Cell Phone Conversation and Automobile Crashes: Relative Risk is Near 1, Not 4

Richard A. Young

Wayne State University School of Medicine, Detroit, Michigan, USA

Abstract

The aim of research into cell phone tasks is to obtain an unbiased estimate of their relative risk (RR) for crashes. This paper re-examines five RR estimates of cellular conversation in automobiles. The Toronto and Australian studies estimated an RR near 4, but used subjective recall to estimate driving times. The OnStar, 100-Car, and a recent naturalistic study used objective measures of driving times and estimated an RR near 1, not 4 – a major discrepancy. Analysis of data from GPS trip studies shows that subjects were in-car only 20% of the time on a previous day, given they were in-car at the same clock time on a subsequent day. Hence, the Toronto estimate of driving time during control windows must be reduced from 10 to 2 min. Given a cell phone call rate about 7 times higher when in-car than out-of-car, and correcting for misclassification of some post-crash calls as pre-crash, the Toronto adjusted RR is 0.61, and the Australian 0.64, agreeing with the OnStar estimate of 0.62. After adjustment for bias, all five RR estimates for cellular conversation while driving in automobiles are near 1, with a pooled RR of 0.61 (95% confidence interval 0.51 to 0.74).

Introduction[1]

Part 1 reviews the data and biases in five previous epidemiological studies of the effect of cell phone conversation on the relative risk (RR)[2] of crashing while driving an automobile. These reviews inform Part 2, which develops new methods to remove bias in these studies, and re-analyzes their data to arrive at an unbiased estimate of the relative risk of cell phone conversation while driving an automobile.

Two of the studies used a "case-crossover" analysis, one a "cohort" analysis, and two others a "case-control" analysis.[3] For present purposes, it is only necessary to understand that, in the absence of bias, all three study designs should yield RR estimates close to the same value when risks are small (Rothman 2012a: 96). If the RR estimates are substantially different, then bias or error must be present in one or more of the studies, and needs to be removed to arrive at a valid estimate of relative risk.

---

[1] *Abbreviations and Symbols*: adj = adjusted; CI = 95% Confidence Interval; CNC = Crash/Near-Crash; GPS = Global Positioning System; HFC = OnStar Hands-Free Calling; N/A = Not Applicable; NDS = Naturalistic Driving Study; No Task = no secondary task; PAR% = Population Attributable Risk %; *PPF%* = Population Prevented Fraction %; P_e% = Percentage of Exposed Controls (an estimate of prevalence in the population); RR = Relative Risk (can refer to Risk Ratio, Rate Ratio, or Odds Ratio as an estimate of relative risk); ρ = ratio of in-car to out-of-car call rate; Call = talking/speaking over a wireless communication device; VTTI = Virginia Tech Transportation Institute.

[2] Under the "rare-disease assumption" (Porta 2008: 207), the risk ratio, rate ratio, and odds ratio all approximate one another; hence, the term "relative risk" will be used here as a synonym to represent any of them, as in common epidemiological practice (Rothman 2012a: 96).

[3] These epidemiological terms may be unfamiliar to many in traffic safety See Rothman (2012a: chapter 5) for an introductory description.

PART 1: Review of Epidemiological Studies of Calls and Crashes and Their Biases

1.1 Toronto Study Review

The Toronto study (Redelmeier and Tibshirani 1997) analyzed the cell phone billing records of 699 study subjects who crashed while driving. It found that 170 of them had calls in the 10-minute *case window* (just before the estimated time of their crash), and 37 of them had calls in the 10-minute *control window* (during the same clock time the day before the estimated time of their crash).[4] This study used a case-crossover analysis method, which estimates the RR with the *discordant pair*,[5] which here is subjects with a call in the case window (but not in the control window), and subjects without a call in the case window (but with a call during the control window).

Table 1A shows the tabulated data from Tibshirani and Redelmeier (1997: Table 4). The first two columns are the standard 2x2 matrix used for a case-crossover analysis. The discordant pair is the cells containing 157 and 24. Dividing those numbers estimates a "crude" (i.e., unadjusted for bias) risk ratio of 6.5, shown with the 95% confidence interval (CI) from 4.5 to 10 as reported by the Toronto study.

Table 1 Toronto study call counts. A. Crude RR estimate. B. RR estimate corrected for driving intermittency bias by the Toronto investigators

A. Crude		Control Window[b]			
		Call (in-car)	No Call (in-car)	(out-of-car)	Total
Case Window[a]	Call	13	157	0	170
	No Call	24	505	0	529
	Total	37	662	0	699
	RR (95% CI)	**6.5**	(4.5 to 10)		
B. Corrected for Driving Intermittency					
Case Window[a]	Call	13	102	55	170
	No Call	24	505	0	529
	Total	37	607	55	699
	RR (95% CI)	**4.3**	(3.0 to 6.5)		

Notes: [a]10 min duration before crash, day of crash [b]10 min duration, day before crash

Driving Intermittency Bias

The Toronto investigators correctly realized that they had to adjust for bias from *driving intermittency* (i.e., drivers who crashed, but who did not drive at all in a control window on a previous day). The Toronto investigators had the complete billing records making it possible for them to know the call rates during a control window vs. not. Subjects who did not drive during a control window would have a lower call rate than subjects

[4] It is likely that more than one call occurred in a given 10-minute window. If so, these would have been tabulated as a single count by the Toronto methods. Thus, technically the counts in Table 1 are not counts of calls, but rather counts of subjects who engaged in at least one call in a given window.

[5] *Discordant* in a matched case-control (or case-crossover) study describes a pair whose members had different exposures to the risk factor under study. Under conventional analytical methods, only the discordant pairs are informative about the association between exposure and disease (Porta 2008: 68).

who did drive there (Young 2012a),, causing a smaller RR denominator and incorrectly elevating the RR The elevated RR is unrelated to call risk; it would simply mean that subjects drove in the case window but did not in the control window. The same imbalance in calls between any driving period vs. any non-driving period would be observed, even with no crash. The Toronto investigators therefore realized that driving intermittency was a major confounding variable, and attempted to correct for it. To do so, they decided to discard the subjects who did not recall driving in a control window. However, they had no objective data as to whether a subject actually drove or not in a control window. Therefore, they surveyed another group of 100 respondents (not in the original study) and found that 35 percent of them remembered not driving during a "selected period" the day before. The Toronto investigators then multiplied all their RR estimates by 0.65 to correct for this driving intermittency. Table 1B illustrates the correction, which is the equivalent of discarding 55 of the 170 subjects because they were out-of-car in the control window. The RR estimate dropped from 6.5 to 4.3 (102 divided by 24), which was the final reported RR.[6]

Unadjusted Biases in the Toronto Study

Part-time driving bias

The Toronto study neither recognized nor controlled for *part-time* driving bias, which arose from those who drove in a control window, but drove only part of the time there. This part-time driving reduced their driving exposure, leading to a lower number of calls in the RR denominator and erroneously inflating the RR (Young 2011a; 2012a,b,c,d; 2013a). Part-time driving bias is not controlled for by the driving intermittency correction in the previous section. Discarding those with no driving at all in a control window does nothing to correct for those who drove there, but for only part of the time. Part-time driving bias is an additional bias besides driving intermittency, and was neither recognized nor controlled for in the Toronto study. Part-time driving upwardly biased all RR estimates in the Toronto study, even after corrections for driving intermittency (Young 2011a; 2012a,b,c,d; 2013a).

Misclassification bias

Another bias arises from misclassification of calls as occurring before the crash, when in fact they occurred after. These call misclassifications arise from errors in estimating crash times. Redelmeier and Tibshirani (1997) recognized this misclassification bias and attempted to correct for it by various means. They did not have objective information on crash times, and so had to rely in large part on police reports. However, these tend to be rounded to the nearest 5-minute mark (Baker 1971), causing misclassification errors when comparing the times in cell phone bills. The Toronto study also used times of calls to emergency services if they appeared in the billing records. However, the billing records showed a sharp increase in calls of all kinds after the crash, not just to emergency services. Some of those calls could have occurred after the crash but before the call to emergency services; leading to their being misclassified

[6] Another method used to estimate the intermittency factor was a survey of the actual subjects a year after the study, asking each subject to remember his or her driving pattern on both the day of the crash and the day before the crash the year before. Among those who responded (the number was not reported in the paper), the investigators limited their analysis to only those subjects who were confident that they had driven during both windows on both days, and obtained a similar RR.

as pre-crash when in fact, they were post-crash.

Usually in epidemiological studies, misclassification bias occurs equally in both directions, and so biases the RR estimates towards the null (i.e., baseline). In call and crash association studies, however, the imbalance in calls after the crash compared to before the crash leads to a differential misclassification bias due to the error in the crash time estimates, where the RR is specifically biased upward (Young and Schreiner 2009a,b; Quinlan 1997; Bellavance 2004, 2005). Bellevance (2004, 2005) developed a model for this misclassification bias in the Toronto study and estimated that it inflated the RR by a factor of 3.

1.2 Australian Study Review

McEvoy et al. (2005) interviewed 744 subjects who gave permission to access their cell phone billing records in a hospital emergency room after a crash in which they were driving and had been injured. They were asked to remember how long they had been driving at the time of the crash up to a maximum time window of 10 min before the crash. The time of crash was estimated from emergency response records, medical records, and subject self-report. The subjects were also asked if they remembered driving during control windows of the same duration and clock time as the crash window, but at 1, 3, or 7 days before the estimated time of the crash. A total of 456 (61%) of the subjects remembered driving during one or more of the three control windows.

From the cellular billing records of those 744 subjects, the investigators found that 40 of them (9%) had a cell phone conversation[7] sometime up to 10 min before the crash during the period in which they said they were driving. If the remembered driving time was less than 10 min, the call was counted only if it occurred during the remembered driving time, relative to the estimated time of the crash. The billing records were then examined for those control periods that the subjects remembered driving in, for the time duration the subject remembered as having driven before the crash. The driving intermittency correction was hence built into the study design and the RR did not need to be corrected after the fact as in the Toronto study.

The results showed that 25 subjects had engaged in a call in at least one of the control windows at 1, 3, or 7 days before the estimated time of the crash. Using a conditional logistic regression method, the Australian study estimated an RR of 4.1 (95% CI 2.2 to 7.7), consistent with the Toronto RR adjusted estimate of 4.3.

The current study could not replicate the conditional logistic regression analysis used in the Australian study with the data provided, because it requires knowing, for each driver, whether he or she called or not in the case window and during each control window. In addition, the data for the discordant pairs[5] were not provided, and these are required for a case-crossover analysis. Therefore, the discordant pairs for each control period were solved for algebraically by using the call counts in its Table 2 plus its RR estimates.

---

[7] Note that billing records contain information only about cell phone conversations and no other aspect of cell phone use, such as reaching for a phone, looking up phone numbers, dialing, etc., so this study and the Toronto study can only make conclusions about conversation and not other aspects of phone use.

Table 2 gives the 1, 3, and 7 day solutions and the combined solution. It can be seen that these solutions adequately replicated the more complex conditional logistic regression analysis results of McEvoy et al. (2005). It would have been desirable to have the 2x2 cell entries from McEvoy et al. directly, but the marginal totals are exactly the same as McEvoy et al. (2005: Table 2), and the RR estimates are the same to the 2 significant digits for the 1, 3, and 7 day solutions that were provided by McEvoy et al.

Table 2 **Australian study call counts, no driving intermittency**

		Control Window[b]		
A. 1 Day Before		Call	No Call	Total
Case Window[a]	Call	4	22	26
	No Call	6	216	222
	Total	10	238	248
RR (95% CI)	Exact	**3.7**	(1.4 to 11.0), *p* = 0.0025	
	McEvoy et al.	**3.7**	(1.5 to 9.0)	
B. 3 Days Before				
Case Window[a]	Call	1	14	15
	No Call	3	209	212
	Total	4	223	227
RR (95% CI)	Exact	**4.7**	(1.3 to 25.3), *p* = 0.0076	
	McEvoy et al.	**4.7**	(1.3 to 16.2)	
C. 7 Days Before				
Case Window[a]	Call	5	27	32
	No Call	6	288	294
	Total	11	315	326
RR (95% CI)	Exact	**4.5**	(1.8 to 13.3), *p* = 0.0003	
	McEvoy et al.	**4.5**	(1.9 to 10.9)	
D. Pooled				
Case Window[a]	Call	20	20	40
	No Call	5	411	416
	Total	25	431	456
RR (95% CI)	Exact	**4.0**	(1.5 to 13.6), *p* = 0.0027	
	McEvoy et al.	**4.1**	(2.2 to 7.7)	
	Pooled	**4.2**	(2.1 to 8.4)	

[a]up to 10 min duration before crash [b]up to 10 min duration,1, 3, or 7 days before
[c]Stata Exact solution [d]McEvoy et al. 2005: Table 2

Unadjusted Biases in the Australian Study

The unadjusted biases in the Australian study were the same as those in the Toronto study, as it neither controlled for part-time driving in control windows, nor did it have objective crash times (and so was also subject to call misclassification bias).

1.3 OnStar Study Review

Young and Schreiner (2009a) conducted a cohort study of the OnStar Hands-Free Calling (HFC) embedded device, which makes only personal calls, not OnStar advisor calls (which had been previously studied by Young, 2001). Like the 2001 advisor call study, the 2009 study employed objective data on driving and call times. Unlike the 2001 study, it had a baseline control – counts of those who made calls without an airbag deployment crash, and so it could calculate relative as well as absolute crash risk. The start and end times of the phone calls for both the automatic emergency calls (which went to known numbers reserved for that purpose), and the personal calls were obtained from de-identified cell phone billing records for the OnStar calls from the cellular provider. There were a total of 30 months of driving by 3 million individual OnStar personal calling subscribers in the U.S. and Canada. There were 91 million personal calls made by an average of 323,994 drivers per month. During the study period, there were 14 crashes severe enough to deploy an airbag[8] during the 524.02 person-years[9] of personal conversation while driving. There were 2,023 airbag-deployment crashes during an estimated 47,125 person-years[10] of driving with no personal conversation.

Table 3 shows the resultant airbag-deployment crash rates per person-year for Call and No Call. The RR is 0.62, indicating that personal conversations using OnStar HFC do not increase the risk of a crash severe enough to deploy an airbag relative to the combined effect of all other activities in normal driving.[11]

Table 3 OnStar Hands-Free Calling (HFC) study data and rate ratio analysis (based on data in Young and Schreiner 2009a: Table III)

	Call (in-car)	No Call (in-car)	Call (out-of-car)
Airbag-deployment crashes	14	2,023	0
Person-years	524.02	47,125	762,542
Rate	0.02672	0.04293	0
RR (95% CI)	0.62	(0.34 to 1.05)	
Rate Difference (95% CI)	**-0.0162**	(-0.0303 to -0.0021)	

Unadjusted Biases in the OnStar Study

Portable cell phone usage in unexposed condition

Braver et al. (2009) argued against the validity of the Young and Schreiner (2009) OnStar result by hypothesizing that: (1) OnStar users may have used portable cell phones while driving when they were not using OnStar, and (2) that such behavior made normal driving periods more risky and hence made OnStar use look relatively less risky

[8] The Toronto study severity criterion was property-damage-only, and the Australian study was being seen in a hospital emergency room, but neither study indicated how many of their crashes had airbag deployments, and the OnStar study had no data on injuries or property damage, so it is indeterminate to what extent the OnStar airbag crash severity criterion overlapped with those other studies.

[9] These were converted to person-years from person-minutes in Young and Schreiner (2009a: Table III).

[10] For comparison, the 47,609 driving-years in the Young and Schreiner (2009a) OnStar study is 12.5 times larger than the expected 3,800 driving-years of the Strategic Highway Research Program 2 Naturalistic Driving Study at its scheduled completion date in 2014.

[11] Although the RR CI overlapped 1, that for the risk difference did not, providing some evidence that OnStar HFC conversations reduce the risk of crashes severe enough to deploy an airbag.

(that is, it biased the OnStar RR low).[12] To be valid, these hypotheses require that portable cell phone usage increases crash risk relative to the embedded OnStar device, and that subscribers to the OnStar device use their portable phones while driving in lieu of the OnStar device.

Young and Schreiner (2009b) agreed in principle, noting that their original article pointed out (p. 187) that embedded hands-free devices have better human factors properties than portable cell phones (e.g. higher quality in-vehicle speakers and microphone; improved signal-to-noise ratio because of an antenna outside the vehicle). Hence, they hypothesized that embedded devices should reduce crash risk relative to portable devices. Indeed, Young and Schreiner (2009a) showed that OnStar use does not increase crash risk (and indeed reduces the absolute number of crashes severe enough to deploy an airbag), relative to their baseline, which represents the combined effect of all forms of inattention during the 47,125 driving-years in the "No Call" in-car baseline in their study (see Table 3).

It would be desirable to test the Braver et al. (2009) hypotheses with direct data from the OnStar cohort. However, unlike a naturalistic driving study (NDS), the OnStar databases contain no information on other secondary tasks that may have been performed in the vehicle, and so the Braver et al. (2009) hypotheses cannot be directly tested with the available OnStar data.

However, Fitch et al. (2013) in a recent NDS of cell phone usage while driving found direct evidence against both the Braver et al. (2009) hypotheses. They found that conversation on hand-held or hands-free portable cell phones had about the same RR near 1 as did embedded hands-free cellular devices such as OnStar. In addition, drivers who were users of one type of cellular device (portable hands-free, portable hand-held, or embedded hands-free) rarely used another.

Some secondary tasks (such as manual dialing of a hand-held cell phone) do increase the odds of a crash, and were likely present in the OnStar baseline data. However, other secondary tasks reduce the odds of a crash, so that the pooled net RR may be near 1, the same as baseline driving without performing those tasks. Indeed, Young (2013b) found that the pooled RR was near 1 for nine commonly-performed secondary tasks (including phone dialing) in the 100-Car NDS, after removing biases in the original RR analyses (Klauer et al. 2006).

Therefore, the hypotheses of Braver et al. (2009) can be rejected based on recent NDS evidence. This evidence indicates that negligible change would be expected in the Young and Schreiner (2009a) RR estimate of 0.62 for personal conversations using the OnStar device whether or not portable cell phone usage were included in the unexposed condition.

Demographic variables

A valid limitation noted by Young and Schreiner (2009a: 200) about their own study is that although it was a nationally representative sample, it did not control for bias from confounding by demographic variables.[13] Stratifying the data by at least age and sex is

[12] Young and Schreiner (2009a) did not have access to portable cell phone billing records and so could not compare embedded cell phone usage to portable cell phone usage.

[13] Young and Schreiner (2009) recognized this limitation prior to publication but were unable to obtain the de-identified demographic data about OnStar subscribers in time to meet the publication date.

required in the analysis of any data collected in a cohort or case-control study design, in order to control for possible bias from those variables. (Such stratification is not required to control for bias in case-crossover designs, which inherently control for demographic variables by using each subject as his or her own control, but it is done in case-crossover studies not to control for confounding factors, but to look at differences of interest between population subgroups.) Bias from confounding by demographic variables can raise or lower crude RR estimates by unknown amounts in any study that does not control for such bias.

1.4 100-Car Study Review

Naturalistic driving studies are based on non-experimental real-world driving and include real-time vehicle data (e.g., speed, braking, and accelerator inputs) as well as video recordings of the driver and roadway. They are particularly useful for cell phone studies because they provide direct, objective driving and calling data associated with safety-relevant events for each study subject, information available in the 100-Car NDS and OnStar study, but not in the Toronto or Australian studies.

Specifically, the 100-Car study (Klauer et al. 2006) examined video data from 241 drivers in about 100 vehicles over a period of one year. It reviewed data on 830 crash and near-crash events, and compared video clips of a driver's behavior at the time of these events to that in randomly-selected video clips from baseline periods driving without a crash, near-crash, or incident of less severity than crashes and near-crashes. It used a case-control analysis to compare the odds of exposure of drivers to cell phone conversation before a crash/near-crash to the odds of exposure to that task in baseline periods. The crash/near-crash RR estimate was 1.29 for a hand-held cell phone conversation, as shown by the 100-Car data in Table 4.

Table 4 RR estimate by Klauer et al. (2006) for hand-held cell phone conversation comparing Call vs. No Task for 100-Car NDS with unmatched baseline. Data from (Hankey 2007: Table 6)

	Call	No Task	Total	Odds	Prevalence
Crash/Near-Crash	44	237	281	0.1857	N/A
Unmatched Baseline	1,299	9,059	10,358	0.1434	N/A
RR (95% CI)	1.29	(0.93 to 1.80)			

Unadjusted Biases in the 100-Car Study

- It used a different driver fault condition for the "Call" and "No Task" crashes and near-crashes, which biased its RR estimates high (Young 2013c).[14]
- It used a non-standard "No Task" condition (meaning "no secondary task") rather than the standard "No Call" for the unexposed category,[15] which biased the prevalence and Population Attributable Risk % (PAR%) estimates high, so the prevalence estimates are shown as "N/A"

[14] The 44 cases in the "Call" group in Table 4 were all crash/near-crash cases where "Call" occurred, regardless of subject driver fault. The 237 cases in the "No Task" group were restricted to cases where the subject driver was at fault with no observed secondary task.

[15] The "No Task" category does not include other secondary tasks, but does include drowsiness, and so does not represent "attentive."

in Table 4 because they are incorrect (Young 2013c).

- It collapsed crashes and near-crashes into a single group. Although near-crashes are also a safety concern, the RR estimates are uniformly lower for secondary task involvement in near-crashes than for crashes (Guo et al. 2010).[16] Hence, combining these lowers the RR estimate vs. crashes alone, giving the false impression that a task may be safer than it really is. For example, combining crashes and near-crashes into one group underestimates the RR for drowsiness as a cause of crashes (Young 2013c).
- It did not adjust its RR estimates by demographic variables, even though it presented data showing that crash/near-crash prevalence varied by age and sex. The RR estimate of 1.29 in Table 4 could be higher or lower after adjustment for these variables.

1.5 Cell Phone NDS Review

Fitch et al. (2013) used instrumented passenger vehicles to record driving behavior for 204 volunteer subjects for an average of 31 days, to investigate their calling, driving, and safety-critical event behavior. It is the first known study to publish data obtained from cell phone billing records for both in-car and out-of-car calls by the same drivers. A total of 187 of the 204 drivers provided their cell phone records for analysis, which were matched to driving times as determined by both GPS and video recordings. Subjects were stratified into those who primarily used portable hand-held, portable hands-free, or embedded hands-free cell phones.

Fitch et al.'s (2013) study was also the first NDS to use a new method for selecting which video clips to analyze for the baseline comparison data. Fitch et al. (2013: xxvi) used a 20-s baseline video clip sample "selected 30 s prior to the start of each sampled cell phone interaction time period."[17] (All other NDS methods to date have used baselines from previous days.) This idea was first proposed by Maclure and Mittleman (1997) in an editorial review of the Toronto study:

> Another way to adjust for the intermittency of driving and at least partially for telephone use outside cars is to compare the risk 1 to 5 minutes after a call with the risk 16 to 20 minutes after it. This adjustment also partially corrects for additional confounding factors, such as weather or traffic conditions that may increase both telephone use and the frequency of collisions.

Young (2011b) was the first to recommend this improved baseline method specifically for use in NDS data analysis. He pointed out that selecting the baseline video clips from the driving time that was before the case video clips during the same trip on the day of the crash, controls for driver, vehicle, weather, and traffic conditions,

---

[16] That is, drivers who crashed had a higher odds of being exposed to the secondary tasks examined by Guo et al. (2010) than those who had near-crashes. This may seem contradictory at first, because there are 10 times more near-crashes than crashes. However, the exposure odds are the probability of the occurrence of an exposure to that of non-occurrence of the exposure. So the RR is estimated as a ratio of those odds, and so it is not a count, and is independent of the absolute number of crashes.

[17] Fitch et al. (2013) were concerned that their results might not be representative of the general public, because only drivers who reported talking on a cell phone while driving at least once per day were recruited, which the investigators say is a relatively high cell phone use rate. However, their baseline sampling method uses each subject as his or her own control, so the absolute rate of cell phone usage per subject has no effect on the RR estimate, just as no other individual difference would.

as well as all other factors that are generally constant for a given trip, such as presence or absence of passengers and time of day.[18]

This baseline method follows a case-crossover design in sampling the baseline video clips. However, Fitch et al. (2013) and Klauer et al. 2010 used a case-control method to analyze their data, and not the case-crossover method, which considers only the discordant pairs.[5] Nonetheless, a case-control analysis method used with matched baseline still has all the advantages of reducing bias from many confounding variables as listed above. However, if a case-crossover analysis method had been used, it would have had more sensitivity than the case-control method to capture small effect sizes.

Table 5 presents the Fitch et al. (2013: Table 2) "Call" data for vehicles moving more than 8 km/h using this new method for matching baselines. Fitch et al. (2013) estimated an RR of 0.75 for conversation combined across hand-held, portable hands-free, and embedded hands-free wireless devices, and combined across crashes, near-crashes, and crash-relevant conflicts.[19] In Table 5, Fitch et al. (2013) compared "Call" to "No Task," a non-standard method that causes errors in any prevalence and population percent estimates that are made from such an analysis method, as shown by Young (2013a). Therefore, these estimates are not made in Table 5, which labels these them as N/A for *Not Applicable*.

Table 5 RR estimate by Fitch et al. (2013: Table 2) for NDS study of cell phone conversation comparing Call vs. No Task with matched baseline

	Call	No Task	Total	Odds	Prevalence
Safety-Critical Event[a]	28	154	182	0.1818	N/A
Matched Baseline	259	1,068	1,327	0.2425	N/A
RR (95% CI)	0.75	(0.49 to 1.15)			

[a]Crashes (including curb strikes), near-crashes, and crash-relevant conflicts

Unadjusted Biases in the Cell Phone NDS

Heterogeneity between severity levels

The study did not test for heterogeneity between crash-relevant conflicts, near-crashes, and crashes before combining them into one group. If there was heterogeneity, then combining those strata into the single group of "safety-critical events" would be misleading at best or invalid at worst (Young 2012e).

As a hypothetical example, the Fitch et al. (2013) study cannot reject the possible heterogeneity that a cell phone task might have between crash severity levels; e.g., an RR above 1 for the crashes and near-crashes, but an RR below 1 for safety-critical incidents. When summed into one group, these opposite effects, which may be important and of possible interest for safety issues, would not be seen.

[18] This case-crossover design is far less resource-intensive than the Klauer et al. (2010) case-crossover method that attempts to match baseline video clips on previous days to case video clips by driver, time of day, GPS location, number of lanes, city vs. highway driving, traffic conditions, weekday/weekend, etc.— a difficult task, and impossible for a trip that was never repeated.

[19] The intersection data, although novel for an NDS, were too noisy and were not included in many of Fitch et al.'s RR estimates.

Even if these three severity levels were somehow homogeneous, Guo et al. (2010) showed that crashes have higher RR estimates than near-crashes for every secondary task they examined, thus biasing downward any RR estimate based on combined crashes and near-crashes. Including crash-relevant conflicts would likely bias the RR even lower. Thus, heterogeneous subgroups should not be combined (Young 2012e).

False negative errors from low number of crashes

Data were collected for only one month rather than a year or more as in other NDS datasets, and resulted in only four crashes. Only 4 events likely lacks sufficient sensitivity (i.e. statistical power) to properly assess the risk effects of cell phone tasks on crashes. The study should have clearly stated this limitation, as it could lead to false negatives (i.e., claiming that the crash risk of a cell phone task risk is no different from baseline risk when in fact it is).

This issue can be overcome if severity levels are homogeneous, and can be merged to create a dataset with a larger number of safety-related events, which is a widely-adopted strategy in NDS analysis. However, such grouping is commonly done without a test for heterogeneity. Thus, there could be an issue with a secondary task that is present in crashes but not near-crashes or other lower severity levels. If the crash events are simply merged with the lower severity events, the crash effect will be "washed out" from the larger number of lower-severity events and a false negative error will occur.

Other unadjusted biases in the cell phone NDS

- It did not test for heterogeneity between subtasks before grouping those subtasks into the larger overall groups for each cell phone type.
- It used "No Task" in all its RR calculations that biases baseline prevalence and population percentage estimates upwards (Young, 2013b).
- It presented results for a rate ratio analysis that correctly used the complement of the exposure (e.g., "No Call" instead of "No Task"), but from the information given in the report, it was not possible to replicate those rate ratios (Fitch et al. 2013: Table 22). There is concern of an unknown bias or error in the study's rate ratio estimates because after re-calculating the RR estimates to use "No Call" (e.g. Part 2, Section 5), they did not agree with the rate ratios (Fitch et al. 2013: Table 22), as they should have.[2]
- Placing the control window only 30 s in advance of the case window may cause the control window to be upwardly biased in its call counts, because of the tendency of people to make several calls in a row (which is evident in a close examination of large numbers of billing records). This tendency would bias the RR low.
- The 30 s gap between control and case windows may bias low the estimated RR for the conversation subtask of conversation. There are multiple subtasks involved before a conversation subtask begins, which could "spill over" into a baseline that was too close to the conversation subtask. These include: possible cognitive processes involved in planning the call (e.g. retrieving the phone number from memory), the possible eyes-off-road time involved in searching for a portable phone, looking up a phone number in an address book, and dialling by any method. Although some of these subtasks were separately analyzed in the

report, it is not clear from the methods what baselines were used for which subtasks, and what baselines were used for each subtask when they were performed in close succession. It is not clear that a 30 s gap would prevent these preparatory tasks from being included in the baseline for the conversation portion of the call, biasing the RR low.

PART 2: New Analyses of Call-Crash Data from Prior Epidemiological Studies

2.1 Objectives

The aim of Part 2 is to develop an unbiased estimate of the RR of cell phone conversation for automobile crashes. In lieu of conducting a new NDS, which is expensive and time-consuming, it is more efficient to eliminate as much bias as possible from previous studies through new analyses.

Evidence of bias in one or more of the studies discussed in Part 1 is apparent. The Toronto and Australian studies found an RR near 4 in passenger vehicles, while the OnStar, 100-Car, and Fitch et al. (2013) studies found an RR near 1, a major discrepancy.[2] A minor discrepancy is that the Klauer et al. (2006) 100-Car study estimated a Call RR above 1 (1.29), while the Young and Schreiner (2009a) OnStar study and the Fitch et al. (2013) cell phone study estimated Call RRs below 1, in the opposite direction thus indicating a preventive[20] rather than causal effect.

The current study therefore investigated the root cause of these discrepancies, extending and improving previous analyses (Young 2011a; 2012a,b,c,d,e; 2013a), in order to arrive at an unbiased estimate of the RR of cell phone conversation for automobile crashes.

2.2 Methods to Estimate and Adjust Bias in Call-Crash Studies

Adjustments and corrections were made to 4 of the 5 studies in Part 1, in an effort to remove the remaining biases that were identified regarding their estimates of the cell phone conversation RR while driving. These adjustment methods have been previously described (Young 2011b; 2012a,b,c,d,e; 2013a,b). The methods are applied in Part 2 to each study in Part 1, along with the results of those adjustments on each of the relative risk estimates of those studies.

The risk ratios for the Toronto and Australian studies in Tables 1 and 2 were converted into rate ratios, to allow for easy adjustment for part-time driving bias in the control windows. An equation to predict RR based on the in-car to out-of-car calling rate ρ value is shown, as first described by Young (2012e,f; 2013a). The adjustments for bias in the Klauer et al. (2006) 100-Car study were also accomplished using methods previously described, most recently by Young (2013b). Data from the recent NDS on cell phone usage and driving (Fitch et al., 2013) were next analyzed using those same methods (Young 2013b). The statistical and epidemiological analysis software package

[20] The term *protective* could be used here as a synonym for *preventive*, but preventive is used as it refers to the technical term in epidemiology *Prevented Fraction (Population)*, or *Population Prevented Fraction %* (*PPF%*). In a situation in which exposure to a given factor is believed to protect against an outcome (such as a crash), the *prevented fraction* is “the proportion of the total load” of the crashes in the population that has been prevented by exposure to the factor. Porta (2008: 192) cautions that, “This measure must be interpreted with caution, as part or all of the apparent protective effect may be due to other factors associated with the apparent protective factor.” (Porta 2008: 192).

Stata/IC 12.1 was used to calculate exact confidence intervals. The methods will be illustrated using the Toronto data; the Results section then applies those methods to the Toronto data in detail to illustrate how the methods are applied, and then more succinctly to that of the other studies discussed in Part 1.

2.2.1 Method to Calculate Part-Time Driving Bias

The adjustments for driving intermittency undertaken in the Toronto study are correct, but do not adjust for part-time driving, which is entirely different from driving intermittency as defined in the Toronto study and discussed in Part 1, Section 1.1. *Part-time driving* is defined as the percentage of time driven during a previous control window, given that some driving occurred in that control window, whereas *intermittency* refers only to not driving at all in a previous control window. If a subject did not drive at all in the control window, the part-time driving percentage is 0%, and the driving would be classified as "intermittent," and the two metrics would agree on a common method of adjustment. If a subject drove for the entire 10 minutes in a 10-minute control window, the part-time driving percentage is 100%, and the driving would not be classified as "intermittent," and again the two metrics would agree. However, if a subject drove for only 2 minutes in a 10-minute control window, the part-time driving would be only 20%, but the Toronto and Australian study methods would not classify that driving as "intermittent," and would treat the subject as having driven for the entire 10 minutes, when their actual driving exposure was only 2 minutes. The bias from such part-time driving in control windows was not recognized or controlled for by the Toronto or Australian correction methods.

Part-time driving bias is in addition to the driving intermittency bias described in Part 1, Section 1. Another example to demonstrate the difference is that even with no driving intermittency, there could still be only 10% driving exposure in the control periods, from part-time driving there. Assume every driver who crashed drove for only 1 minute in the 10-minute control window on a previous day. Further, assume all of subjects interviewed by the Toronto investigators remembered that they drove in that control window, so all of them would have been included in the Toronto study sample. However, there would only be 10% actual driving exposure in the control windows because the subjects drove only 10% of the time in those windows. There would be no driving intermittency at all, and all those people would be assumed (incorrectly) by the Toronto methods to have driven 100% of the time in the control window. This would give rise to 10-fold bias in the driving exposure estimate in control windows. This driving bias then causes biased estimates of calling time in the control windows, because call rates are higher when driving vs. not (Young, 2013a) (see also Section 5 in Part 2 below). The Toronto and Australian studies did not recognize this part-time driving bias, nor did their methods adjust for it, so all their RR estimates are biased.

As evidence of the generality of part-time driving to other locations and driver samples, Young (2012a) analyzed 100 days of GPS data collected from 439 vehicles during 2005-2006 in Puget Sound, Washington, using the same hours and days of the week as the crashes in the Toronto study. About 36% of the vehicles were not driven at all in the corresponding clock time the day before, consistent with the 35% recall-based driving intermittency estimate in the Toronto study. However, only 0.3% of the vehicles were driven for the full 10-minute corresponding window on the day before. The remaining 64% of the vehicles (the 2/3 majority) were driven during only part of the 10-

minute control window. The driving intermittency correction method employed in the Toronto study corrected only for the 35% of subjects who did not recall driving at all in the previous control window. It did not correct for those who remembered driving in the control window, but who did not drive for the full duration of the control window.

On average, only about 20% of the time on any previous day was in-car, given known driving on a subsequent day at that same clock time. These findings were robust for weekdays/weekends or varying comparison days (Young 2012b), or different cities and subjects (Young, 2011a). Indeed, most modern trip analysis studies in GPS-equipped vehicles that have been conducted in dozens of U.S. cities find that drivers in general are far less consistent in their day-to-day driving performance than had been previously believed (Young and Seaman, 2012).[21]

2.2.2 Method to Predict RR from a Call Rate Imbalance In-Car vs. Out-of-Car

The crude RR estimate of 4.3 in Table 1B assumes that subjects who remembered driving in a control window were in the car during that entire control window. However, subjects were estimated to be in the car for only 20% of the control window duration. To adjust for this part-time driving bias, it is necessary to have an estimate of the out-of-car and in-car call rates. Given that driving occurs on average during only 1/5 of the time in control windows, there are three possible effects on the RR estimate as a function of the in-car vs. out-of-car call rates as estimated from billing records:

- If the call rate were lower in-car than out-of-car, then too many calls in control windows would be (incorrectly) observed in the billing records, biasing the RR denominator high, and the RR estimate low.
- If the call rate were equal for in-car vs. out-of-car, then there is no call effect whether in or out of the car, and the RR estimate is valid as is, and the RR would not be confounded by either part-time driving or driving intermittency.
- If the call rate were higher in-car than out-of-car, then too few calls in control windows would be (incorrectly) observed in the billing records, biasing the RR denominator low, and the RR estimate high.

It would thus be useful to have an equation that expresses the predicted RR as a function of the imbalance in call rates for out-of-car vs. in-car during control windows, given that 20% of the control window time was in-car.

Based on work done by Young (2013a), Equation 1, given below, predicts the RR as a function of ρ, the ratio of in-car to out-of-car call rates (given that on average only 1/5 of the time is in the car on a previous day, compared to a subsequent day). The predicted RR varies directly with ρ, as follows:

$$\mathrm{RR} = 3.675/\rho + 0.919 \qquad (1)$$

The predicted RR in Equation 1 depends only on ρ, not on call duration in the billing records, nor sampling window duration. Therefore, it predicts RR for any call or window duration.

Figure 1 plots predicted RR estimates from Equation 1 as a function of ρ.

---

[21] "The underlying assumption that travel is repetitive from day to day is …highly suspect." (Stopher and Zhang 2011).

Point A shows that if $\rho = 1$ (no difference for in-car vs. out-of-car call rate), then Equation 1 predicts an RR of 4.59, identical to the crude Toronto RR estimate of 4.59 (Table 6), as expected since that information was partly used to generate Equation 1.

Point B shows that $\rho = 7$ (objective evidence for which is given in Section 5 of Part 2) predicts an adjusted RR of 1.44.

Point C shows that $\rho = 10$ predicts an RR of 1.29, equal to the 100-Car NDS hand-held conversation RR estimate (Klauer et al. 2006). Young (2012d, 2013a) estimated $\rho = 10$ by combining objective but indirect cellular billing and driving data.

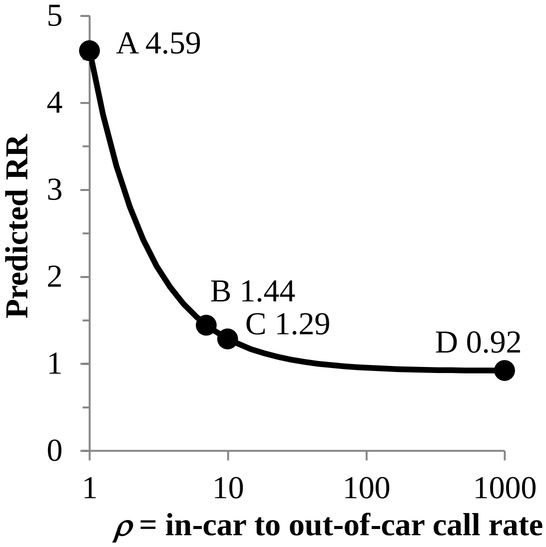

Figure 1 **Graph of Equation 1 for predicting the RR as a function of in-car to out-of-car call rate ρ (based on the Toronto data)**

Point D shows that Equation 1 predicts an RR of 0.92 for an embedded wireless device like OnStar that permits only in-car calls. Young and Schreiner (2009a) found that the rate ratio for OnStar personal calls is 0.62 (95% CI 0.37 to 1.05).[22] The predicted RR of 0.92 for an embedded device falls within that confidence range.

The success of Equation 1 in predicting multiple study results increases confidence that the adjustments for part-time driving bias in the Toronto and Australian RR estimates are valid.

2.2.3 Method to Estimate In-Car vs. Out-of-Car Calls from NDS Data

The ratio of in-car to out-of-car calls (ρ) is critical for valid adjustment of the RR estimate as a function of part-time driving in control windows. An improved estimate of ρ was made from the Fitch et al. (2013) NDS data, based on objective video and billing data for in-car calls, and billing data for out-of-car calls (no video data were available for out-of-car calls in the Fitch et al. study).

Table 6 shows the billing records indicate 51,725 calls (on hand-held, portable hands-free, or integrated hands-free devices) during the 31 study days, of which 14,754

[22] The rate ratio in Table 3 of the current paper for the OnStar device shows a slightly smaller lower confidence limit (0.34) than Young and Schreiner (2009: Table III) (0.37) because the "exact" method in StataIC/12 was used to calculate the confidence limits in Table 3.

(28.5%) occurred while driving. Drivers talked on a cell phone 10.6% of the time when the ignition was on. There were a total of 8,240 hours of driving by 204 drivers. (These driving hours were reduced to 7,553 to correct for the fact that only 187 of the 204 drivers who provided billing records placed or received calls while driving.) There were 139,128 total hours in the study for these 187 drivers. From these numbers ρ is estimated as 7.0 (Table 6).[23]

Table 6 Estimated in-car to out-of-car call rate based on objective direct measurements of call and driving data by Fitch et al. (2013)

	in-car	out-of-car[b]	Total	% in-car
Calls	14,754	36,971	51,725	28.5%[c]
Person-hours	7,553	131,575	139,128	5.40%
Rate	1.95[a]	0.28		
Ratio ρ	7.0			

[a]Consistent with Fitch et al. (2013: 20) estimate of 2 in-car calls per driving hour.

[b]Integrated devices cannot make out-of-car calls and are not included here.

[c]Consistent with Fitch et. al. (2013: xxiii) estimate of 28% in-car calls.

2.2.4 Method to Adjust an RR Estimate for Call Misclassification Bias

A model of call misclassification bias was developed (Young, submitted). This model predicted the effect of crash time error on misclassification of calls before or after the crash. It used objective crash time data from the OnStar system, and compared it to subjective crash time data from police reports for the same crash, using data from Blatt et al. (2009: their Figure 3). This model was then used to make an additional correction to the Toronto estimated RR to adjust for call misclassification bias. The model indicates that the RR estimates in studies that have non-objective measures of crash times and therefore have errors in the crash time estimates that must be further reduced by an adjustment factor of 2.37, consistent with the adjustment factor of about 3 estimated by the Bellavance (2004, 2005) model of the biasing effect of call misclassification in the Toronto study.

2.3 Toronto Study Adjustment Results

To illustrate the method of correcting for part-time driving bias, it is perhaps easiest to use the simple method of dividing the 170 calls in the case window by the 37 calls in the control window, for a crude[24] RR estimate of 4.59.[25]

This simple method shows that there were about 4 times as many calls made just before the estimated time of the crash than in the same clock-time windows on previous days with no crash. This crude estimate is close to the final adjusted RR of 4.3 using the

[23] Because ρ = 7 estimate is derived from direct cell phone records and exact driving times, it is likely more accurate than the ρ = 10 estimate by Young (2012c,d), which was based on objective data, but compiled from indirect sources.

[24] *Crude* is an epidemiological term, referring to an estimate of relative risk before adjustment for bias.

[25] A more elaborate method to control for part-time driving bias was developed and applied to the data arranged in the case-crossover 2x2 table format as per Tables 1 and 2, and yielded similar results, but the explanation is more cumbersome.

case-crossover analysis method (Table 1B).

Table 7A applies a rate ratio method to estimate this same crude Toronto RR of 4.59. That is, the simplified RR method can easily be converted to a rate ratio method (Young 2012c, 2013a) by dividing the observed number of callers in the case and control windows by the number of person-minutes in those case and control windows.

This rate ratio method also approximates the adjusted risk ratio of 4.3 using the case-crossover method (Table 1B), so any effect that biases the rate ratio estimate would plausibly be expected to have an equivalent biasing effect on the risk ratio estimate. For example, if a rate analysis shows that the 170:37 imbalance in the number of calls in the case vs. control windows is entirely due to bias, and the true call ratio between cases and controls is really 1:1, then all metrics and epidemiological analysis methods will agree that the RR is no different than 1.

To arrive at a more bias-free estimate of the RR than provided in Table 7A, both part-time driving bias and call misclassification bias must be eliminated. Using a rate ratio with the denominator of time has the benefit of making it easier to see how to remove part-time driving bias from the RR estimate. This is because the rate ratio involves a function of time (person-minutes of exposure), unlike the risk ratio, which uses counts (e.g. Tables 1 and 2).

Table 7B shows that eliminating the 20% part-time driving bias reduces the RR estimate from 4.59 to 1.44. This adjustment uses the ρ = 7 estimate from Table 6 and the 20% part-time driving estimate from Section 2.2.1.

Table 7C shows that adjusting for call misclassification bias alone using the method in Section 2.2.4 reduces the RR estimate from 4.59 to 1.95.

Table 7D shows that combining both adjustments yields an adjusted RR estimate of 0.61 for the Toronto study, near the OnStar crude RR estimate of 0.62 (Table 3).

In sum, there is little difference between hand-held or embedded hands-free devices in terms of the effects of holding conversations on them – both have a crash RR near 1, or baseline driving, after adjusting for bias.

Table 7 **Toronto study adjusted RR estimates for the data in Table 6. A. Converting risk data into rate ratios. B. Adjustment for 20% part-time driving in control windows, and an in-car call rate 7 times the out-of-car. C. Adjustment for call misclassification bias. D. Both adjustments combined**

	Case Window[a]	Control Window[b]			Adjustments	
A. Original	(in-car)	(in-car)	(out-of-car)	Total		
Calls	170	37	0	37	None	c
Person-minutes	1700	1700	0	1700	None	d
Rate	0.1000	0.0218	undefined		undefined	e
RR (95% CI)	**4.59**	(3.20 to 6.75)				
B. Adjust for 20% part-time driving and ρ = 7 in control window						
Calls	170	23.5	13.5	37	None	c
Person-minutes	1700	340	1360	1700	20%	d
Rate	0.1000	0.0691	0.0099		7	e
RR (95% CI)	1.44	(0.94 to 2.22)				
C. Adjust for call misclassification						
Calls	72	37	0	37	2.37	c
Person-minutes	1700	1700	0	1700	None	d
Rate	0.0424	0.0218	undefined		undefined	e
RR (95% CI)	**1.95**	(1.29 to 2.98)				
D. Both adjustments						
Calls	72	23.5	13.5	37	2.37	c
Person-minutes	1700	340	1360	1700	20%	d
Rate	0.0424	0.0691	0.0099		7	e
RR (95% CI)	**0.61**	(0.38 to 0.98)				

Key: [a]10 min duration, day of crash
[b]10 min duration; 1,3,7 days before crash

Adjustments: [c]misclassification correction
[d]part-time driving
[e]ρ =in-car/out-of-car call rate

2.4 Australian Study Adjustment Results

For simplicity, the Australian study data were placed into a rate ratio framework, and adjusted in the same manner as done for the Toronto study. There were 40 calls in the case window and 25 in the control window (Table 2D). Table 8 estimates an adjusted RR of 0.64 for that data, after controlling for part-time driving and call misclassification bias. The final adjusted crash RR was 0.64, near the adjusted crash RR of 0.61 in the Toronto study and the crude crash RR of 0.62 in the OnStar study.

Table 8 **Australian study adjusted RR estimates. A. Conversion of risk ratio data into rate ratios. B. Adjustment for 20% part-time driving in control windows, and an in-car call rate 7 times the out-of-car. C. Adjustment for call misclassification bias. D. Both adjustments combined**

	Case Window[a]	Control Window[b]			Adjustments	
A. Original	(in-car)	(in-car)	(out-of-car)	Total		
Calls	40	25	0	25	none	c
Person-minutes	400	1200	0	1200	none	d
Rate	0.1000	0.0208	undefined		undefined	e
RR (95% CI)	**4.8**	(2.91 to 7.91)				
B. Adjust for 20% part-time driving and ρ = 7 in control window						
Calls	40	15.9	9.1	25	none	c
Person-minutes	400	240	960	1200	20%	d
Rate	0.1000	0.0663	0.0095		7	e
RR (95% CI)	1.51	(0.84 to 2.70)				
C. Adjust for call misclassification						
Calls	17	25	0	25	2.37	c
Person-minutes	400	1200	0	1200	none	d
Rate	0.0425	0.0208	undefined		undefined	e
RR (95% CI)	**2.04**	(1.10 to 3.78)				
D. Both adjustments						
Calls	17	15.9	9.1	25	2.37	c
Person-minutes	400	240	960	1200	63%	d
Rate	0.0425	0.0663	0.0095		7	e
RR (95% CI)	0.64	(0.32 to 1.27)				

Key: [a]up to 10 min duration, day of crash
[b]up to 10 min duration; 1,3,7 days before crash

Adjustments: [c]misclassification correction
[d]part-time driving
[e]ρ=in-car/out-of-car call rate

2.5 OnStar Study Adjustment Results

All calls must be made inside the vehicle because the OnStar system is embedded in the vehicle; thus, no part-time driving adjustment is needed for the OnStar rate ratio estimate. No demographic data were available[13] to adjust the RR for demographic factors in the OnStar cohort of 3 million drivers. Again, this demographic bias is not unique to the OnStar study, because it is present in any NDS study that does not stratify its results by demographics, or instead uses the case-crossover study design to completely eliminate the possibility of bias from all demographic variables.

2.6 100-Car Study Adjustment Results

Table 9A shows the results of a re-analysis by Young (2013b) of the 100-Car crash/near-crash and unmatched baseline databases (VTTI 2010), using the standard epidemiological method of comparing exposure (Call) to non-exposure (No Call) (Young 2013b), and including all cases regardless of driver fault to remove errors and bias in the assignment of fault (Young 2013b). This method estimated an RR of 0.78 for hand-held cell phone conversation and crash/near-crash risk in the 100-Car data (Young 2013b).

Table 9 **New 100-Car RR estimates for hand-hand cell phone conversation. A. Similar to Table 4 but using all cases regardless of fault. B. Same cases as A, but using matched baseline data for Call task (digitized from Klauer et al. 2010: Figure 16)**

A. Standard design, all cases regardless of fault[a]				
	Call	No Call	Total	Prevalence
Crash/Near-Crash	44	769	813	5.4%
Unmatched Baseline	1,339	18,276	19,615	6.8%
RR (95% CI)	0.78	(0.56 to 1.06)		
B. Same cases as A but matched baseline[b]				
	Call	No Call	Total	Prevalence
Crash/Near-Crash	44	769	813	5.4%
Matched Baseline	930	9,078	10,008	9.3%
RR (95% CI)	**0.56**	(0.41 to 0.76)		

[a]Unmatched baseline data from Young (2013b)

[b]Matched baseline data from Klauer et al. (2010: Fig. 16)

Table 9B shows the same case data, but using a matched baseline that removes bias due to demographics, junction, weather, time of day, etc. which reduced the RR estimate to 0.56 for the same case data. Again, the matched baseline (unlike the unmatched baseline) removes bias for these variables because each subjects acts as their own control, and the proximity of the baseline clips to the case clips ensures that closeness to junction, weather, time of day, etc. are controlled for.

2.7 Cell Phone NDS Adjustment Results

Table 10 uses the standard epidemiological analysis design (Young 2013b) of comparing the exposure (e.g. "Call") to its complement ("No Call") applied to the data of Fitch et al. (2013). In this particular case, the standard analysis method estimates a consistent RR (0.71) to the non-standard method in Table 5 (RR = 0.75). However, the standard method correctly estimates the prevalence and "Population Prevented Fraction % "(*PPF*%),[26] and the non-standard method does not.

[26] *PPF%* is an estimate of the number of safety-critical events in the population that are prevented by the exposure (Young 2013b). The term is also called "Prevented Fraction (Population)" and is formally defined in epidemiology as, "In a situation in which exposure to a given factor is believed to protect against a disease (or other outcome), the prevented fraction is the proportion of the hypothetical total load of the disease (in the population) that has been prevented by exposure to the factor" (Porta 2008: 192). With regard to Call, *PPF%* is an estimate of the amount by which the safety-critical event under investigation (crash, near-crash, etc.) would increase in the population as a whole (not just in the sample of drivers investigated) if the exposure to Call while driving were eliminated.

Table 10 Adjusted RR estimates for Fitch et al. (2013) NDS data for cell phone conversation using standard epidemiological method (Young 2013b) comparing Call vs. No Call with matched baseline

Standard Design	Call	No Call	Total	Odds	Prevalence
Safety-Critical Event[a]	28	314	342	0.0892	8.2%
Matched Baseline	259	2,049	2,308	0.1264	11.2%
RR (95% CI)	0.71	(0.47 to 1.06)			
PPF% (95% CI)	*3.3% (-0.3% to 6.9%)*				

[a]Crashes (including curb strikes), near-crashes, and crash-relevant conflicts

2.8 All 5 Studies: Pooled Relative Risk

Table 11A shows the original data from Part 1 for cell phone conversation RR estimate while driving. Studies 1 and 2 estimate RR near 4, while studies 3-5 estimate RR near 1. Attempting to pool across these data using meta-analysis[27] shows that they are heterogeneous. Therefore, the pooled result is invalid and is not shown.

Table 11 Adjusted relative risk for cell phone conversation while driving for 5 epidemiological studies and pooled meta-analysis estimate. A. Original estimates as published. B. Estimates as adjusted in current study

			A. Original			B. Adjusted		
				95% CI			95% CI	
#	Study	Study	RR	Lower	Upper	RR	Lower	Upper
1	Redelmeier and Tibshirani (1997)	Toronto	**4.3**	3.0	6.5	**0.61**	0.38	0.98
2	McEvoy et al. (2005)	Australia	**4.1**	2.2	7.7	0.64	0.32	1.27
3	Young and Schreiner (2009)	OnStar	0.62	0.37	1.05	0.62	0.37	1.05
4	Klauer et al. (2006, 2010)	100-Car	1.29	0.93	1.80	**0.56**	0.41	0.76
5	Fitch et al. (2013)	Cell Phone	0.75	0.49	1.15	0.71	0.47	1.06
	Pooled		N/A†			**0.61***	0.51	0.74
	p -value for homogeneity		2E-11			0.93		

Table 11B shows that after adjustment for bias, the RR upper confidence limit is either slightly above 1 (studies 2, 3, and 5) or slightly below 1 (studies 1 and 4). Table 11B further shows that after adjustment, the data from all 5 studies are now homogeneous, indicating that they can be validly pooled. The pooled RR estimate is 0.61 (95% CI 0.51 to 0.74), indicating that engaging in cell phone conversation while driving, whether with hand-held, portable hands-free, or embedded hands-free devices, reduces driving safety risk compared to not engaging in a cell phone conversation while driving.

3 Summary and General Discussion

The current results show that after adjustment for bias, all five major epidemiological studies of calls and crashes – despite wide differences in study designs, locations, times,

[27] Calculated with the meta-analysis worksheet in Episheet (Rothman 2012b).

assumptions and methods – yield adjusted RRs near or below 1 for cell phone conversation while driving, regardless of whether the cellular device is hand-held, portable hands-free, or embedded hands-free. This result holds for all levels of crash severity: crashes severe enough to deploy an airbag (Young and Schreiner 2009a); injury crashes necessitating hospital attendance (McEvoy et al. 2005); property-damage crashes (Redelmeier and Tibshirani 1997a); combined minor crashes and near-crashes (Klauer et al. 2006); and combined minor crashes, near-crashes, and crash-relevant conflicts (Fitch et al. 2013).

Studies with valid times of crashes, calls, and driving have largely superseded studies which relied on cell phone billing records and rough estimates of crash times from police reports. It is not recommended that future studies be conducted on this topic that make use only of recall by subjects of their driving and crash times, even if augmented by police accident reports or on-site crash investigations (none of which have precise information on crash times). Future recall-based studies, if conducted, should report the number of people conversing and not conversing on cell phones and tabulate each person's number of calls: (1) before crashes; (2) after crashes; (3) in control periods when definitely driving; and (4) in control periods when not driving. These future studies should then adjust for misclassification bias arising from errors in crash time estimates as required.

Limitations

In the Young (2012a) study of part-time driving in Seattle GPS data was collected at a different time and place from the Toronto and Australian studies, a limitation noted by Young. Therefore, the estimate of part-time driving in control periods might be different for different times, places, or subjects. However, in control studies, Young showed that the Seattle results about the percentage of part-time driving in control periods were robust to day and day of week (Young 2012b), showing only small variations (basically, driving in general is not very consistent from day to day, even on workdays). Indeed, the part-time driving percentage was about the same even when restricted to the exact days, days of the week, and clock times as the Toronto and Australian studies (Young 2012d). Young (2011a) also found the same part-time driving results in a 2007-2008 Chicago GPS study with different subjects and location.

Reasons for Non-Elevation of Crash RR by Cellular Conversation

Numerous experimental studies – whether brain imaging, laboratory, simulator, test track, or on-road – find that cell phone conversation or other primarily "cognitive" tasks increase response times to visual events by about 70-200 ms (e.g. Hsieh et al. 2007; Young et al. 2013),[28] or have various other minor effects on driver performance (e.g., Strayer et al., 2013). However, the current results show that the epidemiological data from 5 real-world driving studies concur (after adjustment for bias) that there is no increase in crash risk from person-to-person cell phone conversations while driving compared to no conversation, whether using hand-held, portable hands-free, or embedded hands-free devices, a result that apparently contradicts the experimental

[28] Indeed, the specific brain regions and dynamic neural mechanisms that give rise to these increases in event response time from conversation during simulated driving have been identified in brain studies using magnetoencephalography (Bowyer et al. 2009) and functional magnetic resonance imaging (Hsieh et al. 2009). They are part of the orienting and executive attention networks in the brain that are the fundamental processes by which event detection occurs (Young 2012g).

studies. Since real-world data are the "gold standard," the question is raised about what biases are present in the experimental studies that lead to such invalid predictions about real-world driving? Assuming the experimental study results themselves are internally valid, why are attempts to generalize to real-world driving apparently invalid?

The compensation hypothesis of Young and Schreiner (2009a: 199) is one explanation. It states that drivers talking on a cell phone during real-world driving tend to change their driving behavior (using cell phones in less-demanding situations, decreasing speed, increasing headway, etc.) such that there is no net safety decrement from the slight increase in event response time from cell phone conversation.[29] For example, drivers may choose to initiate calls at times when road and traffic conditions provide them a margin of safety that can "absorb" the small performance decrements that result from talking, or they may slow down slightly to create an extra safety margin. Such compensatory effects are not seen in experimental studies because the subjects are required to perform the tasks in order to measure the physical or mental demand of those tasks – such studies do not and cannot measure a subject's willingness to perform the task in real-world driving, nor how they would actually drive in such situations. Experimental studies attempting to measure the physical or mental load of secondary tasks – whether in the brain imager, laboratory, or on the track or road –therefore do not and cannot account for self-regulatory behavior by drivers under real-world driving conditions. Such self-regulatory behavior can only be revealed in naturalistic driving studies or by inferences from real-world crash statistics.

Another explanation is that drivers concentrate their gaze on the road more during cell phone conversations or other auditory-vocal tasks that increase cognitive load, and more eyes-on-road time is well-established to improve driving safety (Klauer et al. 2006, Young 2012e).

Another explanation is that cell phone conversation reduces drowsiness[30] while driving, which in turn reduces crash risk, as shown by Young (2013c) in NDS data. Experimental studies where subjects are not drowsy would not see these improvements, and so would (incorrectly) conclude that cell phone conversation in real-world driving increases crash risk.

A related explanation is that drivers when drowsy may adopt a strategy to do simple secondary tasks, or engage in cell phone conversations, to reduce their drowsiness (see Young 2013c). In this manner, cell phone use, along with perhaps many other secondary tasks, could be symptomatic of drowsy driving. If so, cell phone or other secondary task engagement could be symptomatic of this strategy, and not a causal factor of crashes.

---

[29] For example: limiting wireless communication device use to low-risk driving situations, reducing the tendency to engage in risky maneuvers (e.g., lane changes) when in a conversation, increasing headway to a forward vehicle, reducing speed, ending conversation when the driving task demands it.

[30] Drowsiness is caused by a decline in activation in the brain's alerting attention network, an entirely different type of attention decline than a decline in executive and orienting attention caused by cognitive load (Young 2012e,f,g; Young et al. 2013). Declines in alerting attention during driving are a far greater risk factor for crashes than the slight attentional effects on orienting and executive attention from cognitive load (Young 2013c).

Policy Implications

These results indicate that legislation, regulations, or guidelines that prohibit cellular conversation with a portable hand-held, portable hands-free, or embedded hands-free device while driving will not reduce crashes, and may even increase crashes given that, after adjustment for bias, the five major epidemiological studies examined here give rise to a pooled RR estimate less than 1, indicating a protective effect for cell phone conversation while driving in passenger vehicles.

This evidence suggests that efforts to curb secondary tasks in vehicles should be carefully considered to assure that potentially protective effects are not prohibited.

4 Conclusion

Case-crossover studies of the risk of cell phone conversation while driving overestimated the relative risk by about 7 times. After adjusting for bias from part-time driving in control windows and misclassification of calls as pre- or post-crash, all five major epidemiological studies of conversation on cellular devices (whether portable hand-held, portable hands-free or embedded hands-free) find the relative risk to be near or somewhat below 1. The pooled adjusted relative risk estimate for cell phone conversation while driving in passenger vehicles is 0.61 (CI 0.51 to 0.74).

Acknowledgements

This paper was prepared on the author's personal time and reflects no conflicts of interest. I thank several anonymous reviewers for comments on earlier drafts.

References

Baker, S. P. 1971. Digit preference in reported time of collision. *Accident Analysis & Prevention,* 3(1): 77-80. Available at: http://www.sciencedirect.com/science/article/B6V5S-4695BJG-5/2/60fbf1be6f6f6900f3678d0dd5ab1c86 [Accessed: 1 August 2013].

Bellavance, F. 2004. Misclassification bias in the case-crossover design applied to wireless telephones and the risk of road crashes [abstract]. Available at: https://www.mcgill.ca/files/epi-biostat-occh/Bellavance-04.pdf [Accessed December 18, 2010].

Bellavance, F. 2005. *Linking data from different sources to estimate the risk of collision when using a cell phone while driving*. Presented at the International Conference on Distracted Driving. Toronto, Canada. Available at: http://www.distracteddriving.ca/english/documents/FrancoisBellavance_001.pdf [Accessed December 18, 2010].

Blatt, A., Funke, D, McGwin, G, Majka, K, Flannigan, M. and Russell, M. 2009. *Accuracy of reported crash times: A comparison of rural and urban crashes using crash vehicle telemetry data.* Paper presented at the 2009 National Rural ITS Conference, Seaside, OR. Available at: http://www.nritsconference.org/downloads/Presentations09/S1_Flannigan.pdf [Accessed 12 May 2013].

Bowyer, S., Hsieh, L., Moran, J., Young, R., Manoharan, A., Liao, C.-C., Malladi, K., Yu, Y.-J., Chiang, Y.-R. and Tepley, N. 2009. Conversation effects on neural mechanisms underlying reaction time to visual events while viewing a driving scene using MEG. *Brain Research,* 1251: 151–161. Available at: http://sun3.science.wayne.edu/~mttc/conversation_effect_fMRI_Hsieh_2009.pdf [Accessed: 30 April 2013].

Braver, E., Lund, A. and McCartt, A. 2009. Hands-free embedded cell phones and airbag-deployment crash rates (letter). *Risk Analysis*, 29(8): 1069. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19558394 [Accessed: 30 April 2013].

Fitch, G., Soccolich, S., Guo, F., McClafferty, J., Fang, Y., Olson, R., Perez, M., Hanowski, R., Hankey, J. and Dingus, T. 2013. *The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk*. (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration. Available at: http://trid.trb.org/view.aspx?id=1249880 [Accessed: 30 April 2013].

Guo, F., Klauer, S., Hankey, J. and Dingus, T. 2010. Near crashes as crash surrogate for naturalistic driving studies. *Transportation Research Record: Journal of the Transportation Research Board,* 2147: 66-74. Available at: http://trb.metapress.com/content/p67l2436826p5u77/ [Accessed: 30 April 2013].

Hankey, J. 2007. Comparison of odds ratios and population attributable risk for cell phone dialing and talking/listening on a handheld cell phone. (NHTSA-2007-28442). Appendix 3: New VTTl analysis based on 100-Car data, NHTSA Docket No. NHTSA-2007-28442. Available at: http://www.regulations.gov/search/Regs/contentStreamer?objectId=09000064802b42a2&disposition=attachment&contentType=pdf [Accessed: 17 May 2013].

Hsieh, L., Young, R., Bowyer, S. and Moran, J. 2007. *Conversation effects on driving: Neural mechanisms underlying reaction times to visual events*. Paper presented at the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Stevenson, WA. Available at: http://drivingassessment.uiowa.edu/DA2007/PDF/082_HsiehYoung.pdf [Accessed: 15 May 2013].

Hsieh, L., Young, R., Bowyer, S., Moran, J., Genik II, R., Green, C., Chiang, Y., Yu, Y., Liao, C. and Seaman, S. 2009. Conversation effects on neural mechanisms underlying reaction time to visual events while viewing a driving scene: fMRI analysis and asynchrony model. *Brain Research,* 1251: 162–175. Available at: http://sun3.science.wayne.edu/~mttc/conversation_effect_fMRI_Hsieh_2009.pdf [Accessed: 30 April 2013].

Klauer, S., Dingus, T., Neale, V., Sudweeks, J. and Ramsey, D. 2006. *The impact of driver inattention on near-crash/crash risk: an analysis using the 100-Car naturalistic driving study data*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration (NHTSA). Available at: http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash Avoidance/Driver Distraction/810594.pdf [Accessed: 7 May 2013].

Klauer, S., Guo, F., Sudweeks, J., and Dingus, T. 2010. *An analysis of driver inattention using a case-crossover approach on 100-car data: Final report*. (DOT HS 811 334). Washington, D.C.: U.S. Department of Transportation. Available at:

http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811334.pdf [Accessed: 7 May 2013].

Maclure, M. and Mittleman, M. 1997. Cautions about car telephones and collisions [editorial]. *The New England Journal of Medicine*, *336*(7), 501–502. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9017945 [Accessed: 30 April 2013].

McEvoy, S., Stevenson, M., McCartt, A., Woodward, M., Haworth, C., Palamara, P. and Cercarelli, R. 2005. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *BMJ*, 331: 1–15 (online version). Available at: http://www.bmj.com/content/331/7514/428 [Accessed: 20 August 2013].

Porta, M. 2008. *A Dictionary of Epidemiology* (Fifth Ed.). New York: Oxford University Press.

Quinlan, K. 1997. Cellular telephones and traffic accidents [letter]. *The New England Journal of Medicine*, 337(2): 127. Available at: http://www.nejm.org/doi/full/10.1056/NEJM199707103370211 [Accessed: 30 April 2013].

Redelmeier, D. and Tibshirani, R. 1997. Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, 336(7): 453–458. Available at: http://www.nejm.org/doi/full/10.1056/nejm199702133360701 [Accessed: 30 April 2013].

Rothman, K. 2012a. *Epidemiology: An Introduction* (Second Ed.). New York, NY: Oxford University Press.

Rothman, K. 2012b. Episheet: *Spreadsheets for the analysis of epidemiologic data*. Available at: http://krothman.org/Episheet.xls [Accessed: 30 April 2013].

Stopher, P. and Zhang, Y. 2011. *Repetitiveness of daily travel*. Paper submitted to the Transportation Research Board Annual Meeting, Washington, DC, January.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., and Biondi, F. 2013. *Measuring cognitive distraction in the automobile*. AAA Foundation for Traffic Safety. Available at: http://www.multivu.com/assets/60391/documents/60391-AAA-Measuring-Cognitive-Distractions-Full-Report-AAA-Foundation-original.pdf. [Accessed: 18 July 2013].

Tibshirani, R. and Redelmeier, D. A. 1997. Cellular telephones and motor-vehicle collisions: Some variations on matched-pairs analysis. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 25(4), 581-591. Available at: http://dx.doi.org/10.2307/3315349.

VTTI. 2010. 100-Car databases. Available at: http://forums.vtti.vt.edu/index.php?/files/category/3-100-Car-data/ [Accessed: 5 May 2013].

Young, R. 2001. *Recent human factors issues in the use of embedded telematics devices in a vehicle*. First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Available at:

http://drivingassessment.uiowa.edu/DA2001/81_young-richard.pdf [Accessed 22 May 2013].

Young, R. and Schreiner, C. 2009a. Real-world personal conversations using a hands-free embedded wireless device while driving: effect on airbag-deployment crash rates. *Risk Analysis*, 29(2): 187–204. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.2008.01146.x/abstract [Accessed 20 August 2013].

Young, R. and Schreiner, C. 2009b. Response to Braver et al. *Risk Analysis*, 29(8): 1070–1071. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1539-6924.2009.01249.x/abstract [Accessed: 30 April 2013].

Young, R. 2011a. *Driving consistency errors overestimate crash risk from cellular conversation in two case-crossover studies*. Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, 298–305. Available at: http://drivingassessment.uiowa.edu/sites/default/files/DA2011/Papers/043_Young.pdf [Accessed: 7 May 2013].

Young, R. 2011b (unpublished). *Epidemiological analysis of call-crash association in the SHRP 2 naturalistic driving study data*. SHRP 2 Project S08 Proposal. Strategic Highway Research Program, Transportation Research Board.

Young, R. 2012a. Cell phone use and crash risk: evidence for positive bias. *Epidemiology,* 23(1): 116–118. Available at: http://journals.lww.com/epidem/Fulltext/2012/01000/Cell_Phone_Use_and_Crash_Risk___Evidence_for.17.aspx [Accessed: 7 May 2013].

Young, R. 2012b. Supplemental digital content: control analyses. Available at: http://links.lww.com/EDE/A535 [Accessed: 7 May 2013].

Young, R. 2012c. Cell phone use and crash risk [response]. *Epidemiology, 23*(4), 649–650. Available at: http://journals.lww.com/epidem/Fulltext/2012/07000/Cell_Phone_Use_and_Crash_Risk.25.aspx [Accessed: 7 May 2013].

Young, R. 2012d. More on cell phone use and crash risk: the author replies [response]. *Epidemiology,* 23(5): 774–775. Available at: http://journals.lww.com/epidem/Fulltext/2012/09000/More_on_Cell_Phone_Use_and_Crash_Risk.27.aspx [Accessed: 7 May 2013].

Young, R. 2012e. *Seven myths about cognitive distraction and driving*. Paper presented at the Fifth International Conference on Traffic and Transport Psychology, Groningen, The Netherlands. Available at: http://www.icttp2012.com/images/stories/presentations/30th/1330_Driver_attention_and_distraction_Eyes_on_the_road/Young_ICTTP2012_Seven_Myths_about_Cognitive_Distraction_and_Driving.pdf [Accessed: 7 May 2013].

Young, R. 2012f. Cognitive distraction while driving: a critical review of definitions and prevalence in crashes. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst.*, 5(1). Available at: http://saepcelec.saejournals.org/content/5/1/326.short [Accessed: 30 April 2013].

Young, R. 2012g. Event detection: The second dimension of driver performance for visual-manual tasks. *SAE Int. J. Passeng. Cars - Electron. Electr. Syst., 5*(1). Available at: http://saepcelec.saejournals.org/content/5/1/297.abstract [Accessed 20 August 2013].

Young, R. and Seaman, S. 2012. *Improving survey methods using a new objective metric for measuring driving time variability in survey and GPS data.* Paper from the Transportation Research Board 2012 Annual Meeting, Washington, D.C. Available at: http://trid.trb.org/view.aspx?id=1130665 [Accessed: 1 August 2013].

Young, R. 2013a. Cell phone use and crash risk [response]. *Epidemiology, 24*(3), 468–469. Available at: http://journals.lww.com/epidem/Fulltext/2013/05000/Cell_Phone_Use_and_Crash_Risk.27.aspx [Accessed: 7 May 2013].

Young, R. 2013b. *Naturalistic studies of driver distraction: effects of analysis methods on odds ratios and population attributable risk.* Paper from the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, June 17-20. Available at: http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/077_Young_0.pdf [Accessed: 1 August 2013].

Young, R. 2013c. *Drowsiness and crash risk: underestimates of odds ratios from confounding in the 100-Car naturalistic driving studies.* Paper presented at the 3rd International Conference on Driver Distraction and Inattention, September 4-6, Gothenburg, Sweden.

Young, R. A., Hsieh, L. and Seaman, S. 2013. *The Tactile Detection Response Task: Preliminary validation for measuring the attentional effects of cognitive load.* Paper from the 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Bolton Landing, NY, June 17-20. Available at: http://drivingassessment.uiowa.edu/sites/default/files/DA2013/Papers/012_Young_0.pdf [Accessed: 1 August 2013].

Young, R. A. submitted. *A model to predict the effect of call misclassification on the relative risk of cell phone conversation while driving*, Paper submitted to the Society of Automotive Engineers, 2014 Annual Meeting, Detroit, MI.