



**Sam Houston State University
Department of Economics and International Business
Working Paper Series**

Optimal Drunk Driving Penalty Structure

Darren Grant

SHSU Economics & Intl. Business Working Paper No. SHSU_ECO_WP07-07
October 2007

Abstract:

The expected penalty for drunk driving can and does vary by blood alcohol content. This paper outlines the "penalty structure" that does this optimally, using two different metrics, shows how the optimality conditions can be implemented with available data to analyze policy ex ante or ex post, and then uses these findings to investigate four fundamental features of current U.S. drunk driving policy. The paper provides theoretical and empirical support for large penalties at very high alcohol concentrations, but not for reductions in per se blood alcohol thresholds, the most significant recent change in drunk driving policy.

OPTIMAL DRUNK DRIVING PENALTY STRUCTURE¹

Darren Grant
Department of Economics and International Business
Box 2118
Sam Houston State University
Huntsville, TX 77341

dgrant@shsu.edu

Abstract: The expected penalty for drunk driving can and does vary by blood alcohol content. This paper outlines the “penalty structure” that does this optimally, using two different metrics, shows how the optimality conditions can be implemented with available data to analyze policy ex ante or ex post, and then uses these findings to investigate four fundamental features of current U.S. drunk driving policy. The paper provides theoretical and empirical support for large penalties at very high alcohol concentrations, but not for reductions in per se blood alcohol thresholds, the most significant recent change in drunk driving policy.

JEL Codes: K14, I18, C61, D61

¹ This paper arose from an interesting conversation with Morris Coats, and was furthered by an interesting conversation with Richard Cox, who also provided helpful comments. I am grateful to both, to Mishuk Chowdhury and Mohammed Khan for helpful assistance, and to Donald Freeman for thoughtful comments. Through a series of fortunate events, Helen Tashima graciously provided the missing link—elusive information on key variables. A general debt to several papers by Kenkel, cited below, is also acknowledged.

Drunk driving legislation has encountered diminishing returns. Alcohol-related traffic deaths fell after a flurry of federal and state action in the 1980s, which included increasing the minimum legal drinking age to twenty-one. More recently the focus has shifted to lowering per se illegal blood alcohol thresholds, to 0.08 percent for adults and, with “zero tolerance” laws, to nearly zero for among drivers under twenty-one, each now adopted nationwide. Still, of the more than 40,000 annual U.S. traffic fatalities, one-third involve drivers who are legally drunk, a fraction that has remained constant for a decade.

Despite intensive analysis of drunk driving policy by economists, however, it is unclear how to proceed. Partly this is due to a lack of consensus. These studies typically estimate policies’ effects on traffic fatalities using panel data analyses that exploit variation in the dates these policies are adopted by different states. Differences in technique, sample, and controls have often yielded variable or conflicting results, as with the laws mentioned above (Dee, 2001; Eisenberg, 2003; Freeman, 2007; Grant, 2006; Mast, Benson, and Rasmussen, 1999; Ruhm, 1996). These studies are also inherently retrospective, and so cannot be used to analyze proposed changes in the law.

To try to resolve this impasse, we introduce in this paper a complementary, theoretically-grounded method for evaluating existing policy or assessing new policy proposals. We treat drunk driving laws not as a single blood alcohol concentration (BAC) threshold at which one may receive all penalties authorized by statute, but as a menu, or schedule, of penalties associated with different BACs. This perspective mirrors recent changes in the law, as many states have recently moved away from single thresholds by adopting aggravated drunk driving laws that strengthen penalties at very high BACs. (These have received but brief attention in the literature, by Jacobs, 1989, and Kenkel, 1993a, who reach opposite conclusions.) We then characterize the schedule of penalties, or “penalty

structure,” that is socially optimal. This optimality condition can be evaluated with existing epidemiological and economic data to assess current law, recent changes in policy, or new policy proposals. Because severely impaired drivers are one hundred times more dangerous than moderate drinkers, an improved schedule of penalties might significantly improve traffic safety.

Our empirical analysis of existing policy reveals significant violations of the conditions for economic efficiency and a plausible alternate social objective, “expediency,” which generate similar policy recommendations. While some of these accord with recent changes in U.S. drunk driving law, others do not. The first section of the paper introduces the key building blocks of the analysis and derives the optimality conditions; the second introduces the data used to evaluate them; the third section presents results and the last conclusions.

I. Theory.

Setup. Two primitives govern policy design: drivers’ preferences for alcohol and the physiological relationship between alcohol impairment and driving safety (the expected external costs imposed on third parties by any given impaired driver). From these a schedule of drunk driving penalties can be designed so as to maximize some objective. The alcohol consumption then chosen by drivers, mindful of these penalties, ultimately determines the actual costs of drunk driving experienced by third parties.

Let c represent alcohol consumption, in BAC units, c_n^* the privately optimal value chosen by consumer n given existing penalties, and \bar{c} the maximum consumption level for which there is no per se drunk driving violation. There are six variables of interest, each a function of c :

Variable Name	Variable Type	Symbol	Variable Description
Consumer Preferences	primitive	$V_n(c)$	the amount consumer n is willing to pay for permission to purchase c units of alcohol at the market price, drink them, and then drive
Driving Impairment	primitive	$X(c)$	the expected, monetized external damages to life, health, and property generated by a driver who has consumed c units of alcohol
Drunk Driving Penalty Structure	policy	$S(c)$	the expected, monetized value of the legal penalties incurred from driving after consuming c units of alcohol
Resource Costs	policy	$R(c)$	the expected costs, for arrest, conviction, and punishment, expended on a driver who has consumed c units of alcohol
Realized Alcohol Consumption	market-wide outcome	$Q(c)$	the number (technically, density) of drivers possessing c units of alcohol in their blood
Realized External Health Costs	market-wide outcome	$H(c)$	the monetized external costs of life, health, and property loss generated by all drivers who have consumed c units of alcohol

All expectations are unconditional—that is, per driver or per trip, not per arrest or accident. In the theory each variable is assumed continuous and measured in dollars where appropriate; in practice they are measured, in a variety of units, in standard BAC increments of 0.01 percent.

Analysis can proceed on two margins. The extensive margin concerns the resources devoted to drunk driving deterrence and, loosely put, the overall magnitude of penalties for drunk driving. This margin has been adequately studied by Kenkel (1993a, 1993b, 1996), who concludes that existing penalties were inefficiently low and that the social costs of deterrence through alcohol-control policies slightly exceed those of alcohol taxation.

The focus of our paper, instead, is the intensive margin—the “shape” of the penalty structure, or how penalties grow with BAC. For this one can derive equations, analogous to cost-minimization conditions in production theory, that show how best to structure penalties across BAC levels for any

given level of resources/penalty magnitudes. Because these equations do not require specific units of measurement for any variable, they can be evaluated effectively with our data. Upon doing so, deviations from optimality and the policy changes implied are immediately apparent.

Efficiency. We first assume, as is natural, the policy objective is economic efficiency, that is:

$$\max_{c_1^* \dots c_N^*} \sum_{n=1}^N V_n(c_n^*) - X(c_n^*) - R(c_n^*) \quad (1)$$

because harmful internal effects of drunk driving are already captured in V. The optimality condition $V_n' = X' + R'$ is to be met at all c_n^* . We can neglect R' , which is dominated by X' .² Then, assuming expected utility maximization by consumers,³ the policymaker should set $S' = X'$ for all c. This is just the standard Pigovian tax that forces agents to internalize external costs. One could assess policy

² The resource costs per arrest and conviction in Kenkel (1993b) and the average probability of arrest in Kenkel (1993a) yields a resource cost of roughly \$2 per occasion of drunk driving. The latter source also estimates total external costs to be about \$50 per occasion of drunk driving. And data below show that X rises faster than R does. Thus, X' is at least twenty-five times R' .

Equation (1) implicitly assumes punishment involves transfers (fines) rather than deadweight losses (license suspensions or jail time). These cannot be justified under any efficiency criterion. One can easily show, however, that the optimality condition derived below takes the same form if deadweight losses form a constant proportion (in c) of all penalties, which is reasonably accurate.

³ This natural and typical assumption may require more rationality than is present. Studies have shown many people do not “know the applicable laws in their states” (Kenkel and Koch, 2001) and that the certainty of punishment deters youth more strongly than its severity (Grosvenor, Toomey, and Wagenaar, 1999). In the most complete model, based on an extensive literature review (Donovan, Marlatt, and Salzburg, 1983), “drinking behavior, personality traits, [and] emotional stress” combine to produce high risk driving: “When an individual...deficient in those social skills involved in the appropriate expression of anger and the management of stress, frustration or depression...is confronted by an acute emotional stress...the availability of alcohol or an automobile may provide an alternative means...of dealing with their underlying feelings in the absence of more adaptive emotional expression... Such individuals appear to [expect] that both alcohol and driving are effective in decreasing tension [and] increasing feelings of personal power, mastery and control. The probability of drinking and driving together is thus increased.”

efficiency by evaluating this condition at all values of c if S and X are measured in dollars. We cannot do this reliably, as previously discussed, and need not, given our focus in this paper.

For our purpose a weaker condition, necessary but not sufficient for efficiency, allows S and X to be measured in arbitrary units. This condition, in the spirit of Mirrlees (1976), is analogous to the equi-marginal principle in production theory, and requires only that any decrease in drunk driving damages be done at least cost to consumers (the smallest aggregate reduction in V):

$$\max_{c_1^* \dots c_N^*} \sum_{n=1}^N V_n(c_n^*) \quad \text{subject to} \quad \sum_{n=1}^N X(c_n^*) = \Phi \quad (2)$$

with Φ an arbitrary scalar. The optimality condition $V_n' = \lambda X'$, with $\lambda(\Phi)$ the scalar Lagrange multiplier on the constraint, is to be met at all c_n^* ; this is achieved by setting $S' = \lambda X'$ for all c . Then X'/S' should be constant over this domain but the value of the constant need not be specified. For any cost-effective penalty structure S^* , any multiple kS^* is also cost-effective: the “level of penalties” can be separated from the shape of the penalty structure.

We can interpret X'/S' as a ratio of marginal benefits (the reduction in drunk driving damages imposed on third parties) to marginal “costs” (the value of the marginal reduction in alcohol consumption by drivers). When this ratio is not constant, an equal reduction in external damages can be obtained at less aggregate inconvenience to drinkers by raising the marginal penalty at BAC levels where this ratio is large and lowering it where this ratio is small; or a larger reduction in external health damages obtained at the same aggregate inconvenience, in the same way. This condition requires drunk driving damages to be reduced in the most cost-effective manner, so we term this objective “cost-effectiveness,” with the health economics connotation deliberate.

Because this ratio is unitless, we take logarithms so that (modest) deviations from optimality

are expressed in (roughly) percentage terms. This yields the cost-effectiveness optimality condition:

$$\log(X') - \log(S') = \kappa \quad (3)$$

with κ a constant whose value is meaningful only when S and X are measured in identical units: then $\kappa = 0$ implies efficiency, $\kappa > 0$ an inefficiently low penalty, and $\kappa < 0$ an inefficiently high penalty. Even when not so measured, the left-hand side of this equation should be constant for, at least, all $c > c$. If not, strengthening marginal penalties is most desirable in the BAC range where the left-hand side of this equation takes the most positive value.

Expediency. Economic efficiency may not be policymakers' objective, however, because society may not give drunk drivers "standing" in the cost-benefit analysis that underlies efficient policy design. This necessitates balancing marginal external costs with the utility losses derived from reduced alcohol consumption. But society, in branding the drunk driver a criminal, may refuse to incorporate these losses into the determination of policy. Then drunk driving is not longer a typical externality problem. To quote Trumbull (1990, p. 212; see also numerous references therein): "Society has a purpose when it labels certain acts criminal; the label communicates that these acts will not be tolerated or counted in the social weal." Trumbull frames this in terms of rights; the setting of drunk driving thresholds may be construed as identifying a threshold at which the individual no longer has the right to use public roads. Kenkel (1998) contains a thorough discussion of the issues involved in using economic efficiency to design drunk driving policy.

The "rights" perspective can be illustrated by the most recent extended spirited debate on drunk driving policy in Congress, which concerned a 1998 bill impelling the states to adopt a 0.08

per se illegal BAC limit for adult drivers. (*Congressional Record* 144, no. 19-20, 3-4 March 1998:S1236-S1305. Technically an amendment to a larger transportation bill, it passed the Senate 62-32, with bipartisan support and bipartisan opposition, but was dropped in conference. A similar bill became law with little discussion in the next Congress.) Proponents asserted that significant driving impairment occurs at 0.08, that this impairment makes drivers more dangerous, and that the lower limit would reduce traffic fatalities. None of these points were disputed by the opposition.

Proponents also recognized three relevant costs: the external costs of drunk driving just mentioned, the “cost” to the restaurant and beverage industry in lost sales, and the potential cost of limiting social drinking, such as “two beers and a pizza.” The first was presumably recognized by all parties, while the other two were asserted to be nil, with evidence presented that restaurant and alcohol sales were not affected by 0.08 laws in individual states. These assertions were also undisputed. Tellingly, neither side mentioned the key “cost” on which efficiency hinges, consumer surplus lost by reduced alcohol consumption among heavy drinkers intending to drive.

Instead of costs, the debate turned on rights. Three rights were acknowledged: social drinkers’ right to consume modest amounts of alcohol and drive (or large amounts and not drive), citizens’ right to drive on safe roads, and each state’s right to choose its own highway safety legislation. None of these rights were ever disputed either; the question was whether the second or third would prevail. Three quotes frame the two sides’ positions adequately:

Dorgan (D-ND): You have a right to get drunk...in this country, but you do not have a right to get drunk and drive. That ought to be a message from the .08 amendment.

DeWine (R-OH): It is a question of rights...the right to have a fair chance on the highway, not to have someone come at you who has been drinking and driving.

Nickles (R-OK, also see Thomas, R-WY): If I were a state legislator, I would support the lower level [.08]...But I think, again, we are trampling on State's rights.

The record, at least on this legislation, indicates lawmakers' lodestone is not economic efficiency, but the desire to lower drunk driving fatalities as much as possible within this framework of rights, and an implicit citizens' right (provided by the eighth amendment) that punishment not be excessive.

This conclusion leads directly to an alternative mathematical program in which fatalities are minimized subject to penalty bounds, as follows:

$$\min_G \sum_{n=1}^N X(c_n^*) \quad \text{subject to } S(c) \geq 0 \quad \forall c \leq \underline{c}, \quad S(c) \leq S_{MAX} \quad \forall c > \underline{c} \quad (4)$$

with S_{MAX} the penalty maximum. Taking the dual of the cost-effectiveness program, equation (2), the objective remains the same, only the constraint has changed. Therefore the policy implications of the two frameworks, while different, need not be divergent.

This framework has not trivialized the problem. Optimality implies the maximum penalty S_{MAX} will be imposed at the largest conceivable BAC level, c_{max} , which for practical purposes is about 0.40, but not necessarily below that level. Rather, the policymaker will "parcel out" marginal penalties $S' \equiv M$ across the entire domain to maximize efficacy (barring a corner solution). We can derive and evaluate the optimality condition that describes the outcomes associated with this policy.

This condition is most easily expressed using M :

$$\sum_{n=1}^N X' \partial c_n^* / \partial M = \hat{\lambda} \quad \forall \underline{c} < c < c_{max} \quad (5)$$

where $\hat{\lambda}(S_{MAX})$ is a (different) scalar Lagrange multiplier, and c_n^* is chosen using $V_n' = S' \equiv M$, as above. To interpret this condition, consider Figure 1, which presents the optimal marginal penalty

structure $M(c)$ and a small change in that policy (an extremal) at \hat{c} , magnified in size. Drivers Y and Z respond to this change by reducing their consumption, as shown in the figure. The marginal effect of this small increase in penalties equals the change in consumption it inspires, $\partial c^*/\partial M$, converted into external costs using X' , for all individuals affected by the policy, which is proportional to $Q(\hat{c})$, thus generating the following close approximation to equation (5):

$$X'Q(c)\overline{\partial c^*/\partial M} = \hat{\lambda} \quad \forall \underline{c} < c < c_{\max} \quad (6)$$

where the average is taken across all drivers affected by the policy at that value of c . Optimality implies a type of equi-marginal condition: the marginal effect of a small change to the penalty structure, the product of the number of drivers affected, the average change in consumption of those affected, and the rate at which this change reduces external costs, should be equal for all $c > \underline{c}$. Otherwise drunk driving damages could be decreased by rearranging the penalty structure, without increasing the maximum penalty, or one is at a corner solution, in which case marginal penalties should be focused on consumption levels where the marginal effect is greatest.

While Q is not measured, it is proportional to H/X , each of which are measured; substituting and taking logarithms yields a relation that is less easily interpreted but more easily quantified:

$$\log(\log(X)) + \log(H) + \log(\overline{\partial c^*/\partial M}) = \hat{\kappa} \quad \forall \underline{c} < c < c_{\max} \quad (7)$$

with $\hat{\kappa}$ again an arbitrary constant. We call this condition “expediency,” because of its political connotations. An expedient policy reduces the external costs of drunk driving as much as possible given the socio-political constraints limiting the severity of penalties that can be imposed.

With expediency, unlike cost-effectiveness, the magnitude of penalties and the shape of the

penalty structure are intertwined. The former, the extensive margin, is indexed by S_{MAX} . When this is small, penalties are weak, and will focus on heavy drinkers, who cause great damage when driving. In essence, with scarce amounts of “marginal penalty” to distribute, they will be put where they do the most good. Only when penalties are stronger, and drunk driving is diminished, can attention be paid to more modest drinkers. In the extreme, penalties are draconian, and are fully applied at low BACs, which few drinkers will exceed. This evolution is broadly consistent with the reduction in BAC thresholds observed in the U.S. as drunk driving sanctions have become more forceful. Because penalty magnitudes influence the shape of the expedient penalty structure, one cannot delineate a unitless optimality condition directly in terms of penalties, S , as we did for cost-effectiveness. But we can use outcomes, H , which is how equation (7) is expressed.

The expediency and cost-effectiveness conditions are clearly distinct, but are not necessarily at odds. Increasing marginal penalties at \hat{c} will reduce fatalities at that alcohol concentration, $H(\hat{c})$, so both conditions will take smaller values. While existing policy might be more closely aligned with expediency than cost-effectiveness, or vice versa, similar policy adjustments could be called for under both. This would be fortuitous, for economic analysis can more usefully inform policy if its prescriptions are not sensitive to the assumed policy objective.

We can set out the circumstances under which the two conditions imply equivalent penalty structures by solving for X' in each, equating them, and rearranging:

$$Q = \bar{\kappa}(\overline{M \cdot \partial c^* / \partial M})^{-1} = \bar{\kappa}(\log \bar{V}^*)' \quad (8)$$

When $\partial c^* / \partial M$ is constant across BAC levels, which appears to be roughly the case (see below), expediency and cost-effectiveness equate if Q is inversely proportional to M .

II. Data.

To evaluate our optimality conditions, the terms X' , $\partial c^*/\partial M$, H , and M must be measured for all feasible BACs for which driving is outlawed, which as a practical matter, given our data, is $[0.08, 0.24]$. Each optimality condition only requires constancy across values of c , so the units of measurement do not matter. Three of these terms can be evaluated with reasonable precision and the other bounded conservatively, that is, so as not to promote our ultimate policy recommendations.

The marginal external cost primitive, X' , is taken from Blomberg et al.'s (2005) monolithic, exhaustive, NHTSA-funded case-control study of 15,000 drivers in Florida and California, one-third crash-involved, the remainder controls. Relative crash probabilities were determined for each 0.01 unit of BAC using standard case-control methods. Using accident and driver-survey data, these were adjusted for driver demographics and accident characteristics using logistic regression, and then further adjusted for data collection problems stemming from missing covariates, refused breath tests, and hit-and-runs, using supplemental data collected by the investigators. The 95% confidence interval on the crash risks, roughly $\pm 15\%$, is well below the range of variation in the risks themselves.

Figure 2 presents X , the crash risk relative to a non-drinker (whose value is one), and its difference, $X(c) - X(c-.01) \approx X'$. Relative risks increase at an increasing rate, rising dramatically above 0.10 and even more dramatically around 0.20, a range that exceeds the current adult threshold of 0.08 by a wide margin. Compared to legal BACs, relative risks in this neighborhood are one hundred times higher and their rate of increase, X' , ten times higher.

These data do not account for crash severity, and the relative risk of fatal crashes increases even more rapidly with BAC (Zador et al., 2000). This presents no bias if crashes and fatalities are

proportional, as our optimality conditions are unitless. Comparable data on fatal crash risks for the full BAC domain is unavailable, but we can investigate further by comparing crash and fatality numbers by BAC for one state that identifies both, California (Tashima and Daoud, 2007). These are listed in Figure 3 for 2004 in log points. Over the 0.08-0.24 BAC range, the two measures differ by roughly a constant number of log points—they are indeed roughly proportional.

The data also do not distinguish external costs from costs internalized by drunk drivers. It is not even clear how this should be done, whether, for example, X should count harm to drivers that underestimate the risks of drunk driving, or exclude harm to passengers who may rationally recognize the risks of riding with a drunk driver (Kenkel, 1998). Fortunately, Figure 3 shows that driver fatalities and passenger fatalities are also roughly proportional over the 0.08-0.24 BAC range, so there will be no bias however they are weighted.

The second term, $\partial c^*/\partial M$, measures the responsiveness of consumption to penalties. Despite much work on the effects of BAC limits on aggregate fatalities, little is known about their effects on alcohol consumption among different types of drinkers. An evaluation of zero tolerance laws by Wagenaar, O'Malley, and LaFond (2001) indicates similar reductions in mild drinking and heavy drinking, and similar reductions in driving after drinking for both groups. An investigation of the same laws by Carpenter (2004) found greater reductions in heavy drinking than overall drinking and no change in drunk driving conditional on drinking. These studies suggest $\partial c^*/\partial M$ is, at least, no smaller for heavy drinkers.

More plentiful, but less direct, evidence comes from studies of the price elasticity of alcohol among light and heavy drinkers. (Evidence on the price elasticity of drunk driving, rather than drinking, among different types of drinkers is not available.) The relevant estimates indicate price

responsiveness among all but possibly the very heaviest drinkers. Chaloupka, Grossman, and Saffer (2002) review several studies of drinking among young adults that collectively show absolute declines in consumption, in response to price increases, that are at least as large among heavy drinkers as among light or moderate drinkers. Manning et al. (1995) and Kenkel (1996) find the overall price elasticity of heavy drinkers to be about half that of moderate drinkers (at least for men; it is higher for women), which also implies similar absolute changes in consumption, as heavy drinkers consume much more alcohol on average. Finally, Farrell et al. (2003) find a substantial price elasticity of heavy drinking that seems to exceed that for drinking generally. Collectively, these studies also indicate $\partial c^*/\partial M$ is not smaller at high BACs. For simplicity we assume—conservatively given our conclusions—it is constant for all c .

Realized fatalities, H , are measured using the Fatality Analysis Reporting System (FARS) of the National Highway Safety Transportation Administration (NHTSA), which records the BAC of all drivers involved in all fatal crashes on U.S. public roads.⁴ It is unclear how to ascribe responsibility in multivehicle crashes, so we tried two methods: 1) only tabulating fatalities in single-vehicle crashes, and 2) assigning all fatalities to the largest BAC among all drivers involved in the crash. Over the 0.08-0.24 BAC range we focus on, the two gave almost identical results, so the former is used here. Results were tabulated for all BAC levels for all states over the period 1988-2000. The long period was necessary to minimize “sampling error”; over the period many states had per se limits of 0.10, roughly consistent with the data used to calculate S . We did not exclude states with lower BAC limits, preferring to retain a balanced panel. This inclusiveness represented a

⁴ Driver BAC is measured directly in about half of the observations and is imputed using NHTSA’s imputation technique for the others, mostly nondrinkers (Subramanian and Utter, 1998). Only about 5% of the observations involve drinking drivers with unmeasured BAC.

tradeoff between precision, logical consistency, and national representativeness.

Figure 4 presents H and the implied density of driver BAC, $Q \propto H/X$, in arbitrary units.⁵ The prevalence of heavy drinkers in fatal crashes, long emphasized by NHTSA, is apparent here. Heavy drinkers comprise a much smaller fraction of all drivers, however; the difference between the two proportions is attributable to the greatly elevated crash risk of heavy drinkers.

The final variable, S, is the monetary equivalent of the expected drunk driving penalty. Kenkel (1993b) sets this out as follows:

$$S(c) = P_A(c)P_C(c)L(c)(1+r)^{-T(c)} \quad (9)$$

with P_A the probability of arrest, P_C the probability of conviction if arrested, L the monetized penalty if convicted, r the discount rate, and T the interval between crime and punishment. All but r are potentially functions of c , and each can be changed by the allocation of resources, policy, or both.

The only one of these four terms that is directly observed across the full BAC domain is the one that varies the most: P_A , the probability of arrest (see the discussion in Perrine, Peck, and Fell, 1988). This is taken from the classic source, Beitel et al. (1975), based on a field experiment in Kansas City in 1973. (Similar results are obtained in Hause, Voas, and Chavez, 1982, in California, under similar circumstances.) The probability of arrest conditional on BAC is calculated, in 0.05 unit increments (0.01-0.05, etc.) from the unconditional probability of arrest in the field experiment and survey evidence on the distribution of BAC among drivers who were, and were not, arrested for drunk driving. We converted this to 0.01 increments by fitting a cubic spline to the data.

⁵ Because H is not adjusted for demographics, a similarly unadjusted version of X, from the same source, was used in calculating Q. This Q is similar to that reported in Beitel et al. (1975).

Lacking data on T , we assume it is constant in c , which is not unreasonable as discounting should affect S little given the relatively short interval—months, not years—between arrest and sentencing. We assume the same for P_C , given that conviction rates are very high to begin with, and present supportive evidence below. Data on actual penalties conditional on conviction (as opposed to minima or maxima) is sketchier; the only published source we could find is a report by Tashima (1986). This finds these increase only slightly at BACs exceeding 0.20, and personal communication with its author reveals that, since then, “sanctions have become much more homogenous for [drunk driving] offenders.” Thus we initially assume constant L , but later relax this using information on aggravated drunk driving laws adopted by many states. As a practical matter, $dS/dc > 0$ mostly because more inebriated drivers are more likely to get caught.

Both S and $S(c) - S(c-.01) \approx S' \equiv M$ are presented in Figure 5. The probability of arrest, and hence the implied S , increases substantially at all BAC levels above 0.10; the risk of arrest at 0.20 is about ten times that at 0.10, a sizeable difference but much smaller than the relative crash risks illustrated above. While the optimality condition involving S is based not on levels, but rates of change, these too grow much more slowly for arrests than for fatalities. Increases in M , at the bottom of Figure 5, do not keep pace with increases in X' .

We can check the validity of our data by re-calculating S using, instead, the crash and drunk driving conviction (not arrest) variables in Tashima and Daoud (2007), shown in Figure 3. The former and X determines Q ; this and the convictions data determines the probability of being convicted conditional on BAC, $P_A P_C$ (again in arbitrary units). This is an alternative measure of S if T and L are assumed constant in c , as before. It is very similar in Figure 5, suggesting several key variables have been characterized with sufficient accuracy.

Figures 2-5 show that the greatest variation in key variables occurs at BACs far beyond the current threshold of 0.08. This disconnect is intriguing, suggesting that the greatest variation in our optimality conditions may occur at BAC levels greatly exceeding those targeted by existing policy.

III. Results.

Figure 6 graphs the cost-effectiveness condition for the BAC range 0.08-0.24. This is the difference between the value of decreased crash risk resulting from a one unit reduction in driver BAC and (mindful of the equality $V' = S'$) the utility loss resulting from that reduction, measured in logs, and scaled to zero at $c = 0.10$. Larger values are more cost-effective, the region where penalties most need strengthening. Welfare gains from penalty increases here would more than offset losses from penalty decreases elsewhere, were these necessary to satisfy some constraint.

Cost-effectiveness is lowest at BACs below 0.15 and increases steadily thereafter, to a maximum of 1.5 log points, a five-fold difference in cost-effectiveness over this range. By this standard, policy is too heavily weighted toward drivers with BACs under 0.15. Reducing penalties on these drivers and increasing them on heavier drinkers would save more lives at the same aggregate “cost” in withheld consumption. The potential improvement is substantial given the large difference in cost-effectiveness at high and low BAC levels.

Figure 6 also illustrates the expediency condition over the same BAC range. Each point of this figure depicts, again in arbitrary, logged units scaled to zero at $BAC = 0.10$, the reduction in crashes obtained by a one-unit increase in M at that BAC value. Larger values are therefore more expedient than smaller values. This condition, too, is smallest at moderate BACs, below 0.12, rising

rapidly above that point until falling again beyond 0.20. The gap between maximum and minimum is again sizeable, about one log point, similar to that observed for cost-effectiveness. This large gap implies that fatalities could be substantially reduced without changing the maximum penalty, by increasing marginal penalties where expediency is largest and lowering them where it is smallest.

Is current policy more consistent with cost-effectiveness or expediency? It is not easy to choose. While ranging significantly from maximum to minimum, each condition also takes a constant value across a large portion of the domain. In both cases deviations from constancy support increasing marginal penalties at BACs far above the current 0.08, such as those in the 0.15-0.20 range utilized for aggravated drunk driving laws.

In fact, although current policy is neither expedient nor cost-effective, there is considerable harmony between the two conditions. As previously discussed, these will coincide if the marginal penalty, M , is inversely proportional to the fraction of drivers at each BAC level, Q . Fortuitously, this is basically true, at least until BAC exceeds 0.20, as Figures 4 and 5 confirm. One might reasonably conclude that the similarities in these conditions are more noteworthy than their differences. This result is empirical, however, not theoretical, and need not hold elsewhere.

We now use these conditions to analyze four features of U.S. drunk driving policy.

Lower BAC Thresholds. All states have lowered the illegal per se BAC threshold for adults from 0.10 to 0.08. This change increases marginal penalties, S' , in the neighborhood of 0.08 and *decreases* them near 0.10. How does this affect cost-effectiveness and expediency?

For this change to improve either, the value of the condition must be falling over the 0.08-0.11 BAC range, meriting stronger penalties at lower end of this range, where cost-effectiveness and expediency would be highest. Instead, Figure 6 shows that cost-effectiveness is roughly constant

over this range, while expediency is rising, as the elevation in crash risk at higher BACs more than outweighs the increase in expected penalties.

The implication of this result is striking: lowering BAC thresholds from 0.08 to 0.10 should not reduce traffic fatalities, and might increase them. As marginal penalties are shifted downward heavier drinkers relax their efforts to comply with the law, and this will outweigh the effect of reduced consumption from the more modest drinkers the new law focuses on. This conclusion runs counter to two well-known studies supporting the efficacy of 0.08 laws, Dee (2001) and Eisenberg (2003). Analyzing fatalities from 1982 through 1998 (Dee) or 2000 (Eisenberg), both find 0.08 laws reduce overall traffic fatalities by 2-3% compared to a threshold of 0.10. This subject has been recently re-examined by Freeman (2007, especially Figure 3 and Table 3, columns G-I), who extends the sample to 1980-2004, adding several states that recently adopted 0.08 laws, and carefully addresses serial correlation in the residuals, which is substantial. When doing so, Freeman's point estimate is almost identically zero. He then focuses on states passing these laws because of congressional incentives, rather than voluntarily, as a cleaner "natural experiment" less likely to generate biased coefficients. In this group the point estimate, though insignificant, is positive, consistent with the conclusions derived above. It is unlikely that 0.08 laws have any material effect on traffic fatalities, and plausible that they increase them slightly.⁶

Aggravated Drunk Driving Laws. According to Mothers Against Drunk Driving, forty states currently have "aggravated drunk driving" statutes mandating additional penalties or reduced

⁶ One can also bound the effects of a 0.08 law mechanically, using the X and Q variables presented above, by assuming that all drivers with BACs of 0.08 and 0.09 reduce consumption to 0.07 to comply with the new threshold. This implies a reduction in fatalities, with full compliance, of 1%. This calculation further supports Freeman's findings about the efficacy of 0.08 laws.

flexibility in adjudication or sentencing when the driver's BAC exceeds a threshold that ranges from 0.15 to 0.20. These extra penalties range from weak (South Dakota only requires pre-sentencing alcohol evaluations at BACs of 0.17 or higher) to strong (Idaho requires ten days jail time, doubles the maximum fine, and more than doubles the period of license suspension for drivers with BACs of at least 0.20). Several other countries also have similarly graduated penalties (NHTSA 2000).

In Figure 6 we analyze the effect of a hypothetical aggravated drunk driving law, somewhat stronger than the median, that doubles penalties for BACs exceeding 0.15. Cost-effectiveness is dramatically improved. Aggravated drunk driving laws increase marginal penalties in the BAC region where increased penalties have the greatest ratio of benefits to costs. Because drunk driving penalties are inefficiently low (Kenkel, 1993a), economic efficiency increases as well.

We cannot show how aggregated drunk driving laws affect expediency without asserting how they will change the distribution of fatalities, H , which no study has attempted to do. Still, it is clear from Figure 6 that expediency should also increase, as these laws operate in precisely the region of the graph where increased penalties reduce external costs the most. If drunk driving penalties are to increase, this should occur at the BACs targeted by aggravated drunk driving laws: 0.15-0.20.

In an earlier analysis, Kenkel (1993a) argued aggravated drunk driving laws were probably unwarranted under an efficiency criterion because the crash risk and the arrest risk increased with BAC by similar proportions. The more recent, more detailed data employed here indicate otherwise. Accounting for hit-and-runs and breath test refusals makes the difference. Blomberg et al. (2005) track down about one-sixth of all hit-and-run drivers, many of whom have very high BACs, and use passive alcohol sensors to impute the BAC of breath test refusals, many of whom also have high BACs. Accounting for these problems increases the relative risk at a BAC of 0.20 threefold. This

conclusion is reinforced with the California data reported in Figure 3, where the ratio of drunk driving convictions to crashes (the difference in log points) diminishes steadily above a BAC of 0.08. These differences are brought into even sharper relief in our optimality conditions, which are based on marginal changes in these variables, not levels.

Youth vs. Adults. Throughout the 1980s and 1990s states adopted lower BAC thresholds for drivers under twenty-one, so called zero tolerance laws. Is this supported by expediency or cost-effectiveness? Are other differences in the penalty structure between youth and adults so justified?

Efficiency-based age differences in drunk driving penalties can be justified by differences in marginal crash risks, X' . These clearly exist. Zador et al. (2000) find crash risks are twice as high for youth (particularly males, who cause most crashes) as for adults at all BAC levels studied, so $X'_{\text{YOUTH}}(c) \approx 2X'_{\text{ADULT}}(c)$. This proportionality implies the youth and adult penalty structures should have similar shapes but different magnitudes. Current policy has the opposite features: different shapes, because the BAC thresholds differ, but similar magnitudes once the threshold is violated.

An equivalent result obtains for expediency, which depends on $\partial c^*/\partial M$ and H . The studies cited above suggest the rough constancy of $\partial c^*/\partial M$ applies for both youth and adults, while our analysis of the FARS data indicates that the distribution of fatalities, H , is also proportional between youth and adults (for drivers with positive BAC). Expediency also implies the adult and youth penalty structures should be similarly shaped. This finding, coupled with our earlier results concerning the expediency of lower BAC thresholds for adults, implies that zero tolerance laws, too, should not reduce drunk driving fatalities. In a recent, exhaustive, retrospective panel study, the first to focus on assessing zero tolerance laws econometrically, Grant (2006) concludes the same thing.

Uniformity of Laws. A final feature of drunk driving policy is increased uniformity across

the states, partly in response to Congressional incentives to enact 0.08 and zero tolerance laws. Cox (2006) has pointed out this uniformity need not be optimal, and this can be formally investigated here by determining whether the optimality conditions that guide policy vary systematically across states. If so, it is probably inappropriate for all states to have uniform BAC thresholds.

For expediency, this is likely to be so. Based on calculations by the author and NHTSA, both the extent of alcohol involvement in fatal crashes and the average BAC levels of the drivers involved vary across states. In 2002, for example, the percentage of drivers involved in fatal accidents having a BAC of 0.08 or higher had a 0.90/0.10 decile range, across states, of 28% to 16%. This variation suggests the shape of H , and hence of the expediency condition, varies across states, as illustrated in Figure 6 for the five states with the largest values of the aforementioned percentage and the five states with the smallest values. Differences between the two lines are dramatic. Uniform policy is probably not expedient. Whether it is cost-effective is unclear.⁷

IV. Conclusions.

In academia traffic safety policy is generally viewed in hindsight, through retrospective empirical studies of laws' effects on traffic fatalities. Here, however, hindsight is not perfect: these studies are subject to the same estimation problems that plague economists elsewhere, and studies of a given policy often generate conflicting or ambiguous results (Benson, Rasmussen, and Mast,

⁷ Cost-effectiveness does not depend on Q or H , only, possibly, on other factors w that influence traffic safety. When these factors increase the crash risk of all drivers by the same proportion, so that $X(w,c) = f(w)g(c)$, cost-effectiveness is identical across states, policy uniformity is appropriate, and Federal mandates to bring this about may be justified. One cannot determine whether this separability is genuine, but it is plausible.

1999). Furthermore, this type of analysis is necessarily limited to laws that have been adopted fairly widely for long enough that their influence on behavior can be estimated with confidence. This justifies considering an alternative approach not subject to these restrictions.

The alternative proffered here simply sets out the conditions that characterize optimal policy and evaluates them with available data. This allows the whole of the penalty structure to be examined, for the first time, permitting both ex post analysis of existing policy or ex ante analysis of new policy proposals. One can adopt either a traditional efficiency perspective or a politically-oriented expediency perspective; fortunately, here the two are quite harmonious in their implications. This approach is not without weaknesses, however, particularly limitations of the economic analytical approach and data limitations. Still, observed deviations from optimality are large and robust, so our essential conclusions are strongly supported.

Comprehensively, our findings support applying greater marginal penalties on drivers vastly exceeding the current adult threshold of 0.08, whether they are youth or adults, permitting flexibility in the penalty structure across states when doing so. While legislation has evolved in this direction in some respects, such as the adoption of aggravated drunk driving laws, it has diverged in others. This may explain the stasis in drunk driving fatalities observed over the last decade, despite legislation designed to lower BAC thresholds and drunk driving generally. The tools presented here can be used to evaluate future proposals in the quest for optimal drunk driving legislation.

Figure 1. Expedient Policy and a Small Deviation from Expediency (extremal, magnified in size).

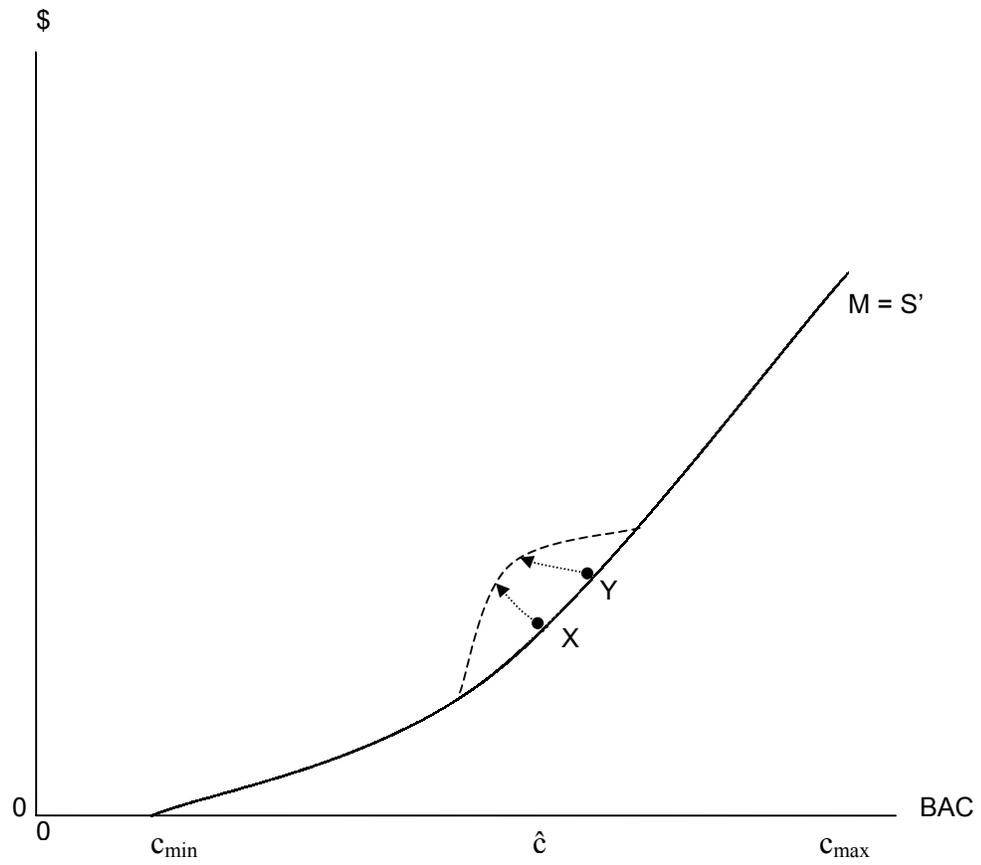


Figure 2: Physiological Relationship between Alcohol Consumption and Crash Probability.

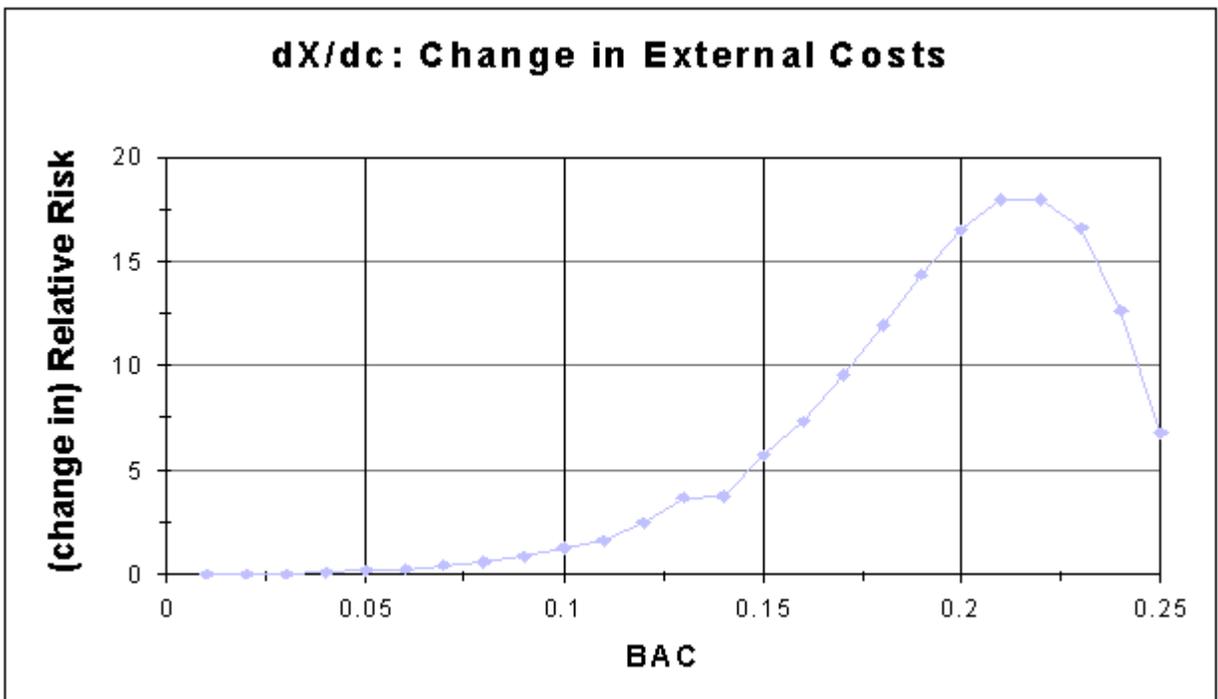
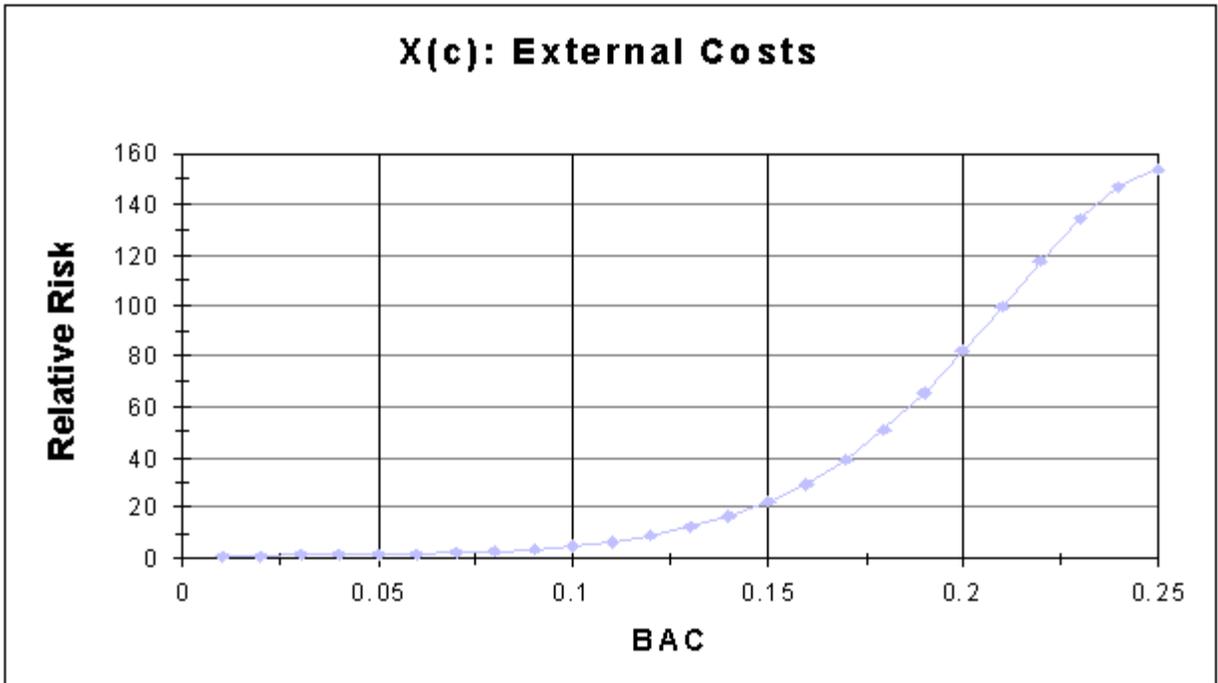
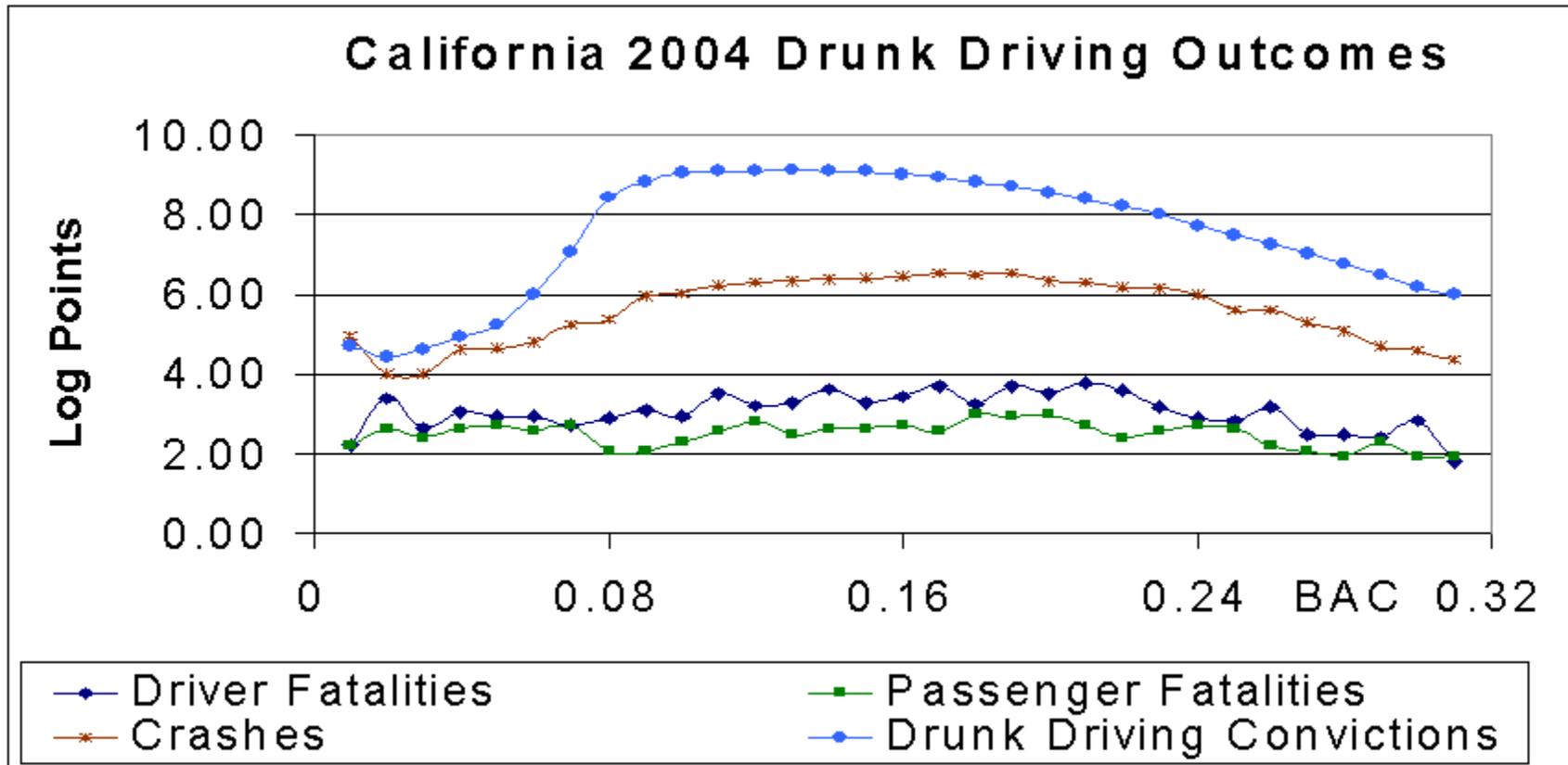


Figure 3. Drunk Driving Outcomes for California, 2004.



Note: Fatalities come from FARS, described in the text, and the other variables from Tashima and Daoud (2007). Drunk driving convictions include DUI and alcohol-related reckless driving convictions. All variables are annual aggregates for the year 2004, logged.

Figure 4. Distribution of Fatalities and Imputed Distribution of Drivers by BAC Level.

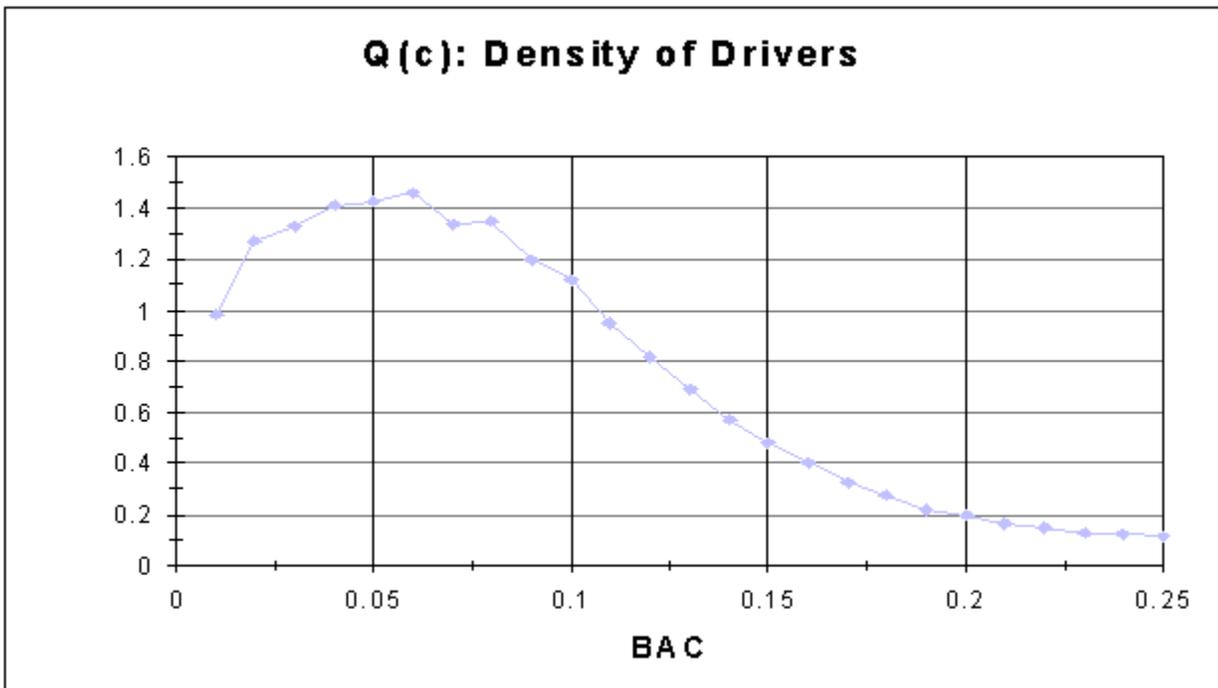
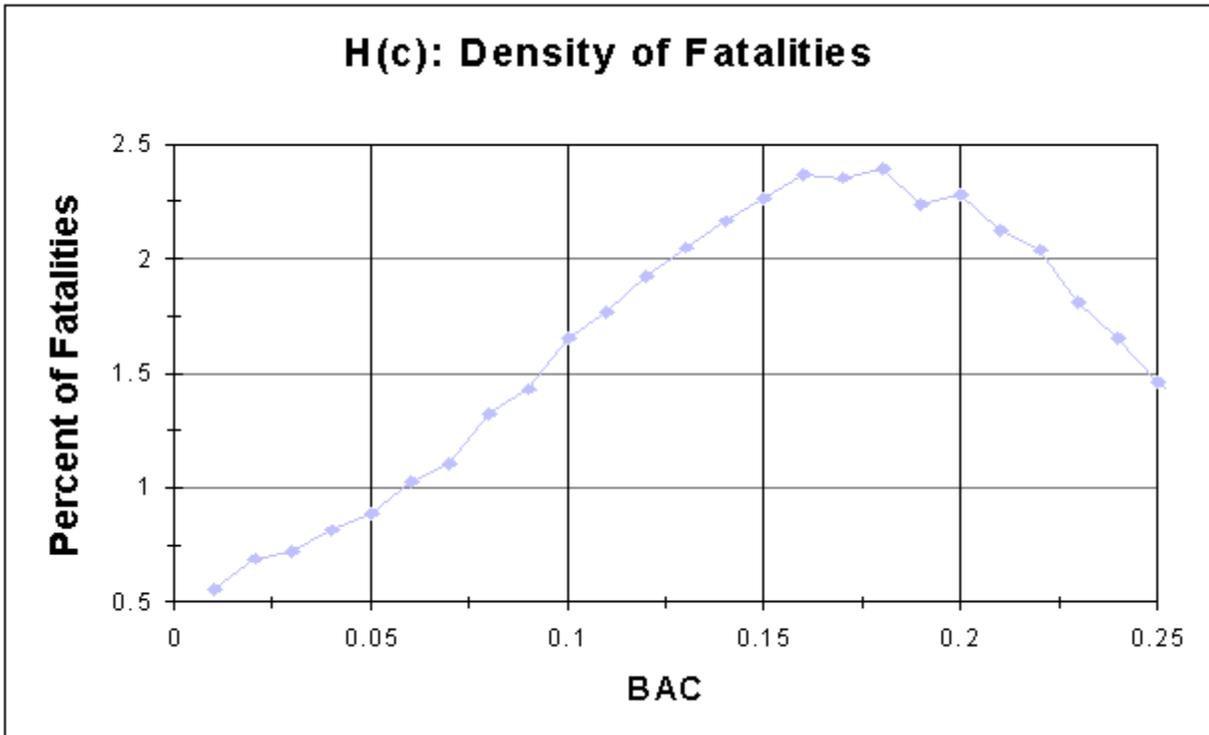


Figure 5: Legal Relationship between Alcohol Consumption and Expected Penalties.

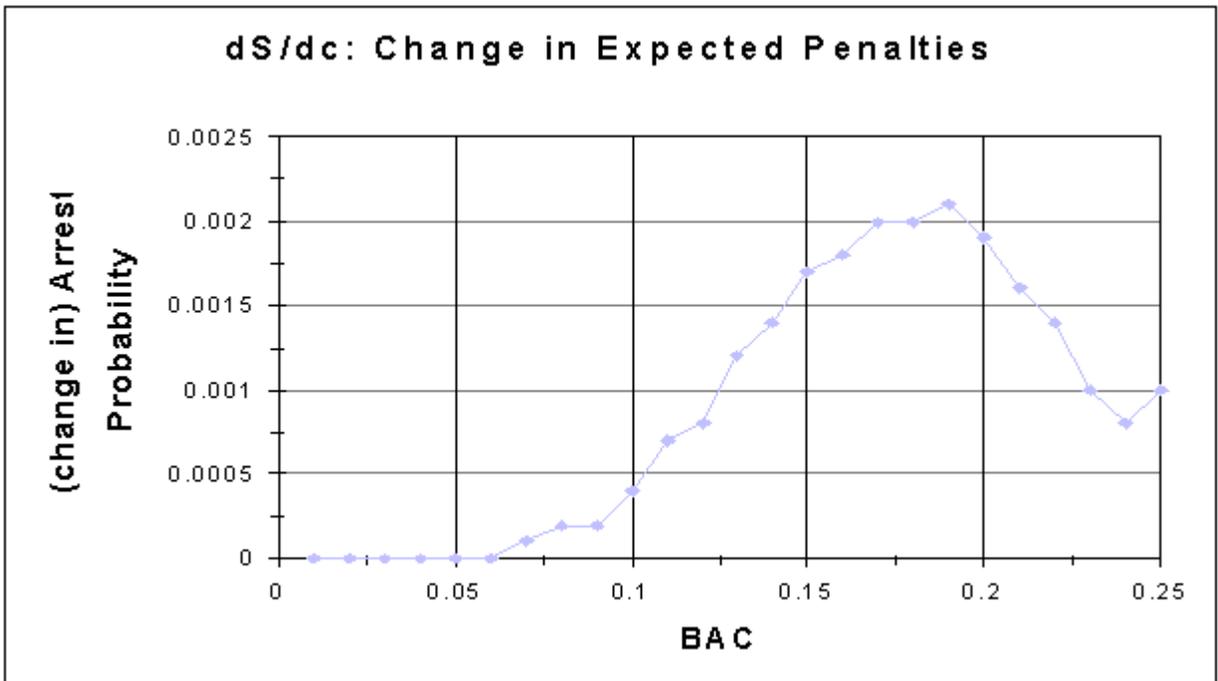
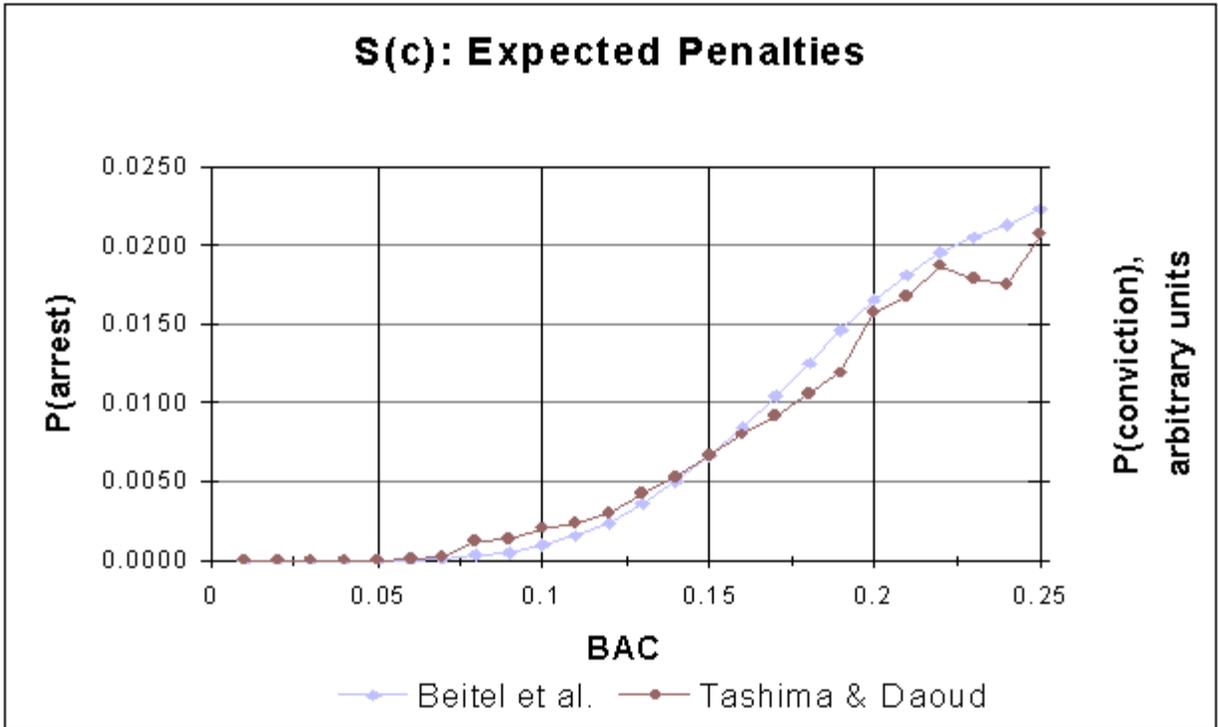
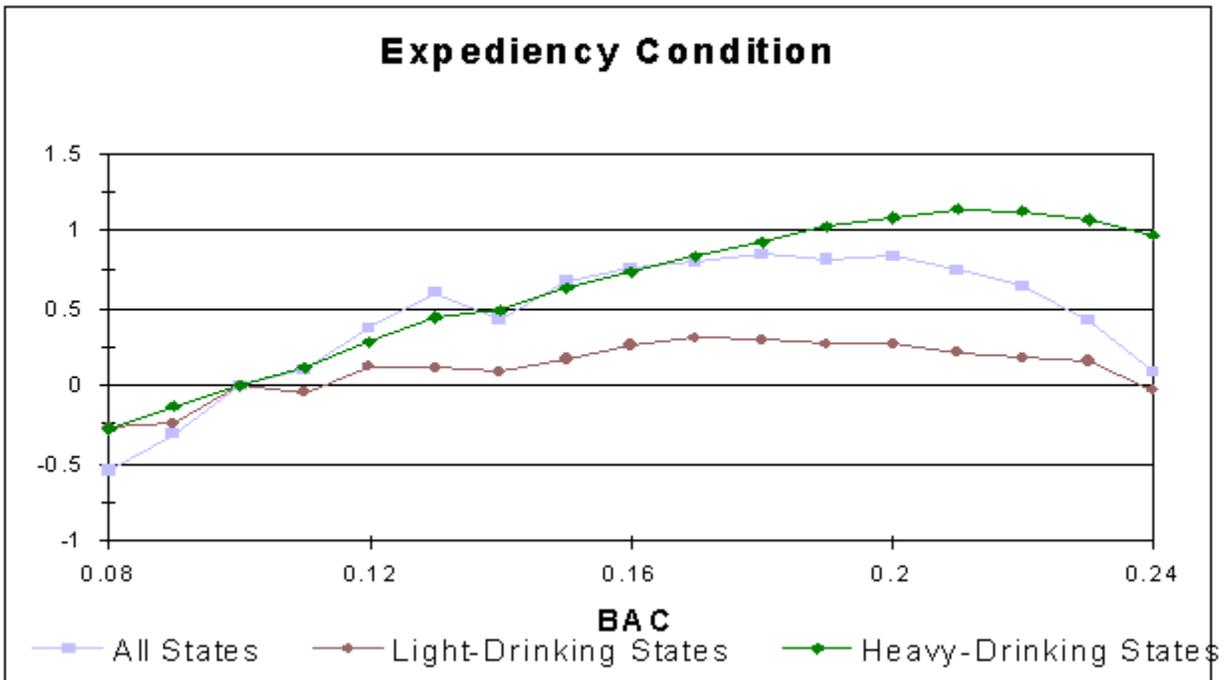
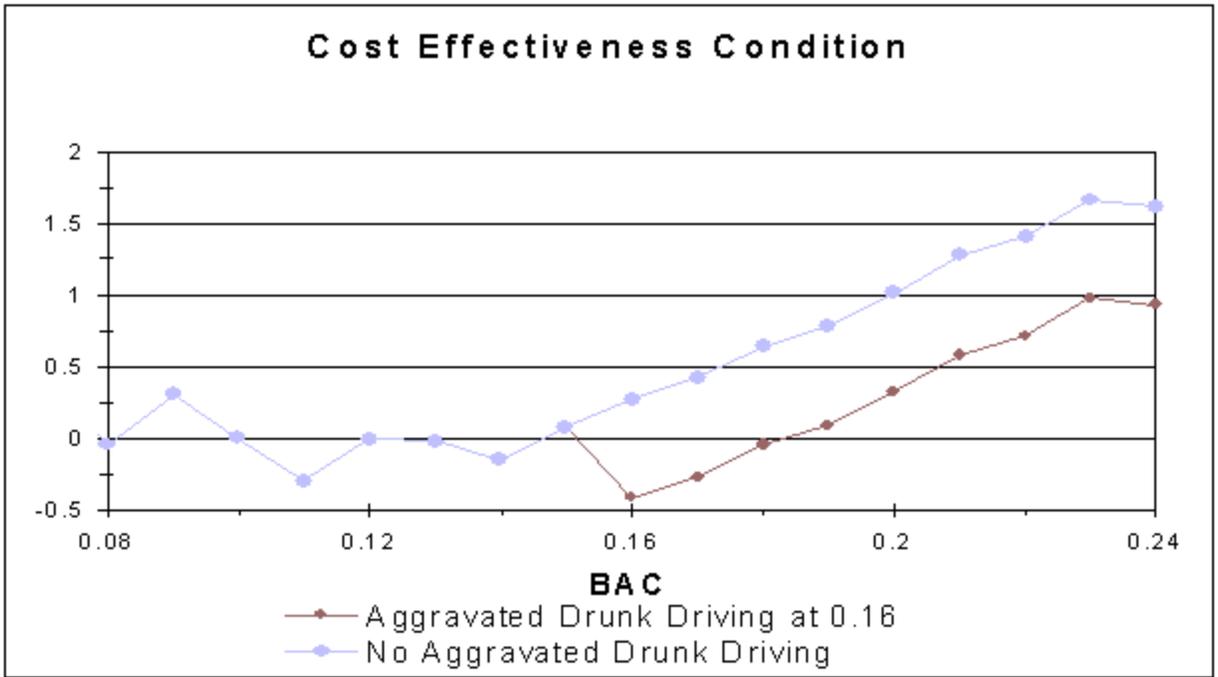


Figure 6. Cost-Effectiveness Condition and Expediency Condition.



REFERENCES

- Benson, B., D. Rasmussen, and B. Mast. "Deterring Drunk Driving Fatalities: An Economics of Crime Perspective," *International Review of Law and Economics* 19:205-225 (1999).
- Blomberg, R., R. Peck, H. Moskowitz, M. Burns, and D. Fiorentino. "Crash Risk of Alcohol Impaired Driving: A Case-Control Study." Manuscript, Dunlap and Associates (2005).
- Carpenter, C. "How Do Zero Tolerance Drunk Driving Laws Work?" *Journal of Health Economics* 23:61-83 (2004).
- Chaloupka, F., M. Grossman, and H. Saffer. "The Effects of Price on Alcohol Consumption and Alcohol-Related Problems," *Alcohol Research and Health* 26:22-34 (2002).
- Cox, R. "A Perverse Effect of Lowering the Threshold Blood Alcohol Content," *Applied Economics Letters* 13:869-871 (2006).
- Dee, T. "Does Setting Limits Save Lives? The Case of 0.08 BAC Laws," *Journal of Policy Analysis and Management* 20:111-128 (2001).
- Donovan, D., G. Marlatt, and P. Salzbeg. "Drinking Behavior, Personality Factors and High Risk Driving: A Review and Theoretical Formulation," *Journal of Studies on Alcohol* 44:395-428 (1983).
- Eisenberg, D. "Evaluating the Effectiveness of Policies Related to Drunk Driving," *Journal of Policy Analysis and Management* 22:249-274 (2003).
- Farrell, S., W. Manning, and M. Finch. "Alcohol Dependence and the Price of Alcoholic Beverages," *Journal of Health Economics* 22:117-147 (2003).
- Freeman, D. "Drunk Driving Legislation and Traffic Fatalities: New Evidence on BAC 08 Laws," *Contemporary Economic Policy* 25:293-308 (2007).
- Grant, D. "Dead on Arrival: Zero Tolerance Laws Don't Work," manuscript, University of Texas at Arlington (2006).
- Grosvenor, D., T. Toomey, and A. Wagenaar. "Deterrence and the Adolescent Drinking Driver," *Journal of Safety Research* 30:187-191 (1999).
- Hause, J., R. Voas, and E. Chavez. "Conducting Voluntary Roadside Surveys: The Stockton Experience," in Valverius, M., ed., *Proceedings of the Satellite Conference to the Eighth International Conference on Alcohol, Drugs, and Traffic Safety*, Stockholm (1982).
- Jacobs, J. *Drunk Driving: An American Dilemma*. Chicago: University of Chicago Press (1989).
- Kenkel, D. "Do Drunk Drivers Pay Their Way?" *Journal of Health Economics*, 12:137-149 (1993a).

- . "Drinking, Driving, and Deterrence: The Effectiveness and Social Costs of Alternative Policies," *Journal of Law and Economics*, 36:877-913 (1993b).
- . "New Estimates of the Optimal Tax on Alcohol," *Economic Inquiry*, 34:296-319 (1996).
- and S. Koch. "Deterrence and Knowledge of the Law: The Case of Drunk Driving," *Applied Economics* 33:845-854 (2001).
- Manning, W., L. Blumberg, and L. Moulton. "The Demand for Alcohol: The Differential Response to Price," *Journal of Health Economics* 14:123-148 (1995).
- Mirrlees, J. "Optimal Tax Theory: A Synthesis," *Journal of Public Economics* 6(4):327-358 (1976).
- National Highway Safety Transportation Agency. "On DWI Laws in Other Countries." Manuscript DOT HS 809 037 (2000).
- Perrine, M., R. Peck, and J. Fell. "Epidemiologic Perspectives on Drunk Driving," in *Surgeon General's Workshop on Drunk Driving: Background Papers*, U.S. Department of Health and Human Services (1988).
- Ruhm, C. "Alcohol Policies and Highway Vehicle Fatalities," *Journal of Health Economics* 15:435-454 (1996).
- Subramanian, R., and D. Utter. "Multiple Imputation of Missing Blood Alcohol Concentration (BAC) Values in FARS," NHTSA Technical Report DOT HS 808 816 (1998).
- Tashima, H. "Final Report to the Legislature of the State of California: The Relationship between Blood Alcohol Concentration Level and Court Sanction Severity in Drunk Driving Cases," manuscript available as NTIS PB87-123618 (1986).
- Tashima, H., and S. Daoud. "2007 Annual Report of the California DUI Management Information System," California Department of Motor Vehicles Report RSS-07-222 (2007).
- Trumbull, W. "Who Has Standing in Cost-Benefit Analysis?" *Journal of Policy Analysis and Management* 9(2):201-218 (1990).
- Wagenaar, A., P. O'Malley, and C. LaFond. "Lowered Legal Blood Alcohol Limits for Young Drivers: Effects on Drinking, Driving, and Driving-After-Drinking Behaviors in 30 States," *American Journal of Public Health* 91(5):801-804 (2001).
- Zador, P., S. Krawchuk, and R. Voas. "Relative Risk of Fatal and Crash Involvement by BAC, Age, and Gender." NHTSA Technical Report DOT HS 809 050 (2000). Selected results presented in "Alcohol-related Relative Risk of Driver Fatalities and Driver Involvement in Fatal Crashes in Relation to Driver Age and Gender: an Update Using 1996 Data," *Journal of Studies on Alcohol* 61:387-395 (2000).