The Effect of Entry-Level Motorcycle Rider Training on Motorcycle Crashes
Supporting Statement for Information Collection Request

TABLE OF CONTENTS

B.1. Describe the potential respondent universe and any sampling or other respondent selection to be used.

Potential Respondent Universe

The theoretical respondent universe for this study is all novice motorcycle riders in the United States, which based on previous research is defined as riders with less than 500 miles of riding experience. It is difficult to determine how many novice motorcycle riders there are in the US. In 2008, 445,295 students were trained in State-run entry-level motorcycle rider training classes (MSF, 2009)[1]. However, not all riders that receive training are novice riders, and not all novice riders receive training.

An ideal study of the effectiveness of motorcycle rider training would randomly assign a subset of novice motorcycle riders to either receive entry-level motorcycle rider training or to a control group that does not receive entry-level motorcycle rider training, and then would compare the crash, injury, and violation rates, as well as other behaviors of interest, between the groups after the time of the intervention. Unfortunately, the resources do not exist to perform a true experiment using random assignment, neither on a national level nor in a single geographic area. Random assignment of riders to training conditions could additionally be considered problematic if the untrained group is never provided training and if training is found to enhance the safety of novice motorcycle riders.

Sample Acquisition

Because random assignment is not feasible, the proposed study will be quasi-experimental. Matched pairs of trained and untrained riders will be formed on the basis of similarity on such variables as age, gender, years of riding experience, miles ridden per year, crash history, and safety behavior. Matching trained and untrained riders on these criteria will ensure that they are as similar as possible on characteristics relevant to crash risk. In particular, matching trained and untrained riders on safety behavior will control for the possibility that safe riders self-select to take training. We will form pairs between novice riders, which will be defined as riders with fewer than 500 miles of riding experience.

It also could be possible to compare riders that have taken training to riders that tried to enroll in a training class but were unable (e.g., because classes in the area had filled). However, in the past year State rider training programs have generally not experienced a large backlog of riders that are unable to enroll in classes because the classes have filled, and we thus do not believe we would be able to survey a sufficient number of these riders. We will ask untrained riders during screening if they have attempted to enroll in a training class, and will use this information to assess if it is possible to match riders on “desire to receive training” as an additional characteristic.

For the survey, 625 trained riders and 625 untrained riders will be matched into pairs (for a total of 1,250 participants). We estimate that we will screen up to an additional 7,375 trained and 7,375 untrained riders (for a total of 14,750 riders) in order to form these 625 matched pairs. The

---

[1] MSF. (2009). State Motorcycle Rider Education Programs – 2009. http://www.msf-usa.org/Downloads/RiderEdProgramCSI2009.pdf

calculation of how many riders will be screened to form the final sample of matched pairs is estimated from the experience of Billheimer (1998)[2], who screened more than 28,000 riders to form approximately 1,200 matched pairs of trained and untrained riders.

Respondents to the survey will be recruited from up to two metropolitan areas. The Los Angeles and Chicago metropolitan areas are currently being considered as potential sites for this study. Riders will be recruited either from the Los Angeles metropolitan area only, or from both the Los Angeles and Chicago metropolitan areas. If riders are recruited from both Los Angeles and Chicago, matched pairs will only be formed between riders of the same location (i.e., Los Angeles trained riders would be matched with Los Angeles untrained riders, and Chicago trained riders would be matched with Chicago untrained riders).

California and Illinois were considered as sites because they do not require all riders to be trained in order to become licensed. The two potential sites also train a sufficiently large amount of riders a year to allow us to recruit the required number of trained riders for each study. In 2008, California trained 69,839 students in entry-level classes, and Illinois trained 14,917 students (MSF, 2009)[1]. Thirty-three percent of California's rider training sites are in the Los Angeles metropolitan area (28 of 84), and 39% of the motorcycles registered in California in 2008 were registered in the Los Angeles metropolitan area. Forty-four percent of Illinois's rider training sites are in the Chicago metropolitan area (23 of 52), and 62% of the motorcycles registered in Illinois in 2008 were registered in the Chicago metropolitan area. From these figures, it can be roughly estimated that between 6,563 and 9,249 riders were trained in the Chicago metropolitan area (between 44% and 62% of total trained in Illinois), and between 23,047 and 27,237 riders were trained in the Los Angeles metropolitan area (between 33% and 39% of total trained in California).

We acknowledge that these figures are an estimation and do not produce a precise measure of the number of riders trained per year in the study area, or the number of these riders that are novices; however, they do demonstrate that the number of riders trained annually in the large metropolitan areas surrounding Los Angeles and Chicago combined are approximately several times the number of trained riders that we desire to screen for this study. These estimates are shown in Table 2.

---

[2] Billheimer, J. W. (1998). Evaluation of California Motorcyclist Safety Program. *Transportation Research Record, 1640*, 100-109.

Site	State	Entry-Level Riders Trained in State in 2008	% of State Rider Training Sites in Metro Area	% of State Motorcycle Registrations in Metro Area	Estimated Number of Trained Riders in Area
Los Angeles metro	California	69,839	33%	39%	23,407 to 27,237
Chicago metro	Illinois	14,917	44%	62%	6,563 to 9,249
	Estimated total trained population				*29,970 to 36,483*

Table 2. Estimated Number of Riders Trained Per Year in Los Angeles and Chicago Metro Areas.

In each location, a convenience sample of motorcycle riders will be recruited through State-run motorcycle training classes, Departments of Motor Vehicles (DMVs), motorcycle dealerships, motorcycle accessory shops, motorcycle events such as expositions and trade shows, and other places where riders congregate. Advertisements for the study will also be placed on local rider websites and on websites such as Craigslist. California and Illinois have agreed to distribute initial surveys to trainees as they enter the State motorcycle training programs to screen and recruit riders to participate.

We are aware that the sample for this study will not be nationally representative. As only 7.7% of Americans with a driver's license also have a motorcycle endorsement (MSF, 2008)[3], it would not be cost effective to recruit riders for this study through traditional nationwide sampling methods, such as through telephone interviewing using random digit dialing. **It is important to note, however, that we are not trying to replicate the population of American riders. Rather, we want to match trained and untrained riders on the basis of preselected criteria so that they are as similar as possible with the exception of their training status, so that we can isolate the relationship between training and outcomes of interest.**

Non-response rate

Billhemier (1998)[2] found that virtually all riders that were given surveys through their training classes agreed to participate in his motorcycle rider training study, and that riders that were approached at expos refused to participate 12% of the time. It is not possible to estimate the response rate for riders who see advertisements for the study but who are not approached by project staff, but it would be expected that a much smaller percentage of riders that see the ad will volunteer to participate than riders who are approached by project staff. However, as noted above, the goal of this study is not to replicate the population of American motorcycle riders.

[3] MSF. (2008). State Motorcycle Operator Licensing – 2008. http://www.msf-usa.org/downloads/State_Motorcycle_Operator_Licensing_CSI_2008.pdf

Because trained and untrained riders will be matched on a number of characteristics, non-response bias should not be as problematic as it would be in a nationally representative survey.

Because this is a longitudinal study, we also expect that there will be attrition in the sample between survey administrations. In Billheimer's (1998)[2] study of trained and untrained motorcycle riders, 37.4% of riders that were matched into pairs responded to requests by mail to complete a follow-up survey. A survey of motorcycle riders in Maryland by Perrino et al. (2002)[4] found similar attrition between three surveys administered over 2 years. Approximately 50% of riders that participated in Perrino et al.'s (2002)[4] baseline survey responded to a follow-up survey one year later when contacted by phone or mail (between Time 1 and Time 2). Of those that responded to the first follow-up survey, 36% additionally responded to a second follow-up survey administered by mail in the subsequent year (between Time 2 and Time 3).

We will use several techniques to mitigate the effects of attrition in this study. Study participants will be contacted by phone if they do not respond to initial letters or e-mails soliciting their participation in follow-up surveys. Using both mail/e-mail and phone calls to reach participants should result in a lower attrition rate than seen by Billheimer (1998)[2] and by Perrino et al. (2002)[4] in their second follow-up survey, where potential respondents were only contacted by mail. Additionally, since a much larger pool of potential respondents will be contacted in the initial survey, we will attempt to replace riders that do not respond to the follow-up surveys with a different suitable match from this larger pool. We will also attempt to replace untrained riders who become trained over the course of the study.

The primary outcomes of interest in this study are crash involvement and citations. If replacements cannot be found for respondents that drop out of the study, crashes and citations will be obtained for these respondents from public records. Regression will then be used to estimate later mileage for these riders on the basis of the mileage they provided in the initial survey. The actual number of crashes and citations and the estimated mileage will be used to compute estimated crash and citation rates for riders that do not respond to follow-up surveys. A similar strategy was used by Billheimer (1998)[2] to estimate mileage for respondents lost through attrition.

B.2. Describe the procedures for the collection of information.

Procedure for survey study

Trained and untrained riders that are matched into pairs on the basis of the initial survey will be sent two additional follow-up surveys: a first follow-up survey 6 months after completing the initial survey, and a second follow-up survey 18 months after completing the initial survey. Thus, respondents that are matched into pairs will complete a total of three surveys over the course of 18 months. The same survey (see Attachment D) will be administered at all three time points (initial, 6 month follow-up, and 18 month follow-up) so that changes over time can be monitored. A link to web-based follow-up surveys will be e-mailed to the respondents that are

[4] Perrino, C. S., Ahmed, A., Callendar, A., Rozier, E., Cantwell, A., Stewart, O., Raleigh, R., & Joyce, J. (2002). The Role of Maryland's Motorcycle Rider Course in Promoting Safer behaviors and Attitudes. Baltimore, MD: National Transportation Center, Morgan State University.

matched into pairs. If a respondent requests to be sent a paper-based survey instead of a web-based one, a paper survey will be mailed.

The surveys will ask about respondents' training experience, license status, riding experience (mileage and years), motorcycle size, riding purpose, behaviors while riding (e.g., wearing a helmet and protective gear, drinking and riding, speeding, lane-splitting), alcoholic beverage consumption, and crash history.

Crash and citation information will be retrieved from public records. The survey will ask for date of birth and driver's license number from respondents so that public records can be located. As mentioned above, riders will additionally be asked to self-report crash history, as many minor crashes are not reported to law enforcement. Combining public record information with self-report information will give objective data on injury crashes and police-reportable property damage only crashes, as well as information on crashes that may possibly have not been reported. Crash rates per mile ridden, citation rates per mile ridden, and self-reported riding behaviors will be compared between groups at each time point.

Degree of accuracy needed for the purpose described in the justification

Since the expected effect of the treatment (training) is an increase in safe rider characteristics, the differences between trained and untrained groups are expected to be unidirectional. Therefore, one-tailed tests are envisioned. Crash rate, the main outcome variable of interest, will be measured on an interval scale. Rider behaviors, such as drinking and riding and wearing helmets and protective gear, will be measured on a categorical scale.

The following describes how a size of 625 matched pairs will achieve sufficient power to detect differences in the crash rates between trained and untrained riders. We used the PAIREDMEANS method in the POWER procedure in SAS to estimate an approximate size of the sample to achieve a power level of 0.900 to discern a mean group difference of 1 crash per 100,000 miles. For the estimates of expected levels of variation and correlation, we relied on observations from Billheimer (1998). We assumed the mean crash rate would be around 5 crashes per 100,000 miles and the expected variation would be around 6 or 7 crashes per 100,000 miles, based on Billheimer (1998)[2] and the fact that variation in crash rates is overdispersed.

In addition, we assumed that variation could differ slightly across the groups. We also have assumed slightly lower levels of correlation between the pairs due to higher levels of uncertainty associated with survey response and quality in reporting. Table 3 features a set of estimates derived under different levels of variation and correlation to detect a mean difference of 1 crash per 100,000 miles as statistically significant.

The approximate sample size levels of 300-325 matched pairs provided a power level of about 0.900 for standard deviation levels around 5-7 crashes per 100,000 miles with a correlation level of 0.4 – 0.5.

625 pairs were selected as a round figure for the whole effort covering both Illinois and California surveys.

Table 3. Sample Size and Power Computations.

	Stan Dev 1 = 5		Stan Dev 2 = 4	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.900	0.934	0.964	0.985
300	0.941	0.965	0.983	0.995
350	0.966	0.982	0.992	0.998
400	0.980	0.991	0.997	>.999
450	0.989	0.995	0.999	>.999
500	0.994	0.998	>.999	>.999
550	0.997	0.999	>.999	>.999
600	0.998	>.999	>.999	>.999
650	0.999	>.999	>.999	>.999
700	>.999	>.999	>.999	>.999

	Stan Dev 1 = 6		Stan Dev 2 = 5	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.776	0.828	0.882	0.934
300	0.839	0.884	0.928	0.965
350	0.885	0.923	0.956	0.982
400	0.919	0.949	0.974	0.991
450	0.944	0.967	0.985	0.995
500	0.961	0.979	0.991	0.998
550	0.973	0.986	0.995	0.999
600	0.982	0.991	0.997	>.999
650	0.987	0.994	0.998	>.999
700	0.991	0.991	0.997	>.999

	Stan Dev 1 = 7		Stan Dev 2 = 5	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.699	0.752	0.811	0.873
300	0.767	0.817	0.869	0.921
350	0.821	0.866	0.911	0.951
400	0.864	0.903	0.940	0.970
450	0.897	0.930	0.960	0.982
500	0.923	0.950	0.973	0.989
550	0.942	0.965	0.982	0.994
600	0.957	0.975	0.988	0.996
650	0.968	0.983	0.993	0.998
700	0.976	0.976	0.988	0.995

	Stan Dev 1 = 8		Stan Dev 2 = 5	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.622	0.671	0.728	0.793
300	0.691	0.740	0.795	0.854
350	0.749	0.796	0.847	0.898
400	0.797	0.841	0.886	0.930
450	0.837	0.877	0.916	0.952
500	0.870	0.905	0.939	0.967
550	0.896	0.928	0.956	0.978
600	0.918	0.945	0.968	0.985
650	0.935	0.958	0.977	0.990
700	0.949	0.949	0.968	0.983

	Stan Dev 1 = 6		Stan Dev 2 = 4	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.823	0.866	0.909	0.948
300	0.880	0.915	0.947	0.974
350	0.919	0.947	0.970	0.987
400	0.946	0.967	0.983	0.994
450	0.965	0.980	0.991	0.997
500	0.977	0.988	0.995	0.999
550	0.985	0.993	0.997	>.999
600	0.990	0.996	0.999	>.999
650	0.994	0.997	>.999	>.999
700	0.996	0.996	0.999	>.999

	Stan Dev 1 = 7		Stan Dev 2 = 4	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.734	0.780	0.828	0.879
300	0.801	0.842	0.884	0.925
350	0.852	0.888	0.923	0.954
400	0.891	0.921	0.949	0.972
450	0.920	0.945	0.967	0.984
500	0.942	0.962	0.979	0.990
550	0.958	0.974	0.986	0.994
600	0.970	0.982	0.991	0.997
650	0.978	0.988	0.994	0.998
700	0.985	0.985	0.992	0.997

	Stan Dev 1 = 8		Stan Dev 2 = 4	
N	Correlation			
	0.3	0.4	0.5	0.6
250	0.647	0.689	0.736	0.788
300	0.716	0.757	0.802	0.850
350	0.773	0.812	0.853	0.894
400	0.820	0.856	0.892	0.927
450	0.858	0.890	0.921	0.949
500	0.889	0.917	0.943	0.965
550	0.913	0.937	0.959	0.977
600	0.932	0.953	0.970	0.984
650	0.948	0.965	0.979	0.989
700	0.960	0.960	0.974	0.985

all For all computations an α-level of 0.05 is assumed.

Stan Dev 1 and Stan Dev 2 refer to the standard deviation of crashes per 100,000 miles for each rider group (i.e., if Stan Dev 1 is for the trained rider group, then Stan Dev 2 is for the untrained rider group).

As noted previously, rider behaviors such as drinking and riding and helmet/protective gear use will be measured on a categorical scale. The effect size in the matched-paired design for these variables can be measured by the level of difference in discordant proportions. Table 4 shows an example of a distribution of a categorical attribute between the pairs. In this example, distribution of 350 pairs by combinations of whether a rider reported a safe or an unsafe rider behavior is presented.

Table 4. Illustration of Comparison of Discordant Proportions.

Number of Observations		Untrained Riders	
		Safe Behavior (i.e., wears helmet always/frequently)	Unsafe Behavior (i.e., wears helmet sometimes/ rarely/ never)
Trained Riders	Safe Behavior	270	45
	Unsafe Behavior	20	15

Proportions		Untrained Riders	
		Safe Behavior	Unsafe Behavior
Trained Riders	Safe Behavior	77.1% (p_{11})	12.9% (p_{10})
	Unsafe Behavior	5.7% (p_{01})	4.3% (p_{00})

The pairs that exhibit the same characteristics are not relevant for testing. However, the differences in the pairs with varying levels of an observed variable are considered (12.9 % vs. 5.7%). The PAIREDFREQ conducts sample and size and power calculations for McNemar's test for paired observations. The procedure evaluates the null hypothesis that states the discordant proportions are equal to each other in the rider population.

Table 5 shows approximate levels of power under different discordant proportions for the given sample size. Three levels of 'Success-Failure' proportions (p_{10}) are shown with the highest possible Failure-Success (p_{01}) proportions which can lead to a detection of a significant effect.

Given a set of 'Success-Failure' and 'Success-Failure' ratios, Table 5 shows how the power changes with sample size in evaluating the null hypothesis. Since we have no evidence from prior studies, we tabulated a range of 'Success-Failure' proportions to demonstrate the relationships between effect size, sample size and power.

Table 5. Power Levels and Effect Sizes for Individual or Combined Samples.

One site (Los Angeles)

	Success - Failure	10.0%	20.0%	30.0%
	Failure - Success	3.75%	10.50%	18.50%
Number of Pairs	250	0.826	0.840	0.815
	275	0.860	0.871	0.848
	300	0.887	0.897	0.876
	325	0.910	0.919	0.899
	350	0.928	0.936	0.919

Two sites (Chicago and Los Angeles)

	Success - Failure	10.0%	20.0%	30.0%
	Failure - Success	5.0%	13.0%	21.0%
Number of Pairs	425	0.824	0.787	0.816
	525	0.892	0.862	0.884
	625	0.936	0.912	0.929
	725	0.963	0.944	0.957

all For all computations an α-level of 0.05 is assumed.

The tables presented above can be used as a guide whether the tabulated sample size levels are capable of supporting hypothesis tests about alcohol use and riding, and about helmet and protective gear use. If we can consider the response in alcohol use and ride, helmet use, and protective gear use frequency questions as categorical (such as ‘Never’, ‘Rarely’ as “Safe” and “Unsafe” otherwise), and if there is any empirical evidence that or the levels of expected Success-Failure (Safe –Unsafe) proportions by professional judgment are similar to those presented at the top row (shaded in blue), then the a sample size of 325 pairs (in the upper table) is able to recognize the ‘Failure-Success’ ratios that are smaller than those in the second row as statistically significant with a power level of minimum 0.899.

B.3. Describe methods to maximize response rates.

This study will attempt to collect data from respondents at more than one time point, and it will thus be important to maximize the response rate to follow-up surveys. If participants do not respond to our contact efforts to respond to the follow-up survey after one month, we will remind them via mail or e-mail, depending on the preference they stated in their initial survey. We will attempt to reach those who still fail to respond by phone. We will make an effort to replace riders that do not respond to the follow-up surveys with a different suitable match from the pool of riders that responded to the initial survey. We will also attempt to replace untrained riders who become trained over the course of the study.

Crash and citation records will be available from public records from riders that do not respond to the follow-up surveys. If replacements cannot be found for these respondents, regression will be used to estimate later mileage for these riders on the basis of the initial mileage estimates that they provided in the initial survey. The actual number of crashes and citations and the estimated mileage will be used to compute an estimated crash rate for riders that do not respond to follow-up surveys. A similar strategy was used by Billheimer (1998)[2] to estimate mileage for respondents lost through attrition. If the attrition rate for the follow-up surveys is too high to allow for sufficient power, we will only analyze data for crashes and citations (which we will have for all respondents from public records) and will not draw conclusions from survey questions on other behaviors (e.g., drinking and riding, helmet use) in the follow-up survey administrations.

B.4. Describe any tests of procedure or methods to be undertaken.

The survey will be pilot tested with no more than 9 novice motorcycle riders to ensure that the questions are easily understood. The survey will additionally be pilot tested with several motorcycle riders that work for the contractor that will be administering this survey in order to test the instrument on a wider sample and familiarize staff with the survey. Changes will be made to the survey based on the results of pilot testing before they are administered to a larger sample.

B.5 Provide the names and telephone numbers of individuals consulted on statistical aspects of the design.

Jessica Cicchino, Ph.D.
Research Psychologist
National Highway Traffic Safety Administration
1200 New Jersey Ave SE, W46-491
Washington, DC 20590
(202) 366-2752
jessica.cicchino@dot.gov

John Billheimer, Ph.D. (consultant to Cambridge Systematics)
Transcribe
170 Erica Way
Portola Valley, CA 94028
(650) 854-2036
john.wbill@batnet.com

Cemal Ayvalik, Ph.D.
Statistician
Cambridge Systematics, Inc.
115 South LaSalle, Suite 2200
Chicago, IL 60603
(312) 665-0209
CAyvalickca@camsys.com