A National Role Delineation Study
of
FMCSA Medical Examiners
August 2007

Conducted for the
U.S. Department of Transportation
Federal Motor Carrier Safety Administration

Prepared by

Axiom Resource Management, Inc.

Axiom
RESOURCE MANAGEMENT, INC.

The National Registry: A Roadmap to Improved Highway Safety

EXECUTIVE SUMMARY

In passing the 2005 Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU), Congress required the Secretary of the United States Department of Transportation (DOT) to establish a national registry of medical examiners. The purpose was to improve highway safety by establishing and maintaining a national registry of medical examiners qualified to examine drivers of commercial motor vehicles (CMVs) and issue medical certificates.

To that end, the Federal Motor Carrier Safety Administration (FMCSA) began developing the National Registry of Certified Medical Examiners (NRCME) program to improve highway safety by producing trained, certified medical examiners who can effectively determine whether a CMV driver's health meets FMCSA standards. In order to realize a sound and representative testing and registry program, FMCSA is developing the NRCME with a focus on those elements of competent job performance common to medical examiners coming from a variety of professional backgrounds and work settings. Therefore, in order to 1) develop a blueprint for the medical examiner certification test and 2) capture demographics for a hidden population sample, FMCSA commissioned the following Role Delineation Study in 2005.

A role delineation study is intended to describe as much as is known about a set of competencies associated with a role in a work setting. In this case, a medical examiner qualifies or disqualifies a CMV driver based on an assessment of the driver's health status. Therefore, researchers examined the role these practitioners played in the driver qualification process. Because this study addressed only a subset of a typical practitioner's job, it is described as a "role delineation study" rather than the more wide-ranging "job analysis."

The following Role Delineation Study was conducted from 2005 to 2007. It was the first national study of medical examiners who conduct CMV driver physical examinations. Axiom Resource Management, Inc. (Axiom) was the primary coordinator of project management activities related to this study. Applied Measurement Professionals, Inc. (AMP) provided research expertise on the project. FMCSA staff was at all times an integral part of study development and implementation.

The Role Delineation Study included the following steps:
- Literature Review and Direct Observations
- Subject Matter Expert Recruitment
- Task List and Survey Development
- Sample Selection and Survey Implementation
- Sampling Bias, Demographic, and Task Analyses
- Detailed Content Outline Development
- Knowledge, Skills, and Abilities Statement Development

Literature Review and Direct Observations. AMP research staff began work on this study by conducting a systematic review of the available relevant literature. Multiple Federal statutes, proposals, reports, and standard and process summaries offered guidance about the approaches and outcomes associated with CMV driver physical examinations. Research staff also reviewed textbooks authored by physicians experienced in applying Federal regulations to the physical certification of CMV drivers.

The next step was to conduct direct observations of medical examiners as they worked in the field. In August and September of 2005, a member of the AMP staff observed a small group of medical examiners as they completed CMV driver physical examinations. The observer witnessed all aspects of the physical examination process when possible, including any preliminary medical testing.

Recruitment of Subject Matter Experts. At this juncture, a major imperative was convening panels of medical examiner experts to assist in the design, implementation, and analysis phases. FMCSA, in conjunction with Axiom, selected a Working Integrated Product Team (WIPT) of medical examiners to serve as subject matter experts. The medical examiner population nationwide includes five known primary groups: (1) advanced practice nurses, (2) chiropractors, (3) doctors of osteopathy, (4) medical doctors, and (5) physician assistants. Staff from FMCSA, Axiom, and AMP met with medical examiners representing all five of the aforementioned medical professions during three regional meetings held in three cities: Chicago, Illinois; Falls Church, Virginia; and San Francisco, California.

FMCSA selected two representatives from each of the five professions for participation in the WIPT, forming a group of 10 members. FMCSA chose WIPT members to represent a variety of practice settings and geographic regions across the United States. Staff from FMCSA, Axiom, and AMP met with the WIPT to discuss and reach consensus about specific study methodology.

Task List and Survey Development. Researchers can use a variety of acceptable methods for a role delineation study, but a task inventory is the most common (Raymond, 2001). AMP staff synthesized data gleaned from the observations and the literature review, creating a preliminary task list of job competencies that would serve as the basis for the survey of medical examiners. During the medical examiner regional meetings, attendees iteratively revised this preliminary task list through brainstorming sessions. Then, using recommendations from these three brainstorming sessions, the WIPT developed the final set of task statements and background questions for a survey to measure the opinions of medical examiners about the criticality of each task.

A task inventory permits flexibility in how specific work behaviors are described. The inventory must be comprehensive, but expressing tasks with greater or lesser specificity permits some control over the length of the list. Since survey response rates tend to decline as survey length increases, choosing to develop tasks in a more general manner gave the WIPT an opportunity to encourage a high response rate by minimizing task list length. When the WIPT finalized the task list for the survey, it contained a total of 146 tasks.

Another advantage of a task inventory is that it yields objective information about each task's criticality for competent practice. In this case, survey respondents ultimately used a rating scale to interact with the task list. The scale permitted each respondent to indicate whether a task was a part of his or her practice; if it was, the respondent was directed to rate the importance of the task. While the task inventory method helps to remove subjectivity from the process, the goal is not to remove subjectivity completely.

This study coupled the collective opinions of WIPT members with objective data to yield content that thoroughly covered the competencies that were necessary to enhance safety but could still be defended as fair for the certification process.

Sample Selection and Survey Implementation. Since medical examiners have not been unified by a specific licensing or professional organization, they are considered a hidden population, one more difficult to reach and study given the absence of a sampling frame. Attempts to study hidden populations using standard sampling and estimation techniques often generate results that contain bias. Therefore, the use of alternative sampling methods is prudent.

One alternative approach that has been used to sample from a hidden population comes from network theory, where it is commonly referred to as respondent-driven sampling or snowball sampling. In snowball sampling, researchers identify individuals in their population of interest and ask these individuals to recommend others of the same population for participation in the study. Using this approach, Axiom staff contacted professional groups that had medical examiners among their members, recruiting an initial base of potential participants. From that foundation, Axiom spread word of the study to a wider range of medical examiners of CMV drivers, using a substantial word-of-mouth campaign, direct mailings, and FMCSA Web site notices.

Given the length of the task list and the time associated with thoughtful responding, the combined team of FMCSA, Axiom, and AMP personnel decided to use a paper survey format to encourage portability and ease of completion for respondents. Ultimately, AMP distributed surveys by mail to 4,082 FMCSA medical examiners identified by Axiom using the recruitment approaches described above. Each medical examiner who received a survey was aware of the proposal for a national registry and had actively agreed prior to the mailing to participate in the study. Therefore, the sample was characterized as a group that had opted into study participation.

Throughout the study, the WIPT and study staff took several steps to encourage a high survey response rate. First, the WIPT opted for a more general list of tasks rather than a longer, more detailed list; the latter required substantially more respondent burden in the form of time and effort. Starting with an opt-in sample also encouraged a high rate of response, ensuring that surveys were sent only to individuals who had expressed an interest in participating. Prior to the survey mailing, study personnel sent a postcard to each volunteer participant to alert him or her to expect the survey. Another measure designed to increase response rate was a follow-up letter that encouraged recipients to complete the survey and directed them to request replacements if they had misplaced their originals. This follow-up letter served a second purpose, that of permitting recipients to opt out of full survey participation by responding to a short survey of demographic questions. This opt-out sample permitted a follow-up study of non-response bias. Axiom staff sent weekly email updates to the entire sample, encouraging them to respond to the survey or to request a replacement survey if needed. Finally, study staff extended the deadline for returning surveys by 2 weeks to permit willing participants more time to comply.

A volunteer sample of 2,297 chose to submit full survey responses. However, 10 individuals contacted AMP and indicated they either were not qualified or were no longer interested in completing the survey. Another 22 surveys were returned without responses because of inaccurate addresses. Therefore, the corrected return rate was

56.7%. When the WIPT reconvened to assess survey responses, AMP staff informed them that this response rate was much higher than is typically observed for other studies of this type; given this outcome, the methods used to encourage responding appeared to have had the intended effect. An additional 891 responses were collected from the follow-up survey, which contained only demographic questions. Including results obtained from both surveys, a total of 3,188 responses (2,297 from the full survey + 891 from the follow-up survey) were obtained from the sample. The total combined response rate for both surveys was 78.7%. However, because survey responses were anonymous, it is possible that some medical examiners completed full- and follow-up surveys, so the combined response rate may be an overestimate.

Sampling Bias Analyses. As indicated, medical examiners who qualify CMV drivers are not conveniently listed for reference in sample surveys. Because study sample members are therefore not selected from a sampling frame (a list of the population from which the sample is drawn), snowball samples are subject to numerous biases. To correct for sampling bias in this study, study staff conducted network analyses as a way of assessing the impact of any potential sampling bias.

Data to assess this effect were collected from the following survey question: "How many people in each of the following groups do you know who also perform CMV physical examinations?" Respondents in the sample were broken down into the five aforementioned subgroups of advanced practice nurses, doctors of medicine and osteopathy, physician assistants, and chiropractors. Linear models were used to adjust the results. Although the network analyses showed some bias, they also showed that this was not disruptive and did not change the overall outcome of the survey results. Adjustment for this effect did not, in any instance, diminish the importance of any task. Therefore, the following results of the survey stand.

Demographic Analyses. Responses to questions in the background information section of the survey helped describe the characteristics of the sample, a particularly important study element given that sample members are part of a hidden population.

Analyses of demographic data produced the first insights into sample characteristics, including the following:
- Respondents had an average of 17.3 years of experience in their current professions.
- They completed a mean of 43.5 FMCSA physical examinations monthly.
- Respondents had been conducting FMCSA CMV driver physical examinations for an average of 12.1 years.
- Approximately two-thirds (66.7%) of respondents reported having had training in occupational health.
- Few (27.7%) had attended training courses for CMV driver physical examinations, supporting the need for training in this area.
- The majority of respondents (95.3%) reported clinical practice as their primary job function.
- Approximately one-half of respondents reported primarily working in occupational health.

- There was roughly equal representation of urban, suburban, and rural communities in the sample.
- Nearly two-thirds (63.7%) of respondents were male.
- The majority of respondents (88.7%) identified their racial and ethnic background as white, non-Hispanic.

WIPT members reviewed responses to all of the survey background questions and indicated that, given their experience, the sample was sufficiently representative to serve as the basis for certification test development.

Task Analyses. Studies of this kind should yield a thorough description of the role in question. It should be clear at the study's conclusion that no content critical to a complete role description was omitted. The study prompted respondents to assess whether the role was adequately covered by the task list. Nearly every respondent (95.8%) found the task list to be adequate, supporting the assertion that no critical content was omitted.

If ratings are to be useful for identifying content that is critical to certification tasks, they must contain a minimum of observed error. Minimal error is associated with high reliability. The WIPT reviewed two types of reliability statistics associated with task ratings: intraclass correlations and coefficient alphas.

The WIPT first reviewed intraclass correlation values for each content domain. During a standard role delineation study, items are typically allocated to content domains after the task quality assessment. These content domains represent the overarching categories within which specific tasks are grouped. The intraclass correlation values for these domains reflected the likelihood that other samples of medical examiners would give the same ratings provided by the medical examiner sample in this study. Intraclass reliability values were very close to the maximum possible value, supporting WIPT members' confidence in using task ratings from this sample.

WIPT members assessed a second reliability coefficient, coefficient alpha, to determine whether task ratings were consistent within each content domain in the task list. Because coefficient alpha values for content domains were also very high, the WIPT had confidence in the outlined organization of tasks; these high coefficient alpha values further supported their confidence in the final set of task ratings.

In addition to evaluating the adequacy of the task list and the reliability of the task ratings, the WIPT evaluated the aforementioned demographic characteristics of respondents in order to understand the target audience of medical examiners. WIPT members were interested in learning whether respondent characteristics were consistent with those of the largely unknown population of medical examiners, so that they could be confident using respondent task ratings. Such confidence was crucial, since a potential use of the ratings would be to determine which non-critical tasks to exclude from the list.

After concluding that the typical respondent seemed well versed in conducting physical examinations of CMV drivers and that the sample included representatives of important population subgroups, the WIPT systematically established rigorous task exclusion rules, the next step in the Role Delineation Study process. First, given that every task was performed by at least two-thirds of the study sample, the WIPT concluded that no tasks should be excluded on the basis of performance frequency. Next, the WIPT

considered potential thresholds in relation to task importance ratings. All tasks received a mean rating of at least "above average"; therefore, they retained all under the second exclusion rule. Finally, the group focused on whether tasks that were critical for the whole sample also were critical for subgroups within the sample. This resulted in a 13-hurdle approach in which surviving tasks would be labeled "critical." Each task was subjected to this 13-step process and, again, the group retained all tasks. Applying all exclusion rules ultimately resulted in the retention of all 146 tasks, indicating that all could become a basis for program competencies defensible as fair for medical examiners who perform physical certifications of CMV drivers. WIPT members also kept the original outline under which they had organized these tasks.

Detailed Content Outline Development. Once the WIPT had identified the competencies the test should cover, the group began test specification development. Test specifications describe the distribution of items on a test form by content domain and complexity level. Assigning cognitive levels to tasks is the accepted manner in which complexity level is established. To assess the potential complexities of each task, the WIPT decided to use the three-level cognitive dimensions of recall, application, and analysis as defined by AMP staff. By limiting the cognition of test items associated with simple tasks and considering cognition when specifying items for the test, test scores are more likely to align with competence. A full consensus of WIPT members assigned a cognitive complexity level to each of the 146 tasks. Next, the WIPT specified the number of items for each content domain and for each cognitive level within each domain. When doing so, they returned to decisions they had made about task complexity to help decide how best to distribute items within content domains. The result was a document called the Detailed Content Outline. It summarized: (1) tasks that could be tested on a certification test, (2) the complexity of test items linked to each task, (3) the number of items by content domain, and (4) the number of items by cognitive level within each domain. The outline will serve for several years as a stimulus for test items and as a blueprint from which certification test forms will be developed.

Knowledge, Skills, and Abilities Statement Development. Because the legislative mandate for a national registry included a training component, as their final activity the WIPT developed statements that described the knowledge, skills, and abilities (KSAs) associated with each task in the Detailed Content Outline. Task statements purposefully describe the behaviors that are observable in medical examiners as they conduct physical examinations of CMV drivers. These statements do not directly identify the KSAs on which examiners draw or that support competent task performance.

The WIPT produced KSA statements to ensure this supplemental content was identified. Training and testing content are expected to align closely. Therefore, the starting point for each KSA statement was a task from the Detailed Content Outline. First, Axiom and FMCSA staff identified knowledge statements that supported the competent performance of each task. The WIPT then verified these knowledge statements. Then, as appropriate, the WIPT developed skill and ability statements to describe the underlying attributes on which competent task performance also relied. In this way, KSAs were tied directly to the Detailed Content Outline, which should encourage standardized training of medical examiners.

Summary of Findings.

Sampling bias analysis:
- Network analyses were conducted as a way of assessing the impact of any potential sampling bias. Although the results of these analyses indicated some bias, they also showed that it was not disruptive and did not change the overall survey results. Adjustment for this effect did not, in any instance, diminish the importance of any task. Therefore, the results of the survey stand.

Demographic analysis:
- Respondents were experienced professionals. Not only did they have an average of 17.3 years of experience in their current professions, they completed a mean of 43.5 FMCSA physical examinations monthly and reported having performed FMCSA CMV driver physical examinations for an average of 12.1 years.
- However, while approximately two-thirds (66.7%) of respondents reported having had training in occupational health, few (27.7%) had attended training courses for CMV driver physical examinations, supporting the need for training in this area.
- Respondents were members of the population of interest. The vast majority of respondents (95.3%) reported clinical practice as their primary job function, and about one-half of respondents reported primarily working in occupational health.
- There was roughly equal representation of urban, suburban, and rural communities in the sample, an indication that medical examiners are accessible across all community sizes and locales.
- Nearly two-thirds (63.7%) of respondents were male; the majority of respondents (88.7%) identified their racial and ethnic background as white, non-Hispanic. WIPT members indicated that these and the other demographic findings were reflective of their anecdotal experience with colleagues.

Task analysis:
- Nearly every respondent (95.8%) found the task list to be adequate, supporting the assertion that all critical content was captured.
- Intraclass reliability values were very close to the maximum possible value, supporting WIPT members' confidence in using task ratings from this sample.
- Because coefficient alpha values for content areas were also very high, the WIPT had confidence in the outlined organization of tasks; these high coefficient alpha values further supported their confidence in the final set of task ratings.

Limitations. Limitations of this study should be acknowledged. As indicated, the population of interest was largely hidden, so no sampling frame existed from which to draw a random study sample. Therefore, researchers used a convenience sample of volunteers. Although responses from approximately 78% of the 4,082 volunteers were obtained, many practitioners who performed CMV driver physical examinations were not included.

For example, FMCSA estimates that 40,000 medical examiners will be needed to qualify CMV drivers. Assuming that there are tens of thousands of practitioners who perform these examinations, the substantial efforts of Axiom staff to recruit volunteers for the Role Delineation Study identified only a small fraction of the population. Given these assumptions, these results can extend only to those individuals who gave survey responses. In order to generalize the results of the Role Delineation Study with confidence, the population would have to be studied further—ideally after the establishment of a definitive sampling frame.

The NRCME would eventually generate a definitive sampling frame, so future studies of the medical examiner role should be able to proceed with fewer unknown sources of potential variability. Further role delineation studies may be used as tools beneficial to maintaining the relevance of certification test content. These studies may also present opportunities for fuller descriptions of the medical examiner population.