Visual Closure subtest, with corresponding ORs.

Visual Perception test, Visual Closure subtest; the Trail-Making test, Part B; the Useful Field of View test, subtest 2; the Delayed Recall test; the Rapid Pace Walk test; and the Head–Neck Rotation test. After the peak valid OR value was determined for each of these measures (by stringently adhering to the requirement for a minimum of five observations per cell in the OR contingency table), a chi-square test was performed to evaluate the level of significance associated with that analysis outcome.

3. Results

The results of the updated analyses are displayed in the form of combined OR plots that permit direct comparison with the baseline data presented in the Final Technical Report. This approach contrasts the peak OR values for each predictor, and graphically illustrates changes in the proportions of crash- (black bars) versus non-crash-involved drivers (white bars) as the cut-point for pass–fail decisions is varied. The dashed line in each plot marks an OR value of 1.0; this result indicates that a driver is equally likely to be crash-involved when passing as when failing a test (i.e., the test has no predictive value in the present context). Results for the four measures of perceptual–cognitive abilities are presented first, followed by the two measures of physical ability.

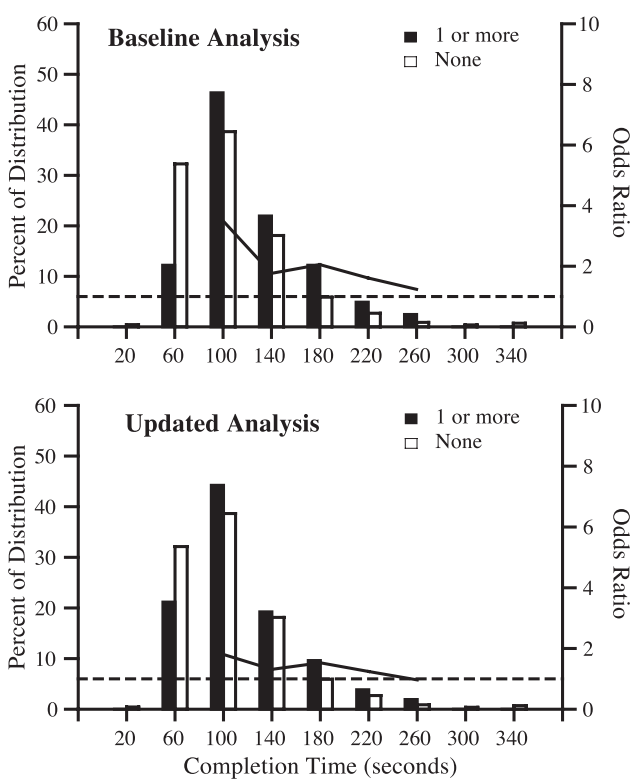

Fig. 2. Distributions of at-fault-crash and non-crash-involved drivers at indicated performance levels on Trail-Making test, Part B, with corresponding ORs.

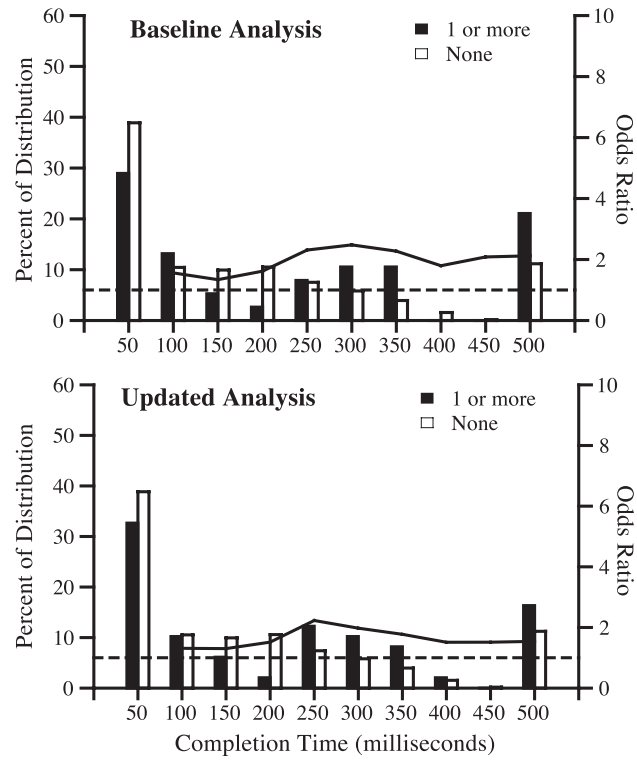

Fig. 3. Distributions of at-fault-crash and non-crash-involved drivers at indicated performance levels on the Useful Field of View, subtest 2, with corresponding ORs.

Results of the OR calculations for the Motor-Free Visual Perception test, Visual Closure subtest are displayed in Fig. 1. Declining functional ability is indicated by an increasing number of incorrect responses, moving to the right along the *x*-axis. The peak valid OR in the updated analysis is obtained at a cut-point of five incorrect responses, the same as in the baseline analysis; however, the magnitude of this value has decreased from 4.96 to 3.60. The chi-square test is significant at this level of performance in the updated analysis, $\chi^2(1) = 17.15$, $p < .001$. It is also interesting to note that, in both plots, the proportion of drivers who are crash-involved begins to exceed the proportion who are crash-free at the same level of functional performance, four incorrect responses. Finally, it may be observed that the distributions of crash-involved drivers appears distinctly bimodal, while the percentages of non-crashing drivers fall off in a linear fashion with declining functional ability.

Next, Fig. 2 shows the results of the OR calculations for the Trail-Making test, Part B. As in the baseline analysis, the peak valid OR is obtained at a completion time of 100 s in the updated analysis; it should be remembered, however, that this is a continuous variable and that 100 s is in fact the midpoint of a data "bin" ranging from 80 to 120 s. The obtained value at this level of performance, 1.80, is markedly lower than the 3.50 resulting from the analysis a year earlier, and this predictor now demonstrates only marginal significance, $\chi^2(1) = 3.02$, $p < .08$.

In Fig. 3, the results of OR calculations for the Useful Field of View, subtest 2, can be compared for the baseline versus updated analyses. As shown, the peak valid OR has fallen from 2.48 to 2.23, and the cut-point at which the peak value is obtained has shifted left along the *x*-axis from a target duration of 300 ms to a target duration of 250 ms. Again, these are binned data for a continuous variable, so the 250-ms value represents the midpoint of a bounded interval beginning at 225 ms and ranging up to 275 ms. This predictor remains significant in the updated analysis, $\chi^2(1) = 7.99$, $p < .005$. In addition, with the exception of the data bars at the far right of the plot (representing a data bin containing all of the outliers requiring target durations in excess of 500 ms to respond on this test), the distribution of crash-involved drivers suggests a bimodal pattern, as observed in the earlier measure.

Fig. 4 displays the results for the Delayed Recall measure. Alone among the measures of perceptual–cognitive abilities, the peak valid OR for this test increased substantially—from 2.92 to 3.34—between the time of the baseline and the updated analysis. As such, this result was again statistically significant, $\chi^2(1) = 9.15$, $p < .002$. In addition, the peak OR was obtained at the same performance level, three incorrect responses, in the latter as in the former analysis. The graph described by the calculated OR values for Delayed Recall shows a prominent linear trend; this may be related to the fact that it is a discrete variable with only four possible outcomes.

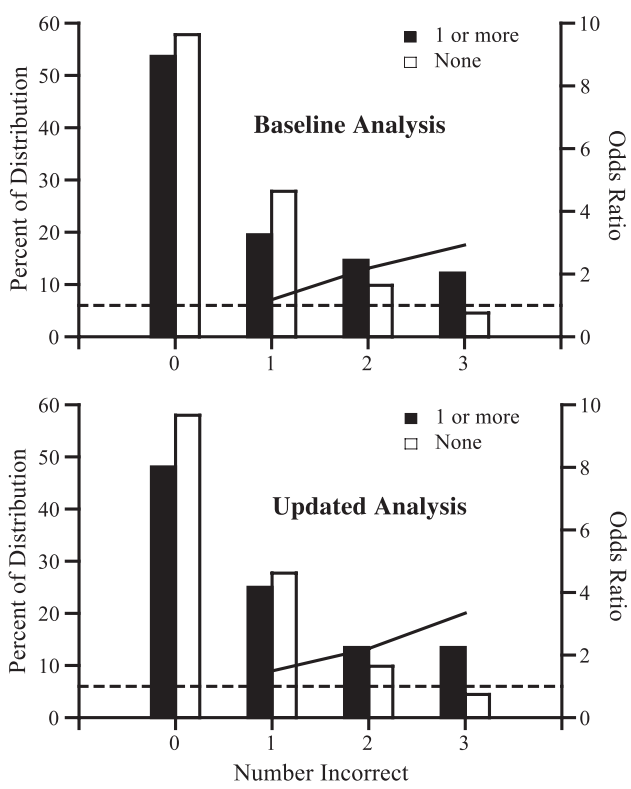

Fig. 4. Distributions of at-fault-crash and non-crash-involved drivers at indicated performance levels on the Delayed Recall test, with corresponding ORs.

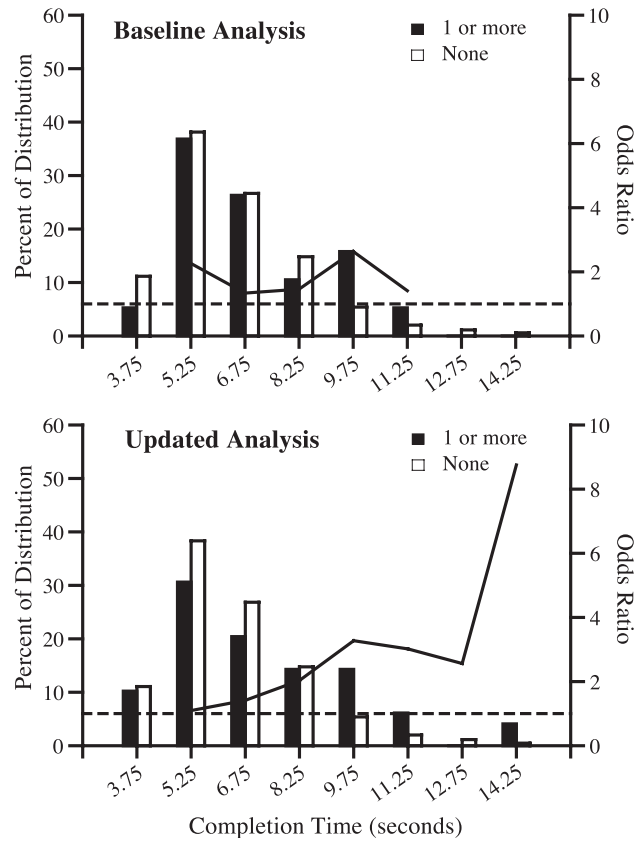

Fig. 5. Distributions of at-fault-crash involved and non-crash-involved drivers at indicated performance levels on the Rapid Pace Walk test, with corresponding ORs.

Results for the first of two measures of physical ability, the Rapid Pace Walk test, are presented in Fig. 5. As a continuous variable, performance data are binned, in this case using 1.5-s intervals. The plot for the updated analysis shows an extraordinarily high peak at the slowest time (signifying the frailest drivers) for this test; there were too few observations in the OR contingency table cell corresponding to drivers who failed the test at this cutoff and were involved in an at-fault crash to permit a valid analysis, however. The peak valid OR obtained in the updated analysis was 3.23, an increase over the 2.64 demonstrated in the baseline analysis. This result was statistically significant, $\chi^2(1) = 13.32$, $p < .001$. In addition, the performance cutoff at which the peak valid OR was calculated remained stable across the two sets of analyses, namely, the walk time data bin beginning at 9.0 s.

The final measure of functional capacity examined in these analyses was the Head–Neck Rotation test. As a binary measure, no choice of cutoffs exists and a plot of OR values is superfluous. Roughly one-third of drivers who were involved in an at-fault crash failed this test, whereas only about half of this proportion (one-sixth) of all drivers tested lacked the necessary flexibility in the neck and upper

Table 3
Summary of changes in peak valid ORs and cut-points for significant predictors in the MaryPODS, with one additional year of driving experience

Functional status measure	Baseline OR[a]	Updated OR[b]	Cut-point change?
Motor-Free Visual Perception Test, Visual Closure subtest	4.96	3.60	no
Trail-Making, Part B	3.50	1.80	no
Delayed Recall	2.92	3.34	no
Useful Field of View, subtest 2	2.48	2.23	yes
Rapid Pace Walk	2.64	3.23	no
Head–Neck Rotation	2.56	2.01	NA[c]

[a] Peak valid ORs reflect 1 year of retrospective crash data plus a mean prospective observation interval of 20 months.
[b] Peak valid ORs reflect 1 year of retrospective crash data plus a mean prospective observation interval of 32 months.
[c] NA = not applicable, binary measure.

torso. The calculated OR of 2.01 in the updated analysis was down from the value of 2.56 obtained in the baseline analysis, however, and this result only approached significance at $p < .08$, $\chi^2(1) = 3.05$.

4. Summary

The comparative results for the peak valid OR calculations in the baseline and updated analyses are summarized in Table 3. The predictive value of each measure of functional ability is indicated, in a general sense, by these results, which convey how much more likely a driver who fails a given test is to cause a crash than a driver who passes the test. An indicator of the stability of the candidate cutoff values, based on whether the peak valid OR for each test was obtained at the same performance level in the updated analysis versus the baseline analysis, also appears in Table 3.

It may be concluded from these results that measured declines in key functional abilities, as detected by the specific set of testing methods providing data for these analyses, in many cases hold diminishing value for discriminating high-risk drivers as the length of the observation period extends beyond 2 years after testing. Results for one perceptual–cognitive measure, Delayed Recall, and one measure of physical frailty, Rapid Pace Walk, ran counter to this trend.

Preliminary inspection of the number of months of driving history data available for analysis for the crash- versus non-crash-involved drivers in the MaryPODS test sample indicated that there was not a significant difference between these groups. Serious reservations about the reliability of self-reported estimates of miles driven by members of the test sample precluded adjustments of raw crash frequencies to account for exposure differences, if any, that may have existed between drivers with and without (at-fault) crashes.

5. Discussion

The analyses reported in this article reinforce the findings of the MaryPODS, which has provided perhaps the best evidence to date that functional capacity screening, conducted quickly and efficiently, in diverse settings, can yield scientifically valid predictions about the risk of driving impairment experienced by older individuals—while at the same time demonstrating its limitations. With one additional year of driving history data available for the MaryPODS test sample in the updated crash analyses, the predictive value of three out of four tests of perceptual–cognitive functional ability has diminished, relative to levels reported in the NHTSA Final Technical Report (Staplin et al., 2003).

Further limitations in interpreting both the baseline and updated analysis results that relate to the choice of analysis techniques and methods may be identified. First, questions may arise about the extent to which the indicated relationships between functional ability predictors and the at-fault crash criterion depend upon the application of OR calculations, versus other available techniques. Epidemiological literature suggests *relative risk* (RR) as a viable alternative approach. When predicting rare events such as crashes, however, OR and RR analyses yields equivalent results (http://www.cche.net/usersguides/overview.asp, 2003).

A potentially greater concern is that false findings may have resulted from the manner in which performance on the continuous measures of functional ability was "binned" for the OR calculations. That is, results reported for the continuous variables in both the baseline and the updated analyses arguably have been influenced simply by the choice of bin size. The finer the bin size, the more precisely a calculated OR can be linked to a specific level of performance on a functional test. Unfortunately, increasing bin size reduces the number of observations per bin, which compromises the validity of the analysis technique. In addition, with more data bins, the apparent level of significance is inflated and the chances of a Type 1 error increase. The choice of how many bins to use to sort and analyze the data for the Useful Field of View, subtest 2; Trail-Making, Part B; and Rapid Pace Walk measures of functional ability attempted to balance these considerations.

Finally, there will always remain some uncertainty about the quality of the crash data itself. A crash typically must involve a tow-away to be reported and stored in the database of the Maryland State Highway Administration,[2] a guideline

[2] Mr. Jack Joyce, Office of Driver Safety Research, Maryland Motor Vehicle Administration (personal communication, January 8, 2002).

that is well-defined and may be applied consistently. However, differences in judgment governing the assignment of fault in a crash may exist both within and between investigating officers. The primary concern in this regard is the underreporting of at-fault crashes for older persons. To the extent that a bias toward leniency in citing an older person for traffic violations may be reflected in these data, peak OR values that were rejected as invalid in this analysis because of low cell counts (i.e., too few drivers who crashed and failed a test) might be ''legitimatized.'' Most affected among the battery of measures addressed here would be the results for the Rapid Pace Walk test, a measure of lower limb strength and mobility and the best indicator of overall physical frailty.

Despite the limitations noted above, the evidence emerging from the MaryPODS helps address a widely recognized need for improved practices in this area. The effectiveness of tests that are currently conducted in most DMVs (visual acuity, sign knowledge) to identify those at increased risk of crashes has been widely questioned, leading some states, in particular California, to consider novel ''multitiered'' approaches (Janke, 2001). The central idea in California's approach is to identify potentially at-risk drivers who need closer scrutiny through office-based tests of performance that can be administered quickly, cheaply, and reliably using standardized procedures. Screening results thus serve as a trigger for further testing and do not lead directly to a licensing decision.

Implicit in a multitiered system is the need for multiple cut-points for the functional tests used as predictor variables—one at a level of mild impairment to trigger education and, hopefully, to initiate remediation actions before more serious decline occurs, and another to trigger in-depth diagnostic testing when a more profound deficit is revealed. Such a system can limit negative response from the driving public due to false positives, because the vast majority of DMV customers who ''fail'' a screen oriented to prevention will be unaffected in terms of their driving privileges. Analyses carried out in the MaryPODS also suggest that a $5 cost-per-driver to administer in-office screening is attainable (Staplin et al., 2003).

In conclusion, functional abilities screening can be seen as part of a prevention model involving not only DMVs, but the broader medical community. Screening tools detect the loss of functional ability; they do not diagnose the underlying problem. For that, medical professionals are needed, particularly with regard to directing individuals to appropriate treatments. Guidelines recently developed by the American Medical Association (2003) highlight the responsibility of physicians to promote safe mobility as a basic aspect of maintaining and protecting their patients' health and well-being and support the use of tools similar to the screening measures employed in this research. Applying functional capacity screening tools to promote ''driving health'' will become paramount in the years ahead, as the dramatic increase in the numbers of older persons in this country will mark a similar increase in the numbers of functionally impaired drivers who experience a significantly higher risk of crashing.

Acknowledgements

The authors of this article acknowledge Robert Raleigh, MD, Chief, Maryland Medical Advisory Board, Maryland Motor Vehicle Administration, for his leadership of the Maryland Pilot Older Driver Study, and Mr. Jack Joyce, Esq., Office of Driver Safety Research, Maryland Motor Vehicle Administration, for facilitating the development of the MaryPODS database examined herein. More broadly, we acknowledge John Eberhard, PhD, of the National Highway Traffic Safety Administration (retired), for his role in bringing the issue of lifelong safe mobility to the forefront of the national research agenda.

References

American Medical Association (2003). *Physician's guide to assessing and counseling older drivers*. Chicago: Author.

Janke, M. (2001). *Medical conditions and other factors in driver risk* (Report No. CAL-DMV-RSS-01-190). Sacramento, CA: California Department of Motor Vehicles.

Staplin, L., & Lococo, K. (2003). *Model driver screening and evaluation program: Guidelines for motor vehicle administrators* (Report No. DOT HS 809 581). Washington, DC: National Highway Traffic Safety Administration.

Staplin, L., Lococo, K., Gish, K., & Decina, L. (2003). *Model driver screening and evaluation program final technical report: Vol. 2. Maryland Pilot Older Driver Study* (Report No. DOT HS 809 583). Washington, DC: National Highway Traffic Safety Administration.

Centre for Health Evidence—Users' Guides to Evidence-Based Practice (2003)Available at: http://www.cche.net/usersguides/overview.asp (Reproduced with permission from *JAMA, Journal of the American Medical Association, 272*(17) (1994), 1367–1371).

Loren Staplin, PhD, is the founder and principal partner of the consulting firm *TransAnalytics, LLC*. His primary focus in recent years has been upon the aging of American society and its implications for transportation policy and practice in the early 21st century. Significant products of work Dr. Staplin has led in this area include the *Safe Mobility for Older People Notebook*, available on the website of the National Highway Traffic Safety Administration; the *Highway Design Handbook for Older Drivers and Pedestrians*, published by the Federal Highway Administration in 2001 and also available online at the FHWA website; and the *Model Driver Screening and Evaluation Program: Guidelines for Motor Vehicle Administration*, a 2003 publication of NHTSA in conjunction with the American Association of Motor Vehicle Administrators. Overall, Dr. Staplin has published 50 articles and technical reports addressing diverse aspects of traffic safety, including two chapters in the pending National Academy of Sciences/Transportation Research Board publication *Transportation in an Aging Society: A Decade of Experience*. Dr. Staplin is a member of long standing of the Transportation Research Board and the Human Factors and Ergonomics Society; and, he serves as a founding member of the newly formed (2002) Human Factors Resources advisory group to the National Committee on Uniform Traffic Control Devices.

Kenneth W. Gish, PhD, is a senior associate at *TransAnalytics, LLC*. He holds a doctoral degree in Experimental Psychology from the Pennsylvania State University. Dr. Gish joined *TransAnalytics* in 2001, following 14 years of experience as a human factors research psychologist in academic, military, and private sector settings. He led the analysis for the NHTSA-sponsored research project, ''Model Driver Screening and Evaluation Program,'' which culminated in the MaryPODS data set. He also has served as statistician guiding major crash database analysis efforts in an FMCSA-sponsored study of the relationship between driver retention and safety. Recently completed NHTSA projects for which Dr. Gish has served as the principal investigator include an in-vehicle observational study of driver distractions and human factors studies evaluating night vision enhancement technology, forward collision warning signals, and head-up displays in automobiles.

Essie Wagner, is a Program Analyst working with NHTSA's Safety Countermeasures Division. There, she works on older driver safety programs and pedestrian safety programs. Ms. Wagner has been the program manager for many studies on older drivers, including the one described in this paper. She received her BA in psychology from the College of Wooster in Ohio and went on to get an MA in Applied Psychology (Human Factors) from George Mason University in Fairfax, VA. Before joining NHTSA in 1998, she worked as a contractor for FHWA's Research and Development division. Her seven years of experience there included conducting and monitoring extensive research on older driver issues related to the roadway environment, such as signs and pavement markings, and responses to emergency driving events.