# Female vs. Male Relative Fatality Risk in Fatal Crashes

Mitchell Z. Abrams, Cameron R. Bass

***Abstract*** Since 1975, the US has seen a decrease in vehicle crash fatalities, noticeably among recent model vehicles. However, a lack of requirement to test motor vehicles using anthropomorphic test devices representative of both the average male and the average female occupant may be driving a disparity in the risk of fatality in severe crash scenarios. The double pair comparison method, first developed by Evans, was applied to many different cross-sections of the US Fatality Analysis Reporting System to investigate sex-dependent differences in relative fatality risk in crashes. Despite a dramatic decline in fatalities over the reporting period, female vehicle occupants aged 20–30 years are 20–25% more likely to die as a result of a fatal crash than males in the same age range. The risk to females and males becomes similar as age increases to 60. These trends hold when looking at subsets of crashes in either rural or urban areas, when looking at vehicles manufactured since 2010, and when isolating by single, two- or multiple-vehicle crashes. The consistency of this age-dependent relative risk emphasises a need to further investigate sex differences in crash-related outcomes.

***Keywords*** Crash analysis, Relative fatality risk, Sex differences, FARS, Vehicle safety trends

## I. INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) maintains the Fatality Analysis Reporting System (FARS)[1,2], which tracks all traffic crashes in the USA since 1975 that involve at least one fatality. FARS data are used to inform safety decisions at the local, state and national levels, and provide key insights into the efficacy of changing vehicle and roadway safety standards [2]. To be included in FARS, a crash must occur on a public road and must result in at least one death within 30 days of the crash. Road fatalities in the US continue to decrease as better advanced safety technologies emerge and become standard features across the board. Occupant fatalities involving vehicles manufactured in the last five or 10 years have decreased steadily, down significantly since 1975 (see Fig. 1).

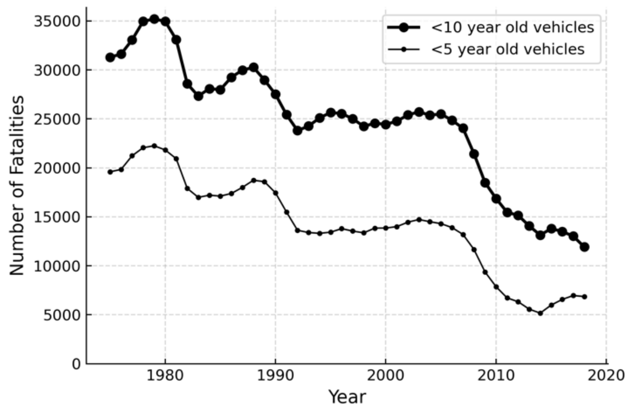


Fig. 1. Annual vehicle occupant fatalities involving vehicles manufactured in the prior five or 10 years.

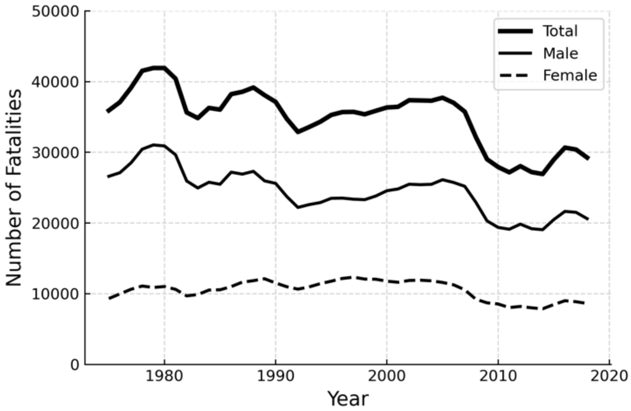


Fig. 2. Annual vehicle occupant fatalities have decreased overall since 1975, yet fatalities among females have remained largely the same year-to-year.

This trend is partially attributable to evolving crash testing and vehicle standards: overall, changes to these

M. Z. Abrams is a PhD Candidate in Biomedical Engineering (e-mail: mitchell.abrams@duke.edu; tel: 919-660-8274). C. R. Bass is an Associate Research Professor of Biomedical Engineering. Both authors are affiliated with Duke University in Durham, NC, USA.

standards are estimated to have prevented ~60% of all potential vehicle fatalities in the USA [3,4]. Of these, a majority of fatalities are male vehicle occupants (Fig. 2). Historically, males drove more miles per licensed driver than females, but that gap is now steadily closing, with decreasing differences in licensure rates and driving exposure [5-7]. Driving exposure for females has increased, and female drivers display similar risk profiles to males when behind the wheel [7-9].

Given these trends, it is important to note that current vehicle crashworthiness tests do not use a crash dummy representative of the average adult female [10]. The current US regulations require crash testing with 50th percentile adult male and 5th percentile adult female anthropomorphic test dummies (ATDs), despite the increase in the number of female drivers over the last 20 years. Previous research has shown that female drivers and vehicle occupants are more likely than males to suffer severe or fatal injuries when involved in a fatal crash [9,11-13].

The goal of this study is to investigate age and sex dependencies on fatality outcomes by matching crash conditions among different sex and age cohorts.

## II. METHODS

### *Data Processing*

Data were downloaded from the NHTSA FARS FTP directory. Data analyses were performed using Python v.3.7.7, with packages installed and managed using Anaconda v4.8.2 on MacOS 10.15.3. Package versions can be found in the environment.yaml file, provided in the GitHub repository for convenience [14]. Files were downloaded using pooch [15], and processed using tools in SciPy [16], namely Pandas (primary package used for data analysis) [17,18], Dask (for parallelised processing) [16], Matplotlib [19] and Seaborn (for visualisations), pyjanitor (for cleanup and recoding of fields), and missingno [20] (for preliminary visualisation of missingness in the dataset).

Data preprocessing was performed using the fars-cleaner package, produced by the authors for this paper and freely available as open source software [21]. Preprocessing merged changes in the FARS dataset over the last 50 years, adjusting outdated and modified codes using the FARS Analytical User Manual as reference [2].

The bulk of the data analysis was performed using Jupyter notebooks [22,23], which are provided as supplemental materials (see Appendix 3 for analysis template) and on Github [14].

### *Double Pair Comparison Method*

The double pair comparison method developed by Evans [24] isolates specific features of fatality risk in a crash. One group of vehicle occupants is selected as the subject occupants, and another is selected as the control. The assessment of fatality risk is performed across the control group. To illustrate the use of the double pair method, if we wish to determine the relative fatality risk of female vs. male drivers, we examine two sets of crashes, choosing a consistent control occupant (for example, male passengers seated in the front right seat of the vehicle). The following can be determined from the dataset:

**A** = Number of female drivers killed in vehicles with a control occupant. (1)
**B** = Number of control occupants killed in vehicles with a female driver. (2)
**C** = Number of male drivers killed in vehicles with a control occupant. (3)
**D** = Number of control occupants killed in vehicles with a male driver. (4)

From these counts, the relative risk of fatality for a female driver vs. the control occupant, $r_1 = A/B$, and the relative risk of fatality for a male driver vs. the control occupant, $r_2 = C/D$, were used to derive the female vs. male relative risk, $R = r_1/r_2$ . It is important to note that the control occupants (B and D) are eliminated from the calculation of relative risk by this technique, serving as a mechanism by which male and female subject occupants can be compared. The original double pair method determines standard error in the estimates of $R$ ($\Delta R$):

$$\Delta R = R\sqrt{\sigma_\mu^2 + \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}} \tag{5}$$

where $\sigma_\mu$ is an estimate of “intrinsic uncertainty,” set to either 0.05 or 0.1 as in [11,24]. The introduction of this constant term forces all variance estimates to be similar, regardless of the pointwise variance in each estimate. Since these are used in the weighted summaries, results may be biased by granting larger weight to samples with more uncertainty. To mitigate this bias, an alternative method was used for describing the variance of the risk ratios [25]. Rather than using *A, B, C,* and *D*, the counts were stratified further:

**A** = Number of female drivers killed in vehicles with a control occupant (also killed). (6)
**B** = Number of female drivers killed in vehicles with a control occupant (not killed). (7)
**C** = Number of control occupants killed in vehicles with a female driver (not killed). (8)
**E** = Number of male drivers killed in vehicles with a control occupant (also killed). (9)
**F** = Number of male drivers killed in vehicles with a control occupant (not killed). (10)
**G** = Number of control occupants killed in vehicles with a female driver (not killed). (11)

Note that the original variables **A**, **B**, **C**, and **D** (Eq. 1-4) can be derived as the sum of the new counts detailed above (Eq. 6-11). Now, the relative risk ratio is given by:

$$R = \frac{{}^{A+B}/_{A+C}}{{}^{E+F}/_{E+G}} \tag{12}$$

As in [25], variance for the log of the relative risk ratio is given with:

$$\Delta R = \frac{[(A\times(A+B+C)+(B\times C))\times(F+G)]+[(E\times(E+F+G)+(F\times G))\times(B+C)]}{(A+B)\times(A+C)\times(E+F)\times(E+G)} \tag{13}$$

Weighted risk ratios ($\bar{R}$) and weighted estimates of variance ($\Delta\bar{R}$) are given by:

$$\bar{R} = \exp\left(\frac{\sum(\ln R\times 1/\Delta R)}{\sum 1/\Delta R}\right) \tag{14}$$

$$\Delta\bar{R} = \frac{1}{\sum 1/\Delta R} \tag{15}$$

Confidence intervals (95%) are derived with a bootstrap method, sampling with replacement many times and calculating values for $\bar{R}$ and $\Delta\bar{R}$ for each new sample set [25]. This procedure was replicated 5000 times for each weighted average. Each bootstrap run produces a distribution of results, from which 95% confidence interval is taken between the 2.5 and 97.5 percentiles.

For all analyses, cases were selected with at least two occupants and with at least one fatality in the vehicle. Fatality was determined as coded within FARS, and includes those declared dead at the crash as well as within 30 days, due to crash-related causes. Initial analyses were performed to compare with Evans [12], matching analyses by considering only cases with no airbag deployment. These comparison analyses are described in Table I. Analyses were then performed with matched airbag deployment, or cases where both the subject and the control occupant experienced the same airbag deployment at their seating position (deployment or no deployment). Cases with unknown airbag deployment were excluded.

As *n* decreases with the granularity of comparison, age ranges were examined in five-year periods for subject occupants, while control occupants were grouped as in previous analyses: ages 16–24, 25–34, 35–54, and 55+ [12]. Table II shows the subject breakdown for 1975–2018 without airbag deployment (*n*=321,320). Breakdowns for the other analyses are shown in Appendix 1.

Fatal crash cases were grouped by vehicle type (car, truck, motorcycle), passenger seating position (front right seat, rear left seat, etc.), seat-belt use, and number of vehicles involved in the crash. In each analysis, a weighted average of the value for *R* was taken within each driver age subset to find the overall risk for a driver. Further, driver age was grouped in five-year chunks to increase sample size for additional analyses.

Additional analyses were performed to ensure the robustness of the results, examining rural vs. urban cases, cases with and without alcohol or drug involvement, and cases involving only late-model vehicles. Results and details can be found in the Appendix.

TABLE I
ANALYSES PERFORMED

| Years | Vehicle Type | Subject Occupants | Air Bag Deployment |
|---|---|---|---|
| 1999–2018 | Car, Light Truck, Motorcycle | Driver, front-right, rear-left/right passengers | None |
| 2010–2018 | Car, Light Truck, Motorcycle | Driver, front-right, rear-left/right passengers | None |
| 1975–2018 | Car, Light Truck, Motorcycle | Driver, front-right, rear-left/right passengers | None |
| 1975–2018 | Car, Light Truck | Driver, front-right, rear-left/right passengers | Deployed |
| 1999–2018 | Car, Light Truck | Driver, front-right, rear-left/right passengers | Deployed |
| 2010–2018 | Car, Light Truck | Driver, front-right, rear-left/right passengers | Deployed |
| 2015–2018 | Car, Light Truck | Driver, front-right, rear-left/right passengers | Deployed |
| 1975–2018 | Car, Light Truck | Driver, front-right passengers | Matched[1] |
| 1999–2018 | Car, Light Truck | Driver, front-right passengers | Matched |
| 2010–2018 | Car, Light Truck | Driver, front-right passengers | Matched |
| 2015–2018 | Car, Light Truck | Driver, front-right passengers | Matched |

[1] Subject and control either both deploy, or do not deploy, airbags.

TABLE II
DISTRIBUTION OF 321,320 FATALLY INJURED SUBJECT OCCUPANTS, 1975–2018, NO AIRBAG DEPLOYMENT

| Vehicle | Subject Occupant | Restraint Use | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|---|
| Car | Driver | Unbelted | 16,994 | 54,196 | 71,190 |
| Car | Right-front passenger | Unbelted | 38,478 | 43,043 | 81,521 |
| Car | Driver | Belted | 8,060 | 15,424 | 23,484 |
| Car | Right-front passenger | Belted | 19,296 | 11,270 | 30,566 |
| Car | Left-rear passenger | Unbelted | 4,395 | 6,298 | 10,693 |
| Car | Right-rear passenger | Unbelted | 5,012 | 7,270 | 12,282 |
| Light truck | Driver | Unbelted | 3,994 | 24,882 | 28,876 |
| Light truck | Right-front passenger | Unbelted | 11,015 | 18,889 | 29,904 |
| Light truck | Driver | Belted | 2,071 | 6,338 | 8,409 |
| Light truck | Right-front passenger | Belted | 5,148 | 4,166 | 9,314 |
| Light truck | Left-rear passenger | Unbelted | 696 | 1,092 | 1,788 |
| Light truck | Right-rear passenger | Unbelted | 790 | 1,153 | 1,943 |
| Motorcycle | Passenger | Helmeted | 3,687 | 1,211 | 4,898 |
| Motorcycle | Passenger | Unhelmeted | 4,496 | 1,956 | 6,452 |
| **Totals** | | | **124,132** | **197,188** | **321,320** |

## III. RESULTS

To best illustrate the results produced by the double pair method, consider the case of belted drivers, aged 23–27 years (25-year-old (yo) drivers), in passenger cars, using belted front-right seat passengers as a control, matching airbag deployment, in crashes between 2010 and 2018. These results are given in Table III for each sex/age cohort under these conditions. Taking the weighted average across the controls presented in Table III for each set of subject occupants (belted/unbelted drivers and front-right passengers, with and without airbag deployment, in cars and light trucks), we produce Table IV, for 25yo vehicle occupants.

From Table III, for example, belted 25yo female drivers with airbag deployment were 8.1% (95% CI [-8.6, 21.9]%) more likely to die in a fatal crash than male drivers under the same conditions from 2010 to 2018. Taken with all other cases for 25yo female occupants, the overall risk (as shown in Table IV), is 10.7% (95% CI [9.7, 11.6]%) higher risk for females. This value is plotted as the corresponding value in Fig. 3. Note that the value is plotted at 23yo, not directly at 25yo. In different analyses, slightly different age bins were used to better capture the distribution of subject and control occupants. Bins were combined into five-year chunks, using the midpoint of the original groupings to determine membership in the final bin. Data were then plotted at the midpoint of each bin. Younger female drivers have a higher fatality risk than younger males when driving or sitting in the front passenger seat, regardless of seat-belt use. Fig. 4 shows the results of repeating this analysis for all crashes from 1975 to 2018. Fig. 5 is representative of the confidence interval estimation utilising the bootstrap method, and shows the distribution of R values calculated with 5000 repetitions for passenger car occupants.

In line with previous findings [11,12], the relative risk for female drivers is higher than that for males until approximately age 60. The highest difference in risk is between ages 20 and 40, with females ~20% more likely to die in a crash. This general trend is apparent when isolating vehicle type, seat-belt use, number of vehicles involved, urban vs. rural road type, and airbag deployment.

TABLE III
FEMALE VS. MALE FATALITY RISK, BELTED 25YO CAR DRIVERS, AIRBAG DEPLOYMENT, 2010–2018
(NOTE: CONTROL OCCUPANTS ARE BELTED, FRONT-RIGHT PASSENGERS)

| Control Occupant Characteristics, Age | Fatalities | | | | Ratios | | | |
|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | $r_1 = A/B$ | $r_2 = C/D$ | $R = r_1/r_2$ | ln ΔR (Eq. 13) |
| Male Passenger, 16–24yo | 24 | 26 | 132 | 118 | 0.923 | 1.119 | 0.825 | 0.059 |
| Male Passenger, 25–34yo | 54 | 48 | 83 | 87 | 1.125 | 0.954 | 1.179 | 0.040 |
| Male Passenger, 35–54yo | 12 | 11 | 21 | 27 | 1.091 | 0.778 | 1.403 | 0.129 |
| Male Passenger, 55+yo | 1 | 6 | 2 | 4 | 0.167 | 0.500 | 0.333 | 1.917 |
| Female Passenger, 16–24yo | 30 | 24 | 110 | 111 | 1.250 | 0.991 | 1.261 | 0.075 |
| Female Passenger, 25–34yo | 24 | 28 | 59 | 77 | 0.857 | 0.766 | 1.119 | 0.065 |
| Female Passenger, 35–54yo | 12 | 25 | 12 | 20 | 0.480 | 0.600 | 0.800 | 0.190 |
| | | | | | Weighted average | | 1.081 | |
| | | | | | 95% Confidence Interval | | | [0.914, 1.219] |

TABLE IV
VALUES OF R, FATALITY RISK TO 25YO FEMALES COMPARED TO 25YO MALES,
ALL SUBJECT OCCUPANTS, 2010–2018, MATCHED AIRBAG DEPLOYMENT

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total Fatalities | R | 95% CI |
|---|---|---|---|---|---|---|
| Car | Unbelted drivers, airbag deployed | 482 | 1,446 | 1,928 | 0.945 | [0.835, 1.079] |
| Car | Unbelted right front passengers, airbag deployed | 938 | 1,137 | 2,075 | 1.379 | [1.211, 1.519] |
| Car | Belted drivers, airbag deployed | 1,747 | 3,330 | 5,077 | 1.081 | [0.902, 1.223] |
| Car | Belted right front passengers, airbag deployed | 3,781 | 2,391 | 6,172 | 0.985 | [0.891, 1.164] |
| Car | Unbelted drivers, airbag not deployed | 300 | 852 | 1,152 | 1.254 | [0.932, 1.960] |
| Car | Unbelted right front passengers, airbag not deployed | 477 | 678 | 1,155 | 1.051 | [0.640, 1.354] |
| Car | Belted drivers, airbag not deployed | 653 | 1,484 | 2,137 | 1.280 | [1.011, 2.453] |
| Car | Belted right front passengers, airbag not deployed | 1,386 | 1,079 | 2,465 | 1.169 | [0.775, 1.450] |
| Light truck | Unbelted drivers, airbag deployed | 211 | 666 | 877 | 1.292 | [0.856, 1.823] |
| Light truck | Unbelted right front passengers, airbag deployed | 433 | 549 | 982 | 1.058 | [0.854, 1.380] |
| Light truck | Belted drivers, airbag deployed | 661 | 1,754 | 2,415 | 0.857 | [0.675, 1.182] |
| Light truck | Belted right front passengers, airbag deployed | 1,643 | 960 | 2,603 | 1.075 | [0.830, 1.441] |
| Light truck | Unbelted drivers, airbag not deployed | 436 | 1,502 | 1,938 | 0.976 | [0.671, 1.441] |
| Light truck | Unbelted right front passengers, airbag not deployed | 759 | 1,081 | 1,840 | 1.476 | [1.238, 2.013] |
| Light truck | Belted drivers, airbag not deployed | 489 | 1,349 | 1,838 | 1.213 | [0.961, 1.549] |
| Light truck | Belted right front passengers, airbag not deployed | 964 | 863 | 1,827 | 0.843 | [0.727, 1.203] |
| | | | **Weighted Average** | | 1.107 | [1.097, 1.116] |
| **Totals** | | **15,360** | **21,121** | **36,481** | | |

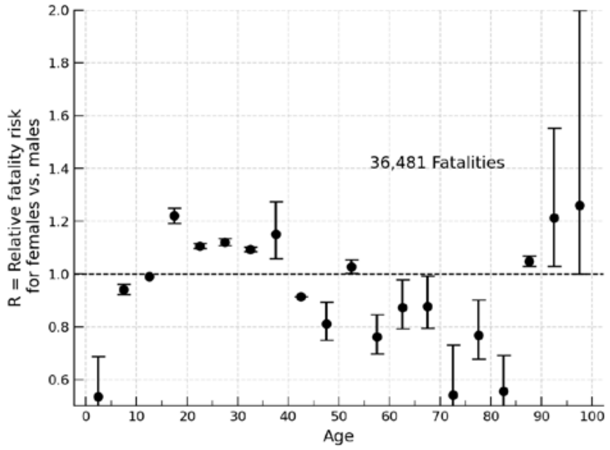


Fig. 3. All passenger car and light truck fatalities, 2010-2018, matched airbag deployment condition.

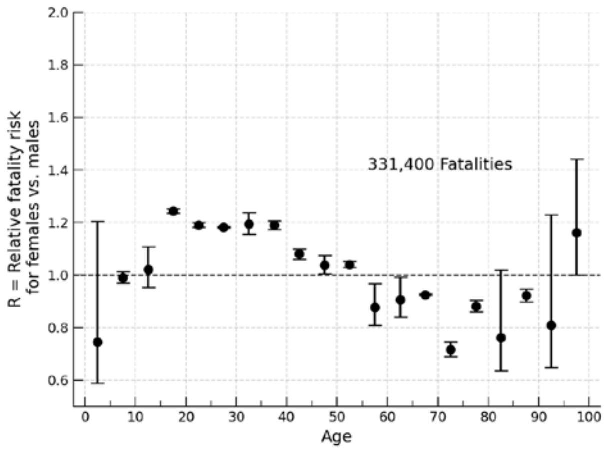


Fig. 4. All passenger car and light truck fatalities, 1975–2018, matched airbag deployment condition.

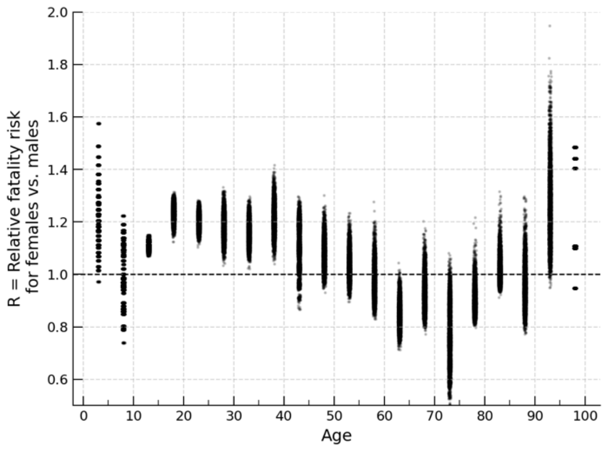


Fig. 5. Bootstrap distributions, passenger car fatalities 1975-2018, matched airbag deployment condition.

## IV. DISCUSSION

Despite significant advances in vehicle safety since 1975, female vehicle occupants involved in fatal crashes have a higher risk of death compared to males in matched circumstances. For example, taking the 25yo driver example introduced above, these results show that when in a fatal crash, the female occupant is approximately 20% more likely to suffer a fatal injury than a male occupant, regardless of seating position, airbag deployment, or seat-belt usage.

The results presented here are similar to those presented by Evans in 2001, for FARS data through 1998 [12]. Notably, the inclusion of data from before 2010 in the analysis presented in Fig. 4 mainly reduces the error assessment, without qualitatively altering the overall characteristics of the risk curve.

We have investigated several potential covariates that might explain these findings, including rural vs. urban crashes, vehicle mass differences by sex, drug and alcohol use by drivers, number of passengers, and number of vehicles involved. Figure 6 shows the distribution of some of these covariates. Figure 6(A) shows that the distribution of vehicles driven by females is similar to that for males, with a slightly higher proportion of female drivers using cars with vehicle masses of ~2,500 lb, but both sexes use heavier vehicles at the same rate. Figure 6(B) shows that male drivers are more frequently involved in single-car crashes compared to female drivers, but the proportion of crashes involving multiple vehicles is similar across sex. Figure 6(C) shows similar distributions with respect to the number of occupants in a vehicle. The lack of a large qualitative difference in these covariates would imply limited effect on the relative risk to drivers.

To verify this conclusion, the above analyses were applied to subsets of vehicle occupants involved in one-, two- and multi-vehicle crashes, as well as cases with one, two, or several vehicle occupants. As there are systematically more crashes of higher severity in rural areas [26], rural and urban crash locations were broken out into separate analyses as well, to serve as a surrogate for crash severity. The trend described above is robust across all of these cases (Appendix 2).

Physiological/anatomical differences are difficult to explore within the FARS dataset, but may provide an explanation. Current testing standards in the US primarily require the use of the 50th percentile Hybrid III (HIII) adult male ATD, with a few tests adding in the 5th percentile adult female as a passenger. Since the 5th percentile female ATD is primarily a dimensionally scaled version of the HIII male [27] thephysiological and anatomical differences between the sexes may not be completely reproduced in testing methods.

These results are not completely consistent with other well-described trends related to sex differences in injury. Females are more likely than males to suffer fractures past the age of 60 due to osteoporosis, and experience bone loss at an earlier age [28], and females tend to have greater age related bone density loss than males [29]. Increased fracture risk in elderly females, therefore, cannot explain the observed trends. Without detailed injury report data (not available within FARS), the cause of death cannot be determined for each case. This information would be valuable in parsing the differences between male and female crash survivability. We posit that there may be unobserved trends in the injury patterns, and therefore outcomes, between male and female occupants. These trends may be the result of unintentional vehicle design issues, or a potentially unexplained lack of biofidelity in the ATDs used in testing.

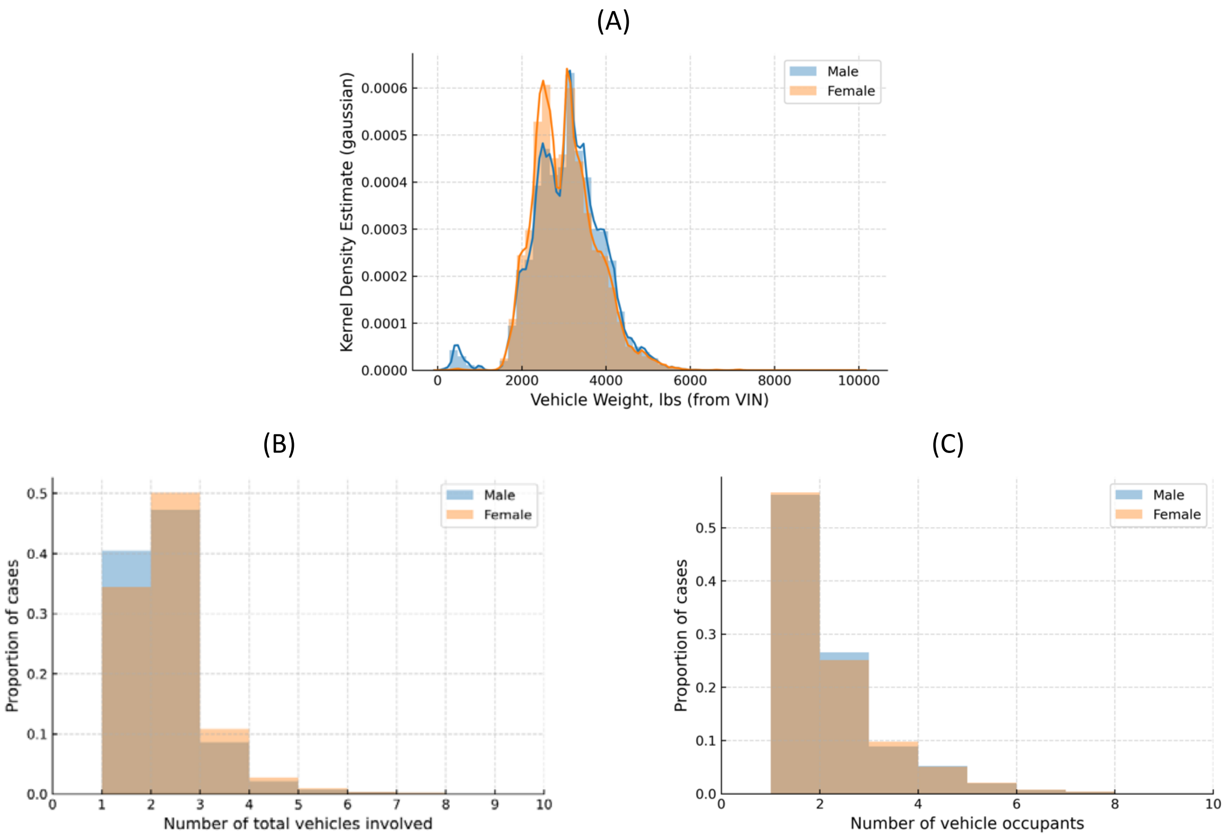


Fig. 6. Breakdown of potential covariates for male vs. female drivers: (A) Gaussian kernel density estimate of vehicle weight (in pounds); (B) normalised histogram of vehicles involved in a crash; (C) normalised histogram of number of vehicle occupants.

## V. CONCLUSIONS

There is an age-dependent risk to females compared with males in vehicle crashes, with younger females at higher risk of death than males. This difference is robust across vehicle type, number of passengers, crash severity, and year. It includes an increased risk peaking in the mid-30s. It is not attributable to vehicle weight, belt usage, or airbag deployment. Given recent advances in vehicle safety technologies, and a general trend towards fewer vehicle fatalities each year, the persistence of this trend is disturbing and requires further study. More work should be done to better understand the differences in female crash fatal and nonfatal outcomes compared to males. The higher risk to females discussed here demands exploration. The results presented above confirm past research, highlighting a decades-long issue that has clearly not been fully addressed. Known underlying physiological differences are insufficient to explain the phenomenon described in this study, and should be examined to ensure all occupants are well-protected in crash situations.

## VI. REFERENCES

[1] National Highway Traffic Safety Administration. (2018). Fatality Analysis Reporting System: Washington, DC.
[2] National Highway Traffic Safety Administration. (2019) *Fatality Analysis Reporting System (FARS) Analytical User's Manual, 1975-2018,* Washington, DC.
[3] Kahane, C.J. (2015) *Lives Saved by Vehicle Safety Technologies and Associated Federal Motor Vehicle Safety Standards, 1960 to 2012 – Passenger Cars and LTVs – With Reviews of 26 FMVSS and the Effectiveness of Their Associated Safety Technologies in Reducing Fatalities, Injuries, and Crashes,* Washington, DC.
[4] Viano, D.C. and Parenteau, C.S. (2016). Effectiveness of the revision to FMVSS 301: FARS and NASS-CDS analysis of fatalities and severe injuries in rear impacts. *Accident Analysis & Prevention*, **89**: pp.1-8
[5] Mayhew, D.R., Ferguson, S.A., Desmond, K.J., and Simpson, H.M. (2003). Trends in fatal crashes involving female drivers, 1975-1998. *Accident Analysis & Prevention*, **35**(3): pp.407-15
[6] Sivak, M. (2013). Female Drivers in the United States, 1963–2010: From a Minority to a Majority? *Traffic Injury Prevention*, **14**(3): pp.259-260
[7] Fucci, N.M.A. (2018) *Summary of Travel Trends: 2017 National Household Travel Survey,* Washington, DC.
[8] Tsai, V.W., Anderson, C.L., and Vaca, F.E. (2008). Young female drivers in fatal crashes: recent trends, 1995-2004. *Traffic Injury Prevention*, **9**(1): pp.65-9
[9] Romano, E., Kelley-Baker, T., and Voas, R.B. (2008). Female involvement in fatal crashes: Increasingly riskier or increasingly exposed? *Accident Analysis & Prevention*, **40**(5): pp.1781-1788
[10] Linder, A. and Svensson, M.Y. (2019). Road safety: the average male as a norm in vehicle occupant crash safety assessment. *Interdisciplinary Science Reviews*, **44**(2): pp.140-153
[11] Evans, L. (1988). Risk of fatality from physical trauma versus sex and age. *Journal of Trauma*, **28**(3): pp.368-78
[12] Evans, L. (2001). Female compared with male fatality risk from similar physical impacts. *Journal of Trauma*, **50**(2): pp.281-8
[13] Bose, D., Segui-Gomez, M., and Crandall, J.R. (2011). Vulnerability of female drivers involved in motor vehicle crashes: an analysis of US population at risk. *American Journal of Public Health*, **101**(12): pp.2368-73
[14] Abrams, M. (2020). mzabrams/fars-ircobi-abrams-2020 v1.0.0.
[15] Uieda, L., Soler, S., et al. (2020). Pooch: A friend to fetch your data files. *Journal of Open Source Software*, **5**(45)
[16] Virtanen, P., Gommers, R., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*, **17**(3): pp.261-272
[17] McKinney, W. Data structures for statistical computing in python. *Proceedings of Proceedings of the 9th Python in Science Conference*, 2010, Austin, TX.
[18] McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, **14**
[19] Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, **9**(3): pp.90-95
[20] Bilogur, A. (2018). Missingno: a missing data visualization suite. *The Journal of Open Source Software*, **3**(22)
[21] Abrams, M. (2020). mzabrams/fars-cleaner v1.0.1.
[22] Kluyver, T., Ragan-Kelley, B., et al. Jupyter Notebooks - a publishing format for reproducible computational workflows. *Proceedings of ELPUB*, 2016.
[23] Perez, F. and Granger, B.E. (2007). IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, **9**(3): pp.21-29
[24] Evans, L. (1986). Double pair comparison—A new method to determine how occupant characteristics affect fatality risk in traffic crashes. *Accident Analysis & Prevention*, **18**(3): pp.217-227
[25] Cummings, P., McKnight, B., and Weiss, N.S. (2003). Matched-pair cohort methods in traffic crash research. *Accident Analysis & Prevention*, **35**(1): pp.131-41
[26] Burgess, M. (2005) *Contrasting Rural and Urban Fatal Crashes 1994-2003,* Washington, DC.
[27] Saul, R.A., Pritz, H.B., et al. (2001). Description and performance of the Hybrid III three-year-old, six-year-old and small female test dummies in restraint system and out-of-position air bag environments. *Proceedings of 17th Enhanced Safety of Vehicles Conference*, 2001, Amsterdam, The Netherlands.

[28] Alswat, K.A. (2017). Gender Disparities in Osteoporosis. *Journal of Clinical Medicine Research*, **9**(5): pp.382-387

[29] Hannan, M.T., Felson, D.T., et al. (2000). Risk factors for longitudinal bone loss in elderly men and women: the Framingham Osteoporosis Study. *J Bone Miner Res*, **15**(4): pp.710-20

## VII. Appendix 1: Fatality Counts

TABLE AI

Distribution of 331,400 Fatally Injured Subject Occupants 1975–2018

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers, airbag deployed | 1,253 | 3,820 | 5,073 |
| Car | Unbelted right front passengers, airbag deployed | 2,408 | 3,157 | 5,565 |
| Car | Belted drivers, airbag deployed | 3,794 | 7,022 | 10,816 |
| Car | Belted right front passengers, airbag deployed | 7,909 | 4,979 | 12,888 |
| Car | Unbelted drivers, airbag not deployed | 16,994 | 54,196 | 71,190 |
| Car | Unbelted right front passengers, airbag not deployed | 38,478 | 43,043 | 81,521 |
| Car | Belted drivers, airbag not deployed | 8,060 | 15,424 | 23,484 |
| Car | Belted right front passengers, airbag not deployed | 19,296 | 11,270 | 30,566 |
| Light truck | Unbelted drivers, airbag deployed | 443 | 1,767 | 2,210 |
| Light truck | Unbelted right front passengers, airbag deployed | 941 | 1,282 | 2,223 |
| Light truck | Belted drivers, airbag deployed | 1,260 | 3,228 | 4,488 |
| Light truck | Belted right front passengers, airbag deployed | 3,011 | 1,862 | 4,873 |
| Light truck | Unbelted drivers, airbag not deployed | 3,994 | 24,882 | 28,876 |
| Light truck | Unbelted right front passengers, airbag not deployed | 11,015 | 18,889 | 29,904 |
| Light truck | Belted drivers, airbag not deployed | 2,071 | 6,338 | 8,409 |
| Light truck | Belted right front passengers, airbag not deployed | 5,148 | 4,166 | 9,314 |
| **Totals** | | **126,075** | **205,325** | **331,400** |

TABLE AII

Distribution of 112,032 Fatally Injured Subject Occupants 1999–2018

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers, airbag deployed | 1,148 | 3,545 | 4,693 |
| Car | Unbelted right front passengers, airbag deployed | 2,220 | 2,893 | 5,113 |
| Car | Belted drivers, airbag deployed | 3,587 | 6,620 | 10,207 |
| Car | Belted right front passengers, airbag deployed | 7,417 | 4,713 | 12,130 |
| Car | Unbelted drivers, airbag not deployed | 1,997 | 5,898 | 7,895 |
| Car | Unbelted right front passengers, airbag not deployed | 3,517 | 5,139 | 8,656 |
| Car | Belted drivers, airbag not deployed | 3,291 | 6,445 | 9,736 |
| Car | Belted right front passengers, airbag not deployed | 7,143 | 4,942 | 12,085 |
| Light truck | Unbelted drivers, airbag deployed | 428 | 1,712 | 2,140 |
| Light truck | Unbelted right front passengers, airbag deployed | 915 | 1,252 | 2,167 |
| Light truck | Belted drivers, airbag deployed | 1,236 | 3,123 | 4,359 |
| Light truck | Belted right front passengers, airbag deployed | 2,945 | 1,826 | 4,771 |
| Light truck | Unbelted drivers, airbag not deployed | 1,546 | 6,829 | 8,375 |
| Light truck | Unbelted right front passengers, airbag not deployed | 3,168 | 5,041 | 8,209 |
| Light truck | Belted drivers, airbag not deployed | 1,486 | 4,098 | 5,584 |
| Light truck | Belted right front passengers, airbag not deployed | 3,177 | 2,735 | 5,912 |
| **Totals** | | **45,221** | **66,811** | **112,032** |

TABLE AIII
DISTRIBUTION OF 36,481 FATALLY INJURED SUBJECT OCCUPANTS 2010–2018

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers, airbag deployed | 482 | 1,446 | 1,928 |
| Car | Unbelted right front passengers, airbag deployed | 938 | 1,137 | 2,075 |
| Car | Belted drivers, airbag deployed | 1,747 | 3,330 | 5,077 |
| Car | Belted right front passengers, airbag deployed | 3,781 | 2,391 | 6,172 |
| Car | Unbelted drivers, airbag not deployed | 300 | 852 | 1,152 |
| Car | Unbelted right front passengers, airbag not deployed | 477 | 678 | 1,155 |
| Car | Belted drivers, airbag not deployed | 653 | 1,484 | 2,137 |
| Car | Belted right front passengers, airbag not deployed | 1,386 | 1,079 | 2,465 |
| Light truck | Unbelted drivers, airbag deployed | 211 | 666 | 877 |
| Light truck | Unbelted right front passengers, airbag deployed | 433 | 549 | 982 |
| Light truck | Belted drivers, airbag deployed | 661 | 1,754 | 2,415 |
| Light truck | Belted right front passengers, airbag deployed | 1,643 | 960 | 2,603 |
| Light truck | Unbelted drivers, airbag not deployed | 436 | 1,502 | 1,938 |
| Light truck | Unbelted right front passengers, airbag not deployed | 759 | 1,081 | 1,840 |
| Light truck | Belted drivers, airbag not deployed | 489 | 1,349 | 1,838 |
| Light truck | Belted right front passengers, airbag not deployed | 964 | 863 | 1,827 |
| **Totals** | | **15,360** | **21,121** | **36,481** |

TABLE AIV
DISTRIBUTION OF 14,796 FATALLY INJURED SUBJECT OCCUPANTS 2015–2018

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers, airbag deployed | 212 | 578 | 790 |
| Car | Unbelted right front passengers, airbag deployed | 414 | 471 | 885 |
| Car | Belted drivers, airbag deployed | 811 | 1,629 | 2,440 |
| Car | Belted right front passengers, airbag deployed | 1,730 | 1,125 | 2,855 |
| Car | Unbelted drivers, airbag not deployed | 96 | 266 | 362 |
| Car | Unbelted right front passengers, airbag not deployed | 161 | 206 | 367 |
| Car | Belted drivers, airbag not deployed | 209 | 485 | 694 |
| Car | Belted right front passengers, airbag not deployed | 415 | 346 | 761 |
| Light truck | Unbelted drivers, airbag deployed | 94 | 234 | 328 |
| Light truck | Unbelted right front passengers, airbag deployed | 192 | 217 | 409 |
| Light truck | Belted drivers, airbag deployed | 321 | 849 | 1,170 |
| Light truck | Belted right front passengers, airbag deployed | 816 | 463 | 1,279 |
| Light truck | Unbelted drivers, airbag not deployed | 139 | 480 | 619 |
| Light truck | Unbelted right front passengers, airbag not deployed | 287 | 339 | 626 |
| Light truck | Belted drivers, airbag not deployed | 186 | 424 | 610 |
| Light truck | Belted right front passengers, airbag not deployed | 331 | 270 | 601 |
| **Totals** | | **6,414** | **8,382** | **14,796** |

TABLE AV

DISTRIBUTION OF 321,320 FATALLY INJURED SUBJECT OCCUPANTS 1975–2018, NO AIRBAG DEPLOYMENT

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers | 16,994 | 54,196 | 71,190 |
| Car | Unbelted right front passengers | 38,478 | 43,043 | 81,521 |
| Car | Belted drivers | 8,060 | 15,424 | 23,484 |
| Car | Belted right front passengers | 19,296 | 11,270 | 30,566 |
| Car | Unbelted left rear passengers | 4,395 | 6,298 | 10,693 |
| Car | Unbelted right rear passengers | 5,012 | 7,270 | 12,282 |
| Light truck | Unbelted drivers | 3,994 | 24,882 | 28,876 |
| Light truck | Unbelted right front passengers | 11,015 | 18,889 | 29,904 |
| Light truck | Belted drivers | 2,071 | 6,338 | 8,409 |
| Light truck | Belted right front passengers | 5,148 | 4,166 | 9,314 |
| Light truck | Unbelted left rear passengers | 696 | 1,092 | 1,788 |
| Light truck | Unbelted right rear passengers | 790 | 1,153 | 1,943 |
| Motorcycle | Helmeted Motorcycle Passenger | 3,687 | 1,211 | 4,898 |
| Motorcycle | Unhelmeted Motorcycle Passenger | 4,496 | 1,956 | 6,452 |
| **Totals** | | **124,132** | **197,188** | **321,320** |

TABLE AVI

DISTRIBUTION OF 75,260 FATALLY INJURED SUBJECT OCCUPANTS 1999–2018, NO AIRBAG DEPLOYMENT

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers | 1,997 | 5,898 | 7,895 |
| Car | Unbelted right front passengers | 3,517 | 5,139 | 8,656 |
| Car | Belted drivers | 3,291 | 6,445 | 9,736 |
| Car | Belted right front passengers | 7,143 | 4,942 | 12,085 |
| Car | Unbelted left rear passengers | 393 | 758 | 1,151 |
| Car | Unbelted right rear passengers | 493 | 967 | 1,460 |
| Light truck | Unbelted drivers | 1,546 | 6,829 | 8,375 |
| Light truck | Unbelted right front passengers | 3,168 | 5,041 | 8,209 |
| Light truck | Belted drivers | 1,486 | 4,098 | 5,584 |
| Light truck | Belted right front passengers | 3,177 | 2,735 | 5,912 |
| Light truck | Unbelted left rear passengers | 280 | 484 | 764 |
| Light truck | Unbelted right rear passengers | 373 | 530 | 903 |
| Motorcycle | Helmeted Motorcycle Passenger | 1,979 | 181 | 2,160 |
| Motorcycle | Unhelmeted Motorcycle Passenger | 2,132 | 238 | 2,370 |
| **Totals** | | **30,975** | **44,285** | **75,260** |

TABLE AVII

DISTRIBUTION OF 17,016 FATALLY INJURED SUBJECT OCCUPANTS 2010–2018, NO AIRBAG DEPLOYMENT

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers | 300 | 852 | 1,152 |
| Car | Unbelted right front passengers | 477 | 678 | 1,155 |
| Car | Belted drivers | 653 | 1,484 | 2,137 |
| Car | Belted right front passengers | 1,386 | 1,079 | 2,465 |
| Car | Unbelted left rear passengers | 53 | 88 | 141 |
| Car | Unbelted right rear passengers | 62 | 131 | 193 |
| Light truck | Unbelted drivers | 436 | 1,502 | 1,938 |
| Light truck | Unbelted right front passengers | 759 | 1,081 | 1,840 |
| Light truck | Belted drivers | 489 | 1,349 | 1,838 |
| Light truck | Belted right front passengers | 964 | 863 | 1,827 |
| Light truck | Unbelted left rear passengers | 64 | 102 | 166 |
| Light truck | Unbelted right rear passengers | 75 | 121 | 196 |
| Motorcycle | Helmeted Motorcycle Passenger | 904 | 55 | 959 |
| Motorcycle | Unhelmeted Motorcycle Passenger | 928 | 81 | 1,009 |
| **Totals** | | **7,550** | **9,466** | **17,016** |

TABLE AVIII

DISTRIBUTION OF 5,216 FATALLY INJURED SUBJECT OCCUPANTS 2015–2018, NO AIRBAG DEPLOYMENT

| Vehicle | Subject Occupant | Female Fatalities | Male Fatalities | Total |
|---|---|---|---|---|
| Car | Unbelted drivers | 96 | 266 | 362 |
| Car | Unbelted right front passengers | 161 | 206 | 367 |
| Car | Belted drivers | 209 | 485 | 694 |
| Car | Belted right front passengers | 415 | 346 | 761 |
| Car | Unbelted left rear passengers | 6 | 8 | 14 |
| Car | Unbelted right rear passengers | 9 | 27 | 36 |
| Light truck | Unbelted drivers | 139 | 480 | 619 |
| Light truck | Unbelted right front passengers | 287 | 339 | 626 |
| Light truck | Belted drivers | 186 | 424 | 610 |
| Light truck | Belted right front passengers | 331 | 270 | 601 |
| Light truck | Unbelted left rear passengers | 15 | 24 | 39 |
| Light truck | Unbelted right rear passengers | 16 | 18 | 34 |
| Motorcycle | Helmeted Motorcycle Passenger | 191 | 17 | 208 |
| Motorcycle | Unhelmeted Motorcycle Passenger | 215 | 30 | 245 |
| **Totals** | | **2,276** | **2,940** | **5,216** |

## VIII. APPENDIX 2: GRAPHICAL RESULTS OF ADDITIONAL ANALYSES

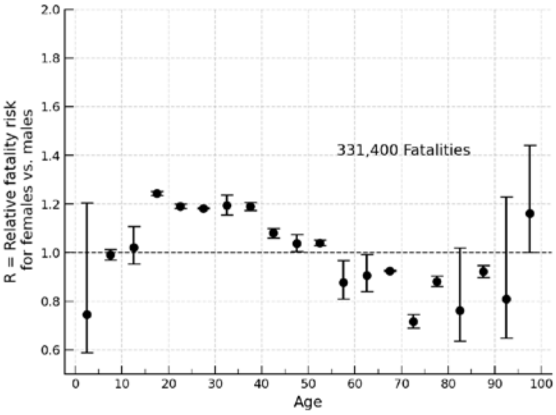


Fig. A1a. Relative risk, females to males, all fatalities 1975–2018, matched airbag conditions.

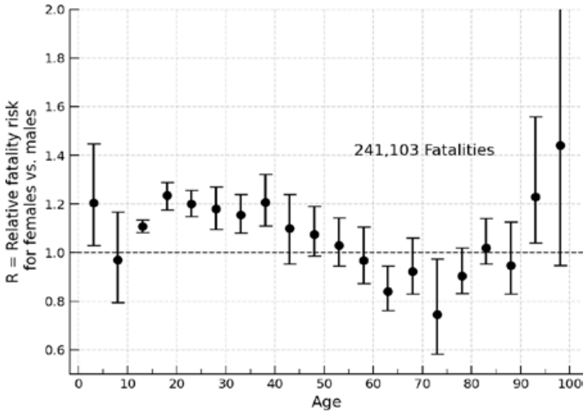


Fig. A1b. Relative risk, females to males, passenger car fatalities 1975–2018, matched airbag conditions.

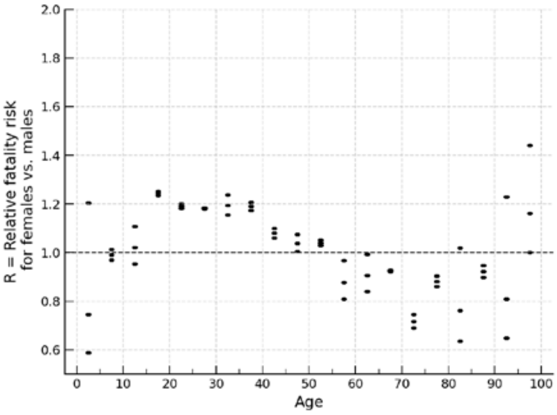


Fig. A1c. Distribution of 5000 bootstrap iterations, all fatalities 1975–2018, matched airbag conditions.

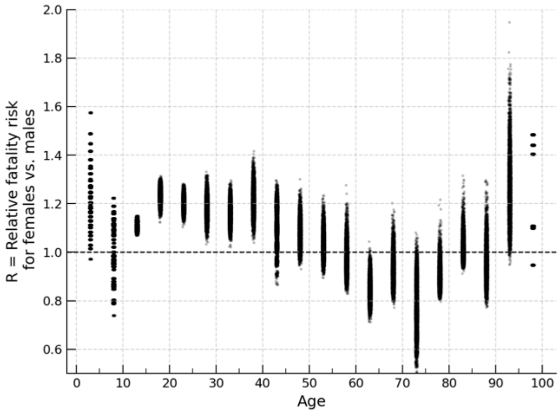


Fig. A1d. Distribution of 5000 bootstrap iterations, passenger car fatalities 1975–2018, matched airbag conditions.

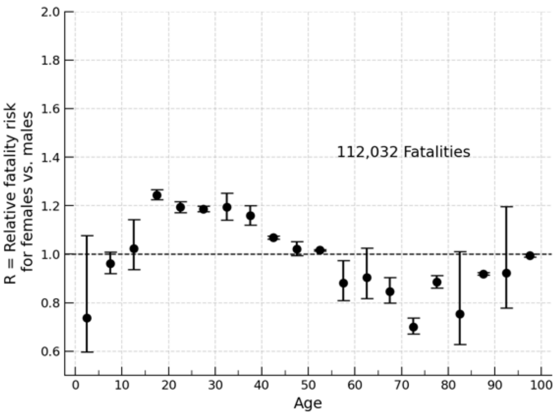


Fig. A2a. Relative risk, females to males, all fatalities 1999–2018, matched airbag conditions.

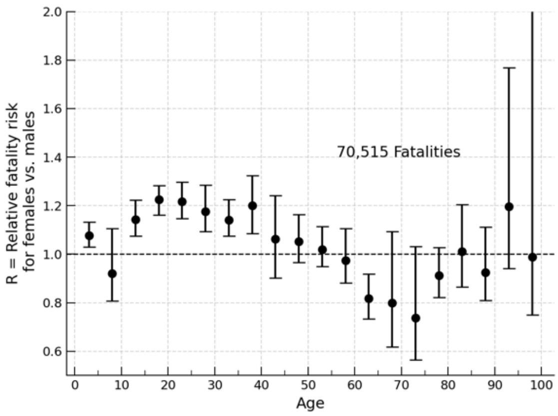


Fig. A2b. Relative risk, females to males, passenger car fatalities 1999–2018, matched airbag conditions.

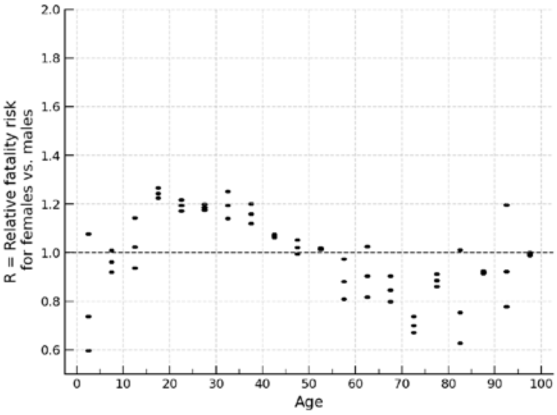


Fig. A2c. Distribution of 5000 bootstrap iterations, all fatalities 1999–2018, matched airbag conditions.

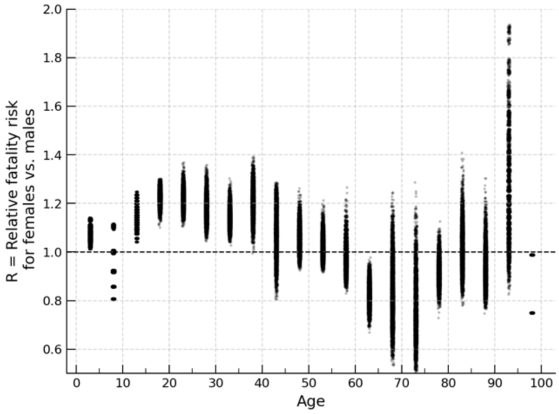


Fig. A2d. Distribution of 5000 bootstrap iterations, passenger car fatalities 1999–2018, matched airbag conditions.

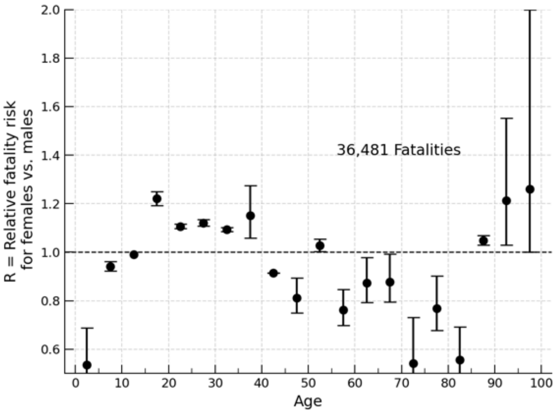


Fig. A3a. Relative risk, females to males, all fatalities 2010–2018, matched airbag conditions.

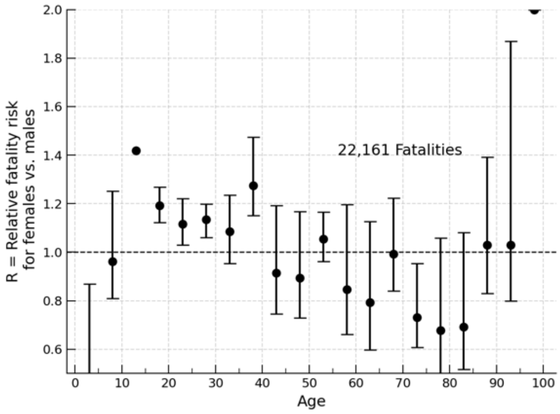


Fig. A3b. Relative risk, females to males, passenger car fatalities 2010–2018, matched airbag conditions.

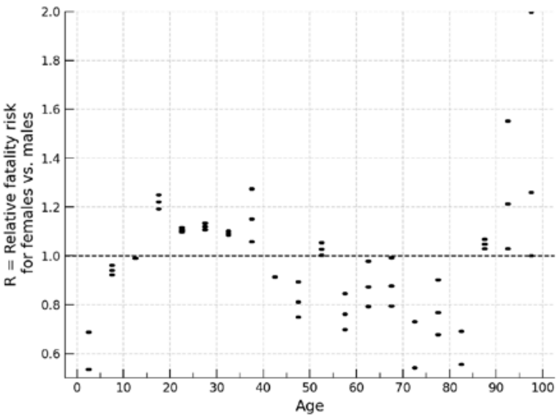


Fig. A3c. Distribution of 5000 bootstrap iterations, females to males, all fatalities 2010–2018, matched airbag conditions.

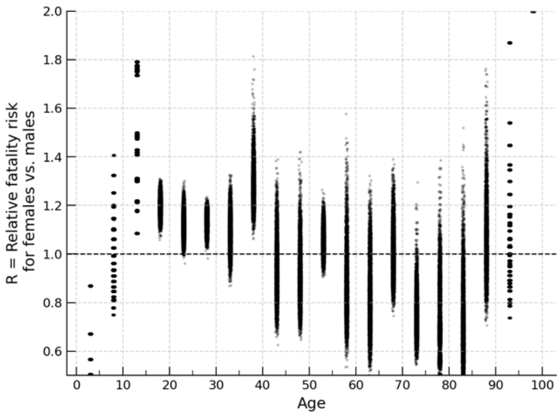


Fig. A3d. Distribution of 5000 bootstrap iterations, passenger car fatalities 2010–2018, matched airbag conditions.

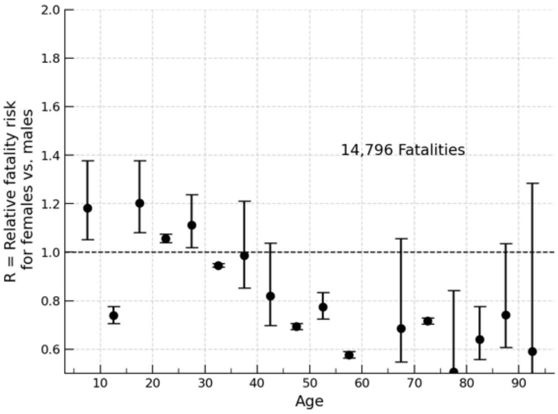


Fig. A4a. Relative risk, females to males, all fatalities 2015–2018, matched airbag conditions.

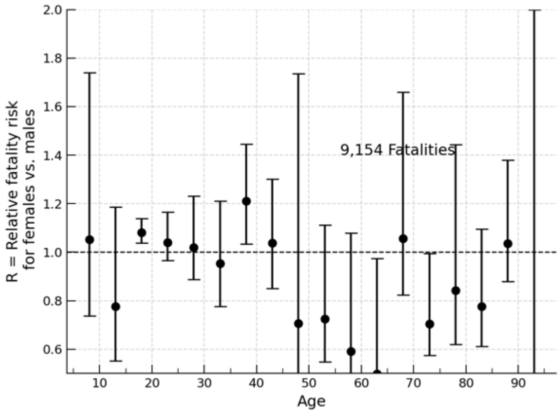


Fig. A4b. Relative risk, females to males, passenger car fatalities 2015–2018, matched airbag conditions.

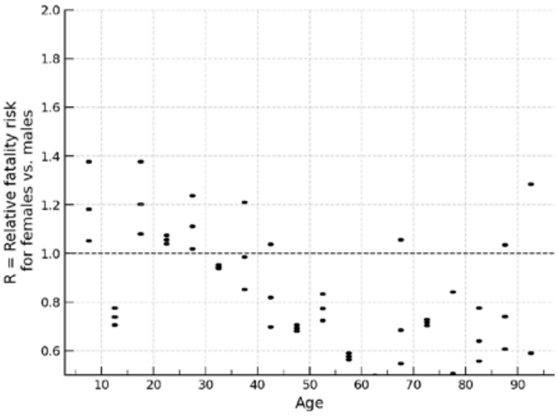


Fig. A4c. Distribution of 5000 bootstrap iterations, females to males, all fatalities 2015–2018, matched airbag conditions.

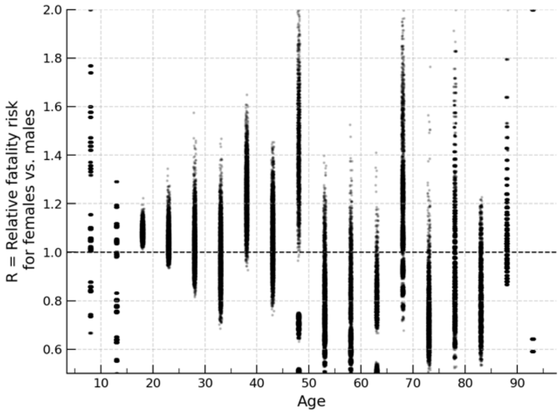


Fig. A4d. Distribution of 5000 bootstrap iterations, passenger car fatalities 2015–2018, matched airbag conditions.

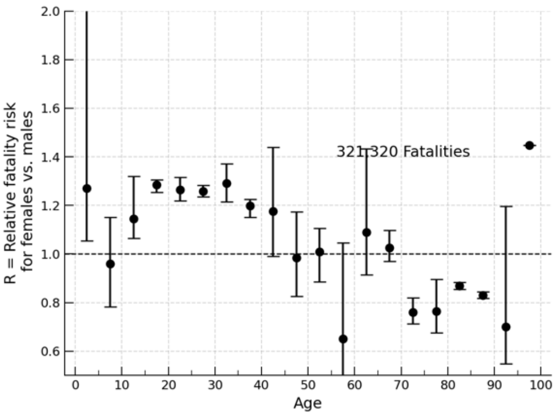


Fig. A5a. Relative risk, females to males, all fatalities 1975–2018, no airbag deployment.

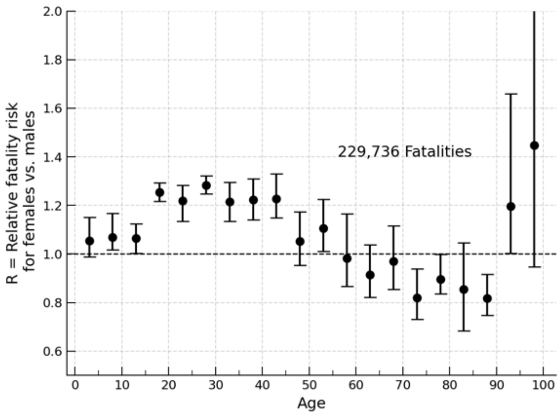


Fig. A5b. Relative risk, females to males, passenger car fatalities 1975–2018, no airbag deployment.

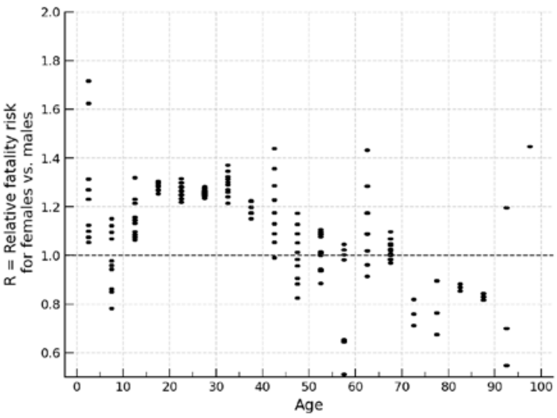


Fig. A5c. Distribution of 5000 bootstrap iterations, females to males, all fatalities 1975–2018, no airbag deployment.

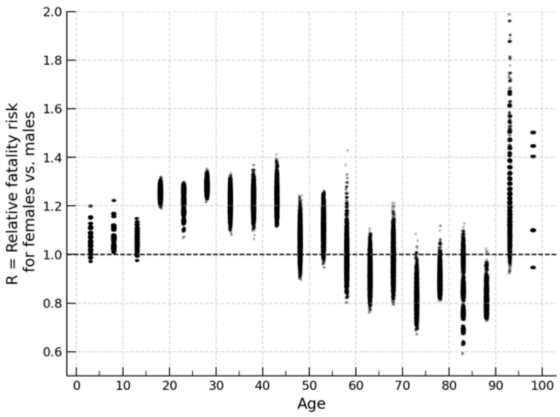


Fig. A5d. Distribution of 5000 bootstrap iterations, passenger car fatalities 1975–2018, no airbag deployment.

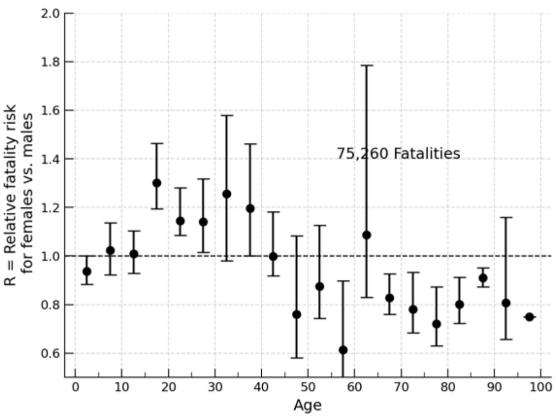


Fig. A6a. Relative risk, females to males, all fatalities 1999–2018, no airbag deployment.

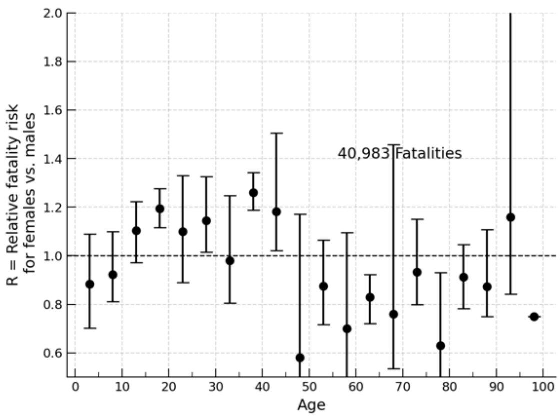


Fig. A6b. Relative risk, females to males, passenger car fatalities 1999–2018, no airbag deployment.

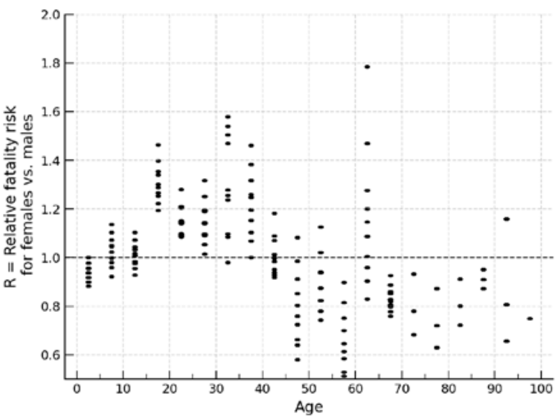


Fig. A6c. Distribution of 5000 bootstrap iterations, females to males, all fatalities 1999–2018, no airbag deployment.

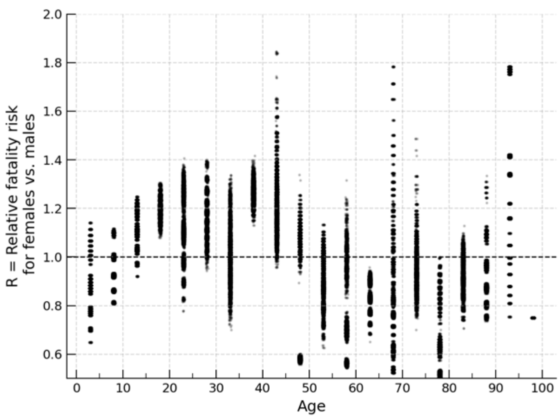


Fig. A6d. Distribution of 5000 bootstrap iterations, passenger car fatalities 1999–2018, no airbag deployment.

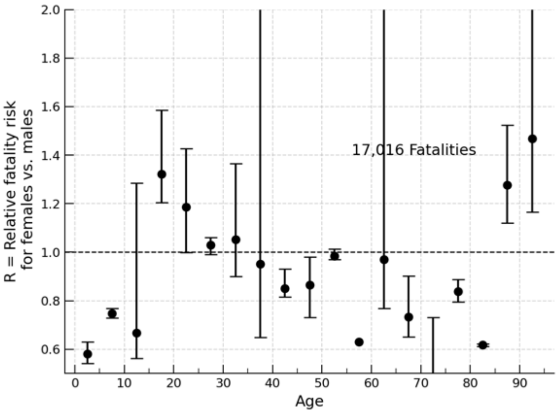


Fig. A7a. Relative risk, females to males, all fatalities 2010–2018, no airbag deployment.

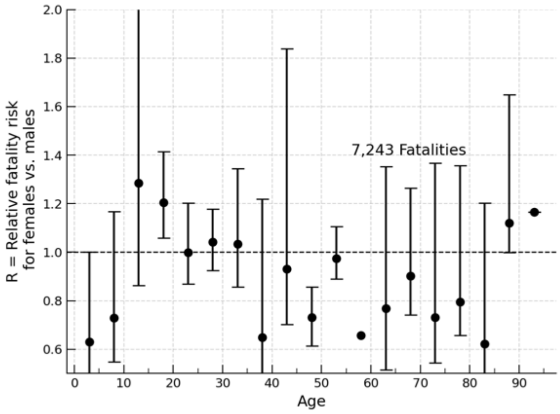


Fig. A7b. Relative risk, females to males, passenger car fatalities 2010–2018, no airbag deployment.

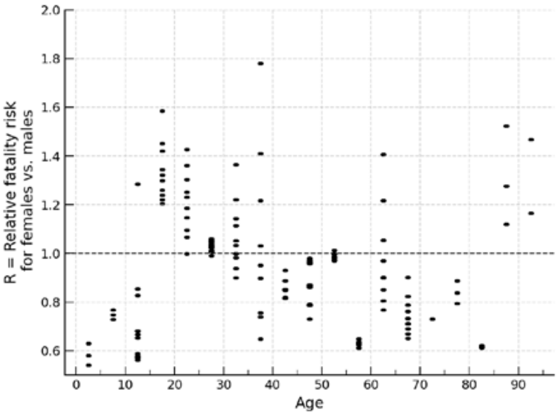


Fig. A7c. Distribution of 5000 bootstrap iterations, females to males, all fatalities 2010–2018, no airbag deployment.

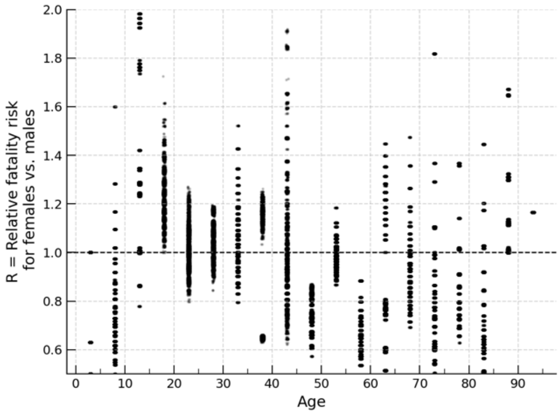


Fig. A7d. Distribution of 5000 bootstrap iterations, passenger car fatalities 2010–2018, no airbag deployment.

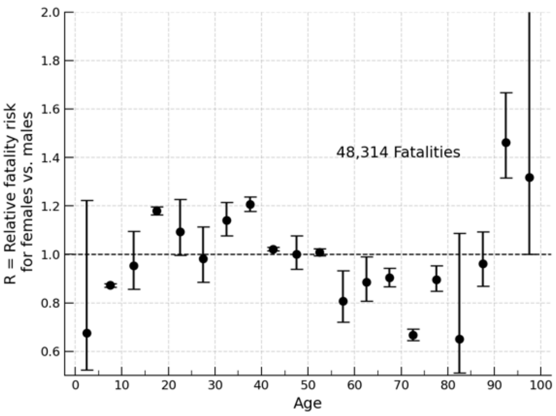


Fig. A8a. Relative risk, females to males, all fatalities 1975–2018, airbag deployment.

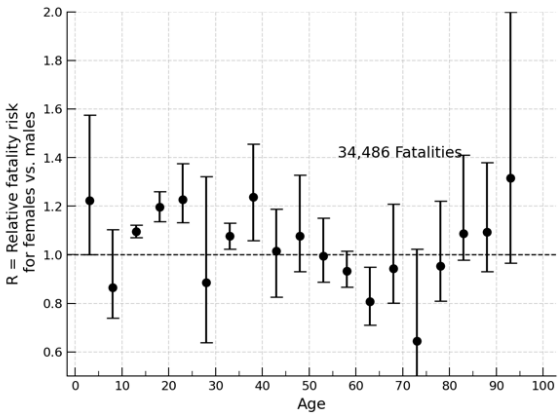


Fig. A8b. Relative risk, females to males, passenger car fatalities 1975–2018, airbag deployment.

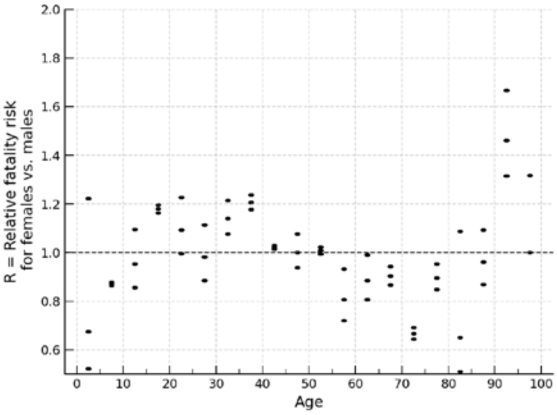


Fig. A8c. Distribution of 5000 bootstrap iterations, females to males, all fatalities 1975–2018, airbag deployment.

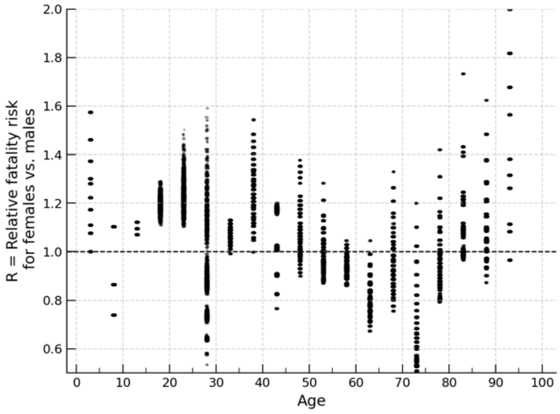


Fig. A8d. Distribution of 5000 bootstrap iterations, passenger car fatalities 1975–2018, airbag deployment.

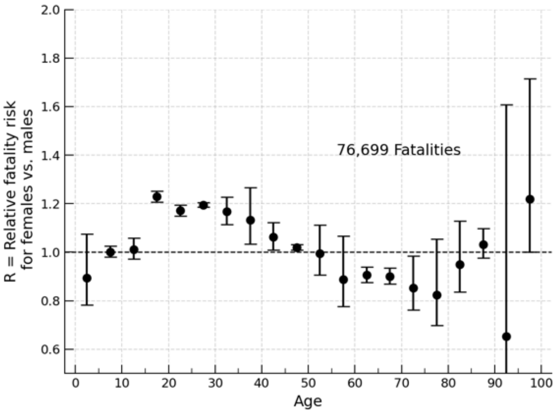


Fig. A9a. Relative risk, females to males, fatalities 2010–2018, rural crashes.

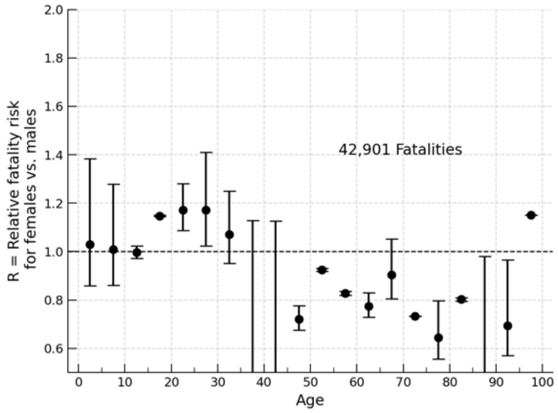


Fig. A9b. Relative risk, females to males, fatalities 2010–2018, urban crashes.

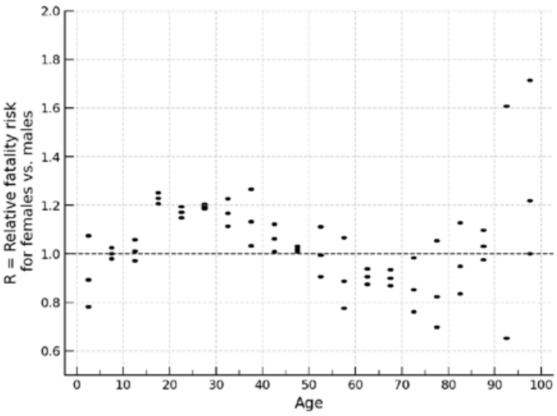


Fig. A9c. Distribution of 5000 bootstrap iterations, females to males, fatalities 2010–2018, rural crashes.

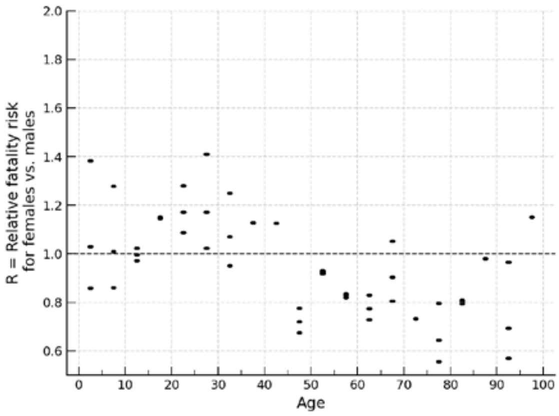


Fig. A9d. Distribution of 5000 bootstrap iterations, fatalities 2010–2018, urban crashes.

## IX. APPENDIX 2: JUPYTER ANALYSIS TEMPLATE

# 0.0 FINAL RUN TEMPLATE

## A. Summary

Explores a subset of the FARS dataset using the double pair comparison method. Notebook prepared for IRCOBI Europe 2020.

## B. Changes¶

- 1/20/2020 v1.0 Mitchell Abrams
- 3/26/2020 v2.0 Mitchell Abrams: update to new loaders, simplify report
- 3/29/2020 v3.0 Mitchell Abrams: modify for Papermill parameterized runs, update figures and output paths
- 4/9/2020 v4.0 Mitchell Abrams: modify with new variance and compositing as suggested by Cummings 2003.

### *Parameters for running with Papermill*

All values in the following cell (tagged `parameters`) are adjustable using Papermill to run specific analyses.

In [ ]:

```
# These are the parameters for runs with Papermill

image_outputs = 'default'

merged = None

driver_only = False

start_year = 1975
end_year = 2018

final_runs = False,

vehicle_types = ['PASSENGER_CAR', 'LIGHT_TRUCK_OR_VAN', 'MOTORCYCLE']

base_filter = f"'PASSENGER_CAR' and (SEX_C != 'Unknown') and (SEX_S != 'Unknown') a
nd "\
                "(AGE_C < 99) and (AGE_S < 99)"

subsets = [
    {'data': "SEAT_POS_C == 'Front Seat - Right Side' and "
                    "RESTRAINTS_C == 'Not Used' and RESTRAINTS_S == 'Not Used'",
     'subject': {'sex': 'SEX_S', 'age': 'AGE_S',
                 'dead': 'DEAD_S', 'id': 'PER_ID_S',
                 'bin': [15] + [x for x in range(22, 101, 5)]
                },
     'control': {'sex': 'SEX_C', 'age': 'AGE_C',
                 'dead': 'DEAD_C', 'id': 'PER_ID_C',
                 'bin': [15, 24, 34, 54, 100]
                },
```

```
 'title': 'Unbelted drivers'
},  #Unbelted drivers
{'data': "SEAT_POS_C == 'Front Seat - Right Side' and "
         "RESTRAINTS_C == 'Not Used' and RESTRAINTS_S == 'Not Used'",
 'subject': {'sex': 'SEX_C', 'age': 'AGE_C',
             'dead': 'DEAD_C', 'id': 'PER_ID_C',
             'bin': [x for x in range(0, 101, 5)]
            },
 'control': {'sex': 'SEX_S', 'age': 'AGE_S',
             'dead': 'DEAD_S', 'id': 'PER_ID_S',
             'bin': [15, 24, 34, 54, 100]
            },
 'title': 'Unbelted right front passengers'
},  #Unbelted front right passengers
{'data': "SEAT_POS_C == 'Front Seat - Right Side' and "
         "RESTRAINTS_C == 'Used' and RESTRAINTS_S == 'Used'",
 'subject': {'sex': 'SEX_S', 'age': 'AGE_S',
             'dead': 'DEAD_S', 'id': 'PER_ID_S',
             'bin': [15] + [x for x in range(22, 101, 5)]
            },
 'control': {'sex': 'SEX_C', 'age': 'AGE_C',
             'dead': 'DEAD_C', 'id': 'PER_ID_C',
             'bin': [15, 24, 34, 54, 100]
            },
 'title': 'Belted drivers'
},  #Belted drivers
{'data': "SEAT_POS_C == 'Front Seat - Right Side' and "
         "RESTRAINTS_C == 'Used' and RESTRAINTS_S == 'Used'",
 'subject': {'sex': 'SEX_C', 'age': 'AGE_C',
             'dead': 'DEAD_C', 'id': 'PER_ID_C',
             'bin': [x for x in range(0, 101, 5)]
            },
 'control': {'sex': 'SEX_S', 'age': 'AGE_S',
             'dead': 'DEAD_S', 'id': 'PER_ID_S',
             'bin': [15, 24, 34, 54, 100]
            },
 'title': 'Belted right front passengers'
},  #Belted front right passengers
{'data': "SEAT_POS_C == 'Second Seat - Left Side' and "
         "RESTRAINTS_C == 'Not Used' and RESTRAINTS_S == 'Not Used'",
 'subject': {'sex': 'SEX_C', 'age': 'AGE_C',
             'dead': 'DEAD_C', 'id': 'PER_ID_C',
             'bin': [x for x in range(0, 101, 5)]
            },
 'control': {'sex': 'SEX_S', 'age': 'AGE_S',
             'dead': 'DEAD_S', 'id': 'PER_ID_S',
             'bin': [15, 24, 34, 54, 100]
            },
 'title': 'Unbelted left rear passengers'
},  #Unbelted rear left passengers
```

```
    {'data': "SEAT_POS_C == 'Second Seat – Right Side' and "
             "RESTRAINTS_C == 'Not Used' and RESTRAINTS_S == 'Not Used'",
     'subject': {'sex': 'SEX_C', 'age': 'AGE_C',
                 'dead': 'DEAD_C', 'id': 'PER_ID_C',
                 'bin': [x for x in range(0, 101, 5)]
                },
     'control': {'sex': 'SEX_S', 'age': 'AGE_S',
                 'dead': 'DEAD_S', 'id': 'PER_ID_S',
                 'bin': [15, 24, 34, 54, 100]
                },
     'title': 'Unbelted right rear passengers'
    }, #Unbelted rear right passengers
]

mcycl_subsets = [{'data': "HELMETED_C == 'Helmeted' and HELMETED_S == 'Helmeted'",
         'subject': {'sex': 'SEX_C', 'age': 'AGE_C',
                     'dead': 'DEAD_C', 'id': 'PER_ID_C',
                     'bin': [x for x in range(0, 101, 5)]
                    },
         'control': {'sex': 'SEX_S', 'age': 'AGE_S',
                     'dead': 'DEAD_S', 'id': 'PER_ID_S',
                     'bin': [15, 24, 34, 54, 100]
                    },
         'title': 'Helmeted Motorcycle Passenger'
        },
          {'data': "HELMETED_C == 'Not Helmeted' and HELMETED_S == 'Not Helmeted'",
         'subject': {'sex': 'SEX_C', 'age': 'AGE_C',
                     'dead': 'DEAD_C', 'id': 'PER_ID_C',
                     'bin': [x for x in range(0, 101, 5)]
                    },
         'control': {'sex': 'SEX_S', 'age': 'AGE_S',
                     'dead': 'DEAD_S', 'id': 'PER_ID_S',
                     'bin': [15, 24, 34, 54, 100]
                    },
         'title': 'Unhelmeted Motorcycle Passenger'
        }]
bootstrap_iters = 1000
```

***Setup environment***

Prepares the output paths, loads required libraries, prepares plotting stylesheet.

In [ ]:

```
import os
import sys
from pathlib import Path
module_path = os.path.abspath(os.path.join('../..'))
if module_path not in sys.path:
    sys.path.append(module_path)

if final_runs:
    fig_out = Path(module_path) / "reports" / "final" / "figures" / image_outputs
```

```
else:
    fig_out = Path(module_path) / "reports" / "figures" / image_outputs
fig_out.mkdir(parents=True, exist_ok=True)
```

In [ ]:

```
%reload_ext autoreload
%autoreload 2

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
plt.style.use(['ircobi'])

from matplotlib.ticker import MultipleLocator
from sklearn.utils import resample

%matplotlib inline

import logging
logger = logging.getLogger("distributed.utils_perf")
logger.setLevel(logging.ERROR)

from src.data.data_loader_pooch import *
from src.data.fars_utils import createPerID, getWeightedAvg, standard_error

from tqdm.notebook import tqdm

pd.set_option('precision', 3)
pd.set_option('display.max_columns', 999)
```

## C. Load pickles

Data must be written out as a pre-merged dataset prior to running this script.

In [ ]:

```
data = pd.read_pickle(Path(merged))
data = data.query(f"{start_year} <= YEAR <= {end_year}")
```

## D. Method Definitions

Defines methods for plotting, combining analyses, and executing the double pair comparison

In [ ]:

```
def plot_doublepair(vehicle_results):

    results = vehicle_results['results']
    titles = vehicle_results['titles']
    counts = vehicle_results['counts']
    nres = len(results)
    print(nres)
    if nres <= 6:
        fig, axes = plt.subplots(3,2, figsize=(10,9), sharex=True, sharey=True)
    else:
```

```
        fig, axes = plt.subplots(int(nres/2),2, figsize=(10,int(3*nres/2)), sharex=
True, sharey=True)
    a=[]
    for i, row in enumerate(axes):
        for j, ax in enumerate(row):
            a.append(ax)

    for i, ax in enumerate(a):
        ax.axhline(1, color='black', lw=1, ls='--')
        #print(results)
        if not results[i].empty:
            x = results[i]['Age']
            y = results[i]['R']
            errUp = results[i]['UpLim'] - y
            errDown = y - results[i]['DownLim']
            ax.scatter(x, y, c='black', marker='o')
            ax.errorbar(x, y, yerr=[errDown, errUp], capsize=5, fmt='ko', linestyle
='')

            ax.set_ylim(.5, 2)
            """results[i].plot(x='Age', y='R', yerr='DeltaR', capsize=5, fmt='ko',
                      ylim = (.5, 2), title=titles[i],
                       ax = ax, legend=False)"""
            ax.text(55, 1.85, f"{counts[i]: ,} Fatalities", fontsize=12)
        else:
            ax.plot(0, 0)


        ax.xaxis.set_major_locator(MultipleLocator(10))
        ax.xaxis.set_minor_locator(MultipleLocator(5))
        ax.tick_params(which='major', length=7)
        ax.tick_params(which='minor', length=4)
        #axis.yaxis.set_major_locator(MultipleLocator(10))
        ax.yaxis.set_minor_locator(MultipleLocator(.1))
    fig.text(0.0, 0.5, "R = Relative fatality risk for females vs. males", va='cent
er', rotation='vertical',
             fontsize='large')
    plt.show()
    return fig
```

In [ ]:

```
def plot_driver_only(vehicle_results):
    fig, axes = plt.subplots(1,2, figsize=(10,3), sharex=True, sharey=True)
    results = vehicle_results['results']
    titles = vehicle_results['titles']
    counts = vehicle_results['counts']
    a=[]
    for i, ax in enumerate(axes):
        a.append(ax)

    for i, ax in enumerate(a):
```

```
        ax.axhline(1, color='black', lw=1, ls='--')
        results[i].plot(x='Age', y='R', yerr='DeltaR', capsize=5, fmt='ko',
                      ylim = (.5, 2), title=titles[i],
                       ax = ax, legend=False)
        ax.text(55, 1.85, f"{counts[i]: ,} Fatalities", fontsize=12)

        ax.xaxis.set_major_locator(MultipleLocator(10))
        ax.xaxis.set_minor_locator(MultipleLocator(5))
        ax.tick_params(which='major', length=7)
        ax.tick_params(which='minor', length=4)
        #axis.yaxis.set_major_locator(MultipleLocator(10))
        ax.yaxis.set_minor_locator(MultipleLocator(.1))
    fig.text(0.0, 0.5, "R = Relative fatality risk for females vs. males", va='cent
er', rotation='vertical',
             fontsize='large')
    plt.show()
    return fig
```

In [ ]:

```
def combine_results(results, bootstrap_iters):

    combined = pd.concat(results,sort=True).reset_index()
    new_idx = pd.interval_range(start=0, end=100, freq=5)
    replacer = []

    for x in combined['Subject Age']:
        curavg = x.mid
        for i in new_idx:
            if curavg in i:
                replacer.append(i)

    combined['Subject New'] = replacer

    combined['Age'] = combined['Subject New'].apply(lambda x: x.mid + 0.5)
    combined = combined.groupby('Age')
    combined = combined.apply(bootstrap, bootstrap_iters=bootstrap_iters).reset_ind
ex()
    #merged_res = merged_res.set_index('NoOverlap')


    return combined
```

In [ ]:

```
def plot_single(result, counts, title):
    fig, ax = plt.subplots()
    ax.axhline(1, color='black', lw=1, ls='--')
    x = result['Age']
    y = result['R']
    errUp = result['UpLim'] - y
    errDown = y - result['DownLim']
    ax.scatter(x, y, c='black', marker='o')
    ax.errorbar(x, y, yerr=[errDown, errUp], capsize=5, fmt='ko', linestyle='')
```

```
    ax.set_ylim(.5, 2)
    """results[i].plot(x='Age', y='R', yerr='DeltaR', capsize=5, fmt='ko',
              ylim = (.5, 2), title=titles[i],
               ax = ax, legend=False)"""

    ax.text(55, 1.4, f"{sum(counts): ,} Fatalities", fontsize=12)
    plt.ylabel("R = Relative fatality risk\n for females vs. males")
    plt.xlabel("Age")
    plt.show()
    return fig
```

In [ ]:

```
def run_double_pair(test, data, base_filter):

    filtered = data.query(base_filter)
    subset = filtered.query(test['data']).reset_index()
    subject = test['subject']
    control = test['control']

    control_bins = pd.cut(x=subset[control['age']], bins=control['bin'])
    subject_bins = pd.cut(x=subset[subject['age']], bins=subject['bin'])

    subset['AGE_C_Bin'] = control_bins
    subset['AGE_S_Bin'] = subject_bins

    grouped = subset.groupby(['AGE_C_Bin', 'AGE_S_Bin', control['sex']])

    soln = []

    for name, group in grouped:

        a = group.loc[(group[subject['sex']] == 'Female')
                      & (group[subject['dead']])]
        b = group.loc[(group[subject['sex']] == 'Female')
                      & (group[control['dead']])]
        c = group.loc[(group[subject['sex']] == 'Male')
                      & (group[subject['dead']])]
        d = group.loc[(group[subject['sex']] == 'Male')
                      & (group[control['dead']])]

        A = a[subject['id']].nunique()
        B = b[control['id']].nunique()
        C = c[subject['id']].nunique()
        D = d[control['id']].nunique()

        if (A > 0) and (B>0) and (C > 0) and (D > 0):
            r1 = A/B
            r2 = C/D
            R = r1/r2
            serr = np.sqrt((1/A) + (1/B) + (1/C) + (1/D))
```

```
            #serr = standard_error(R, A, B, C, D)
            soln.append([name[0], name[1], name[2], A, B, C, D, r1, r2, R, serr])

    result = pd.DataFrame(soln, columns = ['Control Age', 'Subject Age', 'Control S
ex', 'A', 'B', 'C', 'D', 'r1', 'r2', 'R', 'DeltaR'])
    weighted_result = result.groupby(['Subject Age']).apply(getWeightedAvg).reset_i
ndex()

    weighted_result['Age'] = weighted_result['Subject Age'].apply(lambda x: x.mid
+ .5)
    this_count = result['A'].sum() + result['C'].sum()
    return weighted_result.copy(), result.copy(), test['title'], this_count
```

In [ ]:

```
def double_pair(test, data, base_filter, bootstrap_iters):
    filtered = data.query(base_filter)
    subset = filtered.query(test['data']).reset_index()
    subject = test['subject']
    control = test['control']

    control_bins = pd.cut(x=subset[control['age']], bins=control['bin'])
    subject_bins = pd.cut(x=subset[subject['age']], bins=subject['bin'])

    subset['AGE_C_Bin'] = control_bins
    subset['AGE_S_Bin'] = subject_bins

    grouped = subset.groupby(['AGE_C_Bin', 'AGE_S_Bin', control['sex']])

    soln = []

    for name, group in grouped:
        # print(control['dead'])
        a = group.loc[(group[subject['sex']] == 'Female')
                      & (group[subject['dead']])
                      & (group[control['dead']])]
        b = group.loc[(group[subject['sex']] == 'Female')
                      & (group[subject['dead']])
                      & (~group[control['dead']])]
        c = group.loc[(group[subject['sex']] == 'Female')
                      & (~group[subject['dead']])
                      & (group[control['dead']])]

        e = group.loc[(group[subject['sex']] == 'Male')
                      & (group[subject['dead']])
                      & (group[control['dead']])]
        f = group.loc[(group[subject['sex']] == 'Male')
                      & (group[subject['dead']])
                      & (~group[control['dead']])]
        g = group.loc[(group[subject['sex']] == 'Male')
                      & (~group[subject['dead']])
                      & (group[control['dead']])]
```

```
        # print(a.head())
        A = a[[subject['id'], control['id']]].drop_duplicates().shape[0]
        B = b[subject['id']].nunique()
        C = c[control['id']].nunique()
        E = e[[subject['id'], control['id']]].drop_duplicates().shape[0]
        F = f[subject['id']].nunique()
        G = g[control['id']].nunique()

        if (((A + B) > 0) and ((A + C) > 0) and
                ((E + F) > 0) and ((E + G) > 0)):
            r1 = (A + B) / (A + C)
            r2 = (E + F) / (E + G)
            R = r1 / r2

            varlR = ((((A * (A + B + C) + (B * C)) * (F + G)) +
                       ((E * (E + F + G) + (F * G)) * (B + C))) /
                      ((A + B) * (A + C) * (E + F) * (E + G)))

            soln.append([name[0], name[1], name[2], A + B, A + C, E + F, E + G, r1,
r2, R, varlR])

    result = pd.DataFrame(soln, columns = ['Control Age', 'Subject Age', 'Control S
ex', 'A', 'B', 'C', 'D', 'r1', 'r2', 'R', 'DeltaR'])
    weighted_result = result.groupby(['Subject Age']).apply(bootstrap, bootstrap_it
ers=bootstrap_iters).reset_index()

    weighted_result['Age'] = weighted_result['Subject Age'].apply(lambda x: x.mid
+ .5)
    this_count = result['A'].sum() + result['C'].sum()
    return weighted_result.copy(), result.copy(), test['title'], this_count
```

In [ ]:

```
def bootstrap(group, bootstrap_iters, includeTotal = False, includeCounts=True):
    n = len(group.index)
    bs_stats = []
    bs_deltaR = []
    for i in range(1, bootstrap_iters):
        working = resample(group, replace=True, n_samples=n)
        #weight_avg = np.exp(np.sum(np.log(working['R'])/working['DeltaR'])/np.sum
(1/working['DeltaR']))
        weight_avg = np.exp(np.divide(np.sum(np.divide(np.log(working['R']), workin
g['DeltaR'], out=np.zeros_like(working['R']), where=working['DeltaR']!=0)), np.sum
(np.divide(1, working['DeltaR'], out=np.zeros_like(working['DeltaR']), where=workin
g['DeltaR']!=0))))
        delta_R_bar = 1/np.sum(np.divide(1, working['DeltaR'], out=np.zeros_like(wo
rking['DeltaR']), where=working['DeltaR']!=0))
        bs_stats.append(weight_avg)
        bs_deltaR.append(weight_avg)
    working = group
```

```
    #weight_avg = np.exp(np.sum(np.log(working['R'])/working['DeltaR'])/np.sum(1/wo
rking['DeltaR']))
    weight_avg = np.exp(np.divide(np.sum(np.divide(np.log(working['R']), working['D
eltaR'], out=np.zeros_like(working['R']), where=working['DeltaR']!=0)), np.sum(np.d
ivide(1, working['DeltaR'], out=np.zeros_like(working['DeltaR']), where=working['De
ltaR']!=0))))
    #delta_R_bar = 1/(np.sum(1/working['DeltaR']))
    #delta_R_bar = 1/np.sum(np.divide(1, working['DeltaR'], out=np.zeros_like(worki
ng['DeltaR']), where=working['DeltaR']!=0))
    delta_R_bar = np.median(bs_deltaR)
    lowlim = np.percentile(bs_stats, 2.5)
    uplim = np.percentile(bs_stats, 97.5)

    if includeTotal:
        ttl = (working['A'] + working['B'] + working['C'] + working['D'])
        return pd.Series({'R': R_bar, 'DeltaR': delta_R_bar, 'Total': ttl})
    elif includeCounts:
        return pd.Series({
            'A': working['A'].sum(),
            'B': working['B'].sum(),
            'C': working['C'].sum(),
            'D': working['D'].sum(),
            'R': weight_avg,
            'DeltaR': delta_R_bar,
            'UpLim': uplim,
            'DownLim': lowlim,
            'BS': bs_stats})
    else:
        return pd.Series({'R': R_bar, 'DeltaR': delta_R_bar})
```

## E. Full Analysis

For each vehicle type, run the double pair comparison on the specified subsets, and append this information for plotting and reporting.

In [ ]:

```
vehicle_results = {}
all_results = []
total_count = []
for veh_type in vehicle_types:
    results = []
    titles = []
    counts = []
    all_res = []
    if veh_type == 'MOTORCYCLE':
        cur_subset = mcycl_subsets
    else:
        cur_subset = subsets
    for test in cur_subset:
        wr, res, title, count = double_pair(test,
                                            data.query(f"{veh_type}"),
                                            base_filter,
```

```
                                                        bootstrap_iters)
        results.append(wr)
        all_res.append(res)
        titles.append(title)
        counts.append(count)

    vehicle_results[veh_type] = {'results': results,
                                 'titles': titles,
                                 'counts': counts,
                                 'all_res': all_res}
```

## F. Results: Passenger Cars

In [ ]:

```
if 'PASSENGER_CAR' in vehicle_types:
    if not driver_only:
        fig = plot_doublepair(vehicle_results['PASSENGER_CAR'])
        fig.savefig(fig_out / "passenger_car_subsets.png")
    else:
        fig = plot_driver_only(vehicle_results['PASSENGER_CAR'])
        fig.savefig(fig_out / "passenger_car_subsets.png")
```

In [ ]:

```
if 'PASSENGER_CAR' in vehicle_types:
    cur_results = vehicle_results['PASSENGER_CAR']
    car_result = combine_results(cur_results['results'], bootstrap_iters)
    car_counts = cur_results['counts']
    fig = plot_single(car_result, car_counts, "Car Fatalities")
    fig.savefig(fig_out / "passenger_cars.png")
    all_results.append(car_result)
    total_count.append(sum(car_counts))
```

## G. Results: Light Truck

In [ ]:

```
if 'LIGHT_TRUCK_OR_VAN' in vehicle_types:
    fig = plot_doublepair(vehicle_results['LIGHT_TRUCK_OR_VAN'])
    fig.savefig(fig_out / "truck_subsets.png")
```

In [ ]:

```
if 'LIGHT_TRUCK_OR_VAN' in vehicle_types:
    cur_results = vehicle_results['LIGHT_TRUCK_OR_VAN']
    truck_result = combine_results(cur_results['results'], bootstrap_iters)
    truck_counts = cur_results['counts']
    fig = plot_single(truck_result, truck_counts, "Light-Truck Fatalities")
    fig.savefig(fig_out / "light_trucks.png")
    all_results.append(truck_result)
    total_count.append(sum(truck_counts))
```

## H. Results: Motorcycles

In [ ]:

```
def mcycl_plotter(results, counts, titles):
    fig, axes = plt.subplots(1,2, figsize=(10,3), sharex=True, sharey=True)

    a=[]
```

```
    for i, ax in enumerate(axes):
        a.append(ax)

    for i, ax in enumerate(a):
        ax.axhline(1, color='black', lw=1, ls='--')
        results[i].plot(x='Age', y='R', yerr='DeltaR', capsize=5, fmt='ko',
                        ylim = (.5, 2), title=titles[i],
                         ax = ax, legend=False)
        ax.text(55, 1.85, f"{counts[i]: ,} Fatalities", fontsize=12)

        ax.xaxis.set_major_locator(MultipleLocator(10))
        ax.xaxis.set_minor_locator(MultipleLocator(5))
        ax.tick_params(which='major', length=7)
        ax.tick_params(which='minor', length=4)
        #axis.yaxis.set_major_locator(MultipleLocator(10))
        ax.yaxis.set_minor_locator(MultipleLocator(.1))
    plt.ylabel("R = Relative fatality risk for females vs. males")
    plt.show()
    return fig
```

In [ ]:

```
if 'MOTORCYCLE' in vehicle_types:
    fig = mcycl_plotter(vehicle_results['MOTORCYCLE']['results'],
                        vehicle_results['MOTORCYCLE']['counts'],
                        vehicle_results['MOTORCYCLE']['titles'])
    fig.savefig(fig_out / "motorcycle_subsets.png")
```

In [ ]:

```
if 'MOTORCYCLE' in vehicle_types:
    mcycl_result = combine_results(vehicle_results['MOTORCYCLE']['results'], bootst
rap_iters)
    mcycl_counts = vehicle_results['MOTORCYCLE']['counts']
    fig = plot_single(mcycl_result, mcycl_counts, "Motorcycle Fatalities")
    fig.savefig(fig_out / "motorcycles.png")
    all_results.append(mcycl_result)
    total_count.append(sum(mcycl_counts))
```

## *I. All Occupants*

In [ ]:

```
final_result = pd.concat(all_results)
final_result['Age'] = pd.cut(x=final_result['Age'], bins=[x for x in range(0, 101,
5)]).apply(lambda x: x.mid)
final_result = final_result.groupby('Age')
final_result = final_result.apply(bootstrap, bootstrap_iters=bootstrap_iters).reset
_index()
fig = plot_single(final_result, total_count, "All Fatalities")
fig.savefig(fig_out / "all_fatalities.png")
```

## *J. Tables*

Tables produced for review, also output as .xlsx (NOT IMPLEMENTED) and .csv files for easy transfer to papers.

In [ ]:

```
if final_runs:
```

```
    csv_out = Path(module_path) / "reports" / "final" / "tables" / image_outputs / 
"csv"
    xlsx_out = Path(module_path) / "reports" / "final" / "tables" / image_outputs / 
"xlsx"
else:
    csv_out = Path(module_path) / "reports" / "tables" / image_outputs / "csv"
    xlsx_out = Path(module_path) / "reports" / "tables" / image_outputs / "xlsx"

csv_out.mkdir(parents=True, exist_ok=True)
xlsx_out.mkdir(parents=True, exist_ok=True)
```

In [ ]:

```
for veh_type, cur_veh in vehicle_results.items():

    veh_csv = csv_out / veh_type
    veh_xlsx = xlsx_out / veh_type

    veh_csv.mkdir(parents=True, exist_ok=True)
    car_result.to_csv(veh_csv / "full_weighted.csv")

    for title, res, all_res in zip(cur_veh['titles'],
                                   cur_veh['results'],
                                   cur_veh['all_res']):
        cur_csv = veh_csv / title
        cur_xlsx = veh_xlsx / title
        cur_csv.mkdir(parents=True, exist_ok=True)
        cur_xlsx.mkdir(parents=True, exist_ok=True)

        #Write out table of weighted results
        res.to_csv(cur_csv / 'weighted_results.csv', float_format='%.3f')
        #res.to_excel(cur_xlsx / 'weighted_results.csv')

        #Write out table of all results
        all_res.to_csv(cur_csv / 'full_results.csv', float_format='%.3f')

        #Loop through each subject interval in all_res, write out this table
        for subj_age in all_res['Subject Age'].unique():
            cur_chunk = all_res.loc[all_res['Subject Age'] ==
                                    subj_age].sort_values('Control Sex', ascending=
False)
            cur_chunk.to_csv(cur_csv / f"subjects_{subj_age}.csv", float_format='%.
3f')
```

In [ ]:

```
if 'PASSENGER_CAR' in vehicle_types:
    car_result.to_csv(csv_out / "full_car_weighted.csv")
if 'LIGHT_TRUCK_OR_VAN' in vehicle_types:
    truck_result.to_csv(csv_out / "full_truck_weighted.csv")
if 'MOTORCYCLE' in vehicle_types:
    mcycl_result.to_csv(csv_out / "full_motorcycle_weighted.csv")
final_result.to_csv(csv_out / "full_weighted_result.csv")
```

In [ ]:

```
summary = []
if 'PASSENGER_CAR' in vehicle_types:
    car = vehicle_results['PASSENGER_CAR']
    car['Female Fatalities'] = [res['A'].sum() for res in car['all_res']]
    car['Male Fatalities'] = [res['C'].sum() for res in car['all_res']]
    summary.append(pd.DataFrame.from_dict(car).drop(columns=
                                                    ['results',
                                                     'all_res',
                                                    ]).assign(Vehicle='Car'))
if 'LIGHT_TRUCK_OR_VAN' in vehicle_types:
    truck = vehicle_results['LIGHT_TRUCK_OR_VAN']
    truck['Female Fatalities'] = [res['A'].sum() for res in truck['all_res']]
    truck['Male Fatalities'] = [res['C'].sum() for res in truck['all_res']]
    summary.append(pd.DataFrame.from_dict(truck).drop(columns=
                                                      ['results',
                                                       'all_res',
                                                       ]).assign(Vehicle='Light tru
ck'))
if 'MOTORCYCLE' in vehicle_types:
    mcycl = vehicle_results['MOTORCYCLE']
    mcycl['Female Fatalities'] = [res['A'].sum() for res in mcycl['all_res']]
    mcycl['Male Fatalities'] = [res['C'].sum() for res in mcycl['all_res']]
    summary.append(pd.DataFrame.from_dict(mcycl).drop(columns=
                                                      ['results',
                                                       'all_res',
                                                       ]).assign(Vehicle='Motorcycl
e'))
summaries = pd.concat(summary).rename(columns={
    'titles': 'Subject Occupant',
    'counts': 'Total'}).reindex(columns=['Vehicle',
                                         'Subject Occupant',
                                         'Female Fatalities',
                                         'Male Fatalities',
                                         'Total'])
summaries.to_csv(csv_out / "summary.csv")
```

In [ ]:

```
def do_violin(cur_data):

    fig, ax = plt.subplots()
    ax.axhline(1, color='black', lw=1, ls='--')
    ax.violinplot(cur_data.BS,
                  positions=cur_data.Age,
                  points=500, vert=True, widths=4,
                  showextrema=True,
                  bw_method=0.5)
    ax.scatter(cur_data.Age, cur_data.R)
    ax.scatter(cur_data.Age, cur_data.UpLim, marker=10)
    ax.scatter(cur_data.Age, cur_data.DownLim, marker=11)
    #ax.xaxis.set_major_locator(MultipleLocator(10))
    #ax.xaxis.set_minor_locator(MultipleLocator(5))
```

```
    #ax.tick_params(which='major', length=7)
    #ax.tick_params(which='minor', length=4)
    ax.set_ylim(.5, 2)

    return fig
```

In [ ]:

```
def do_box(cur_data):

    fig, ax = plt.subplots()
    ax.axhline(1, color='black', lw=1, ls='--')
    ax.boxplot(cur_data.BS,
               positions=cur_data.Age, widths=4,
               whis=(5, 95))
    ax.scatter(cur_data.Age, cur_data.R)
    ax.scatter(cur_data.Age, cur_data.UpLim, marker=10)
    ax.scatter(cur_data.Age, cur_data.DownLim, marker=11)
    ax.xaxis.set_major_locator(MultipleLocator(10))
    ax.xaxis.set_minor_locator(MultipleLocator(5))
    ax.tick_params(which='major', length=7)
    ax.tick_params(which='minor', length=4)
    ax.set_ylim(.5, 2)

    return(fig)
```

In [ ]:

```
def do_stripplot(cur_data, msize=2, malpha=.2, jitter=.5, tick_mods=False):

    fig, ax = plt.subplots()
    work_data = cur_data.explode("BS")
    work_data['Age'] = work_data["Age"] + jitter * np.random.rand(len(work_data["Ag
e"])) - jitter/2

    #sns.stripplot(x="Age", y="BS",data=work_data, ax=ax, alpha=0.25, size=2, color
="k")
    ax.scatter(work_data.Age, work_data.BS, s=msize, alpha=malpha, color="k")
    ax.axhline(1, color='black', lw=1, ls='--')
    if tick_mods:
        ax.xaxis.set_major_locator(MultipleLocator(10))
        ax.xaxis.set_minor_locator(MultipleLocator(5))
        ax.tick_params(which='major', length=7)
        ax.tick_params(which='minor', length=4)
    plt.ylabel("R = Relative fatality risk\n for females vs. males")
    plt.xlabel("Age")

    ax.set_ylim(.5, 2)

    return(fig)
```

In [ ]:

```
jitters = [.5, 1, 1.5]
tick_mods = [True, False]
msizes = [1.75]
```

```
malphas = [.18]


if 'PASSENGER_CAR' in vehicle_types:

    cur_title = str(image_outputs).split(' ', 1)[1]

    cur_title = cur_title.replace("_", "\n")
    fname = cur_title.replace("\n", " ")
    cur_title = cur_title + " Bootstrap Iterations"

    cur_data = car_result#pd.read_csv(cur_dir / "csv" / "full_car_weighted.csv")
    #cur_data.BS = cur_data.BS.apply(fix_arr)

    #fig, ax = plt.subplots()
    #do_violin(cur_data, fig, ax)
    fig = do_violin(cur_data)
    plt.title(cur_title + " Car Results")

    #sns.violinplot(x="Age", y="BS", data=cur_data)
    plt.show()
    #print(cur_data.head())
    fig.savefig(fig_out / f"{fname}, Full Car Violin.png")
    plt.close()

    fig = do_box(cur_data)
    plt.title(cur_title + " Car Results")
    fig.savefig(fig_out / f"{fname}, Full Car Box.png")
    plt.close()

    for jitter in jitters:
        for tick_mod in tick_mods:
            for msize in msizes:
                for malpha in malphas:
                    fig = do_stripplot(cur_data,
                                       tick_mods=tick_mod,
                                       msize=msize,
                                       malpha=malpha,
                                       jitter=jitter)
                    plt.title(cur_title + " Car Results")
                    fig.savefig(fig_out / f"{fname}, Full Car j{jitter} t{tick_mo
d}.png")
                    plt.close()

if 'LIGHT_TRUCK_OR_VAN' in vehicle_types:

    cur_title = str(image_outputs).split(' ', 1)[1]

    cur_title = cur_title.replace("_", "\n")
    fname = cur_title.replace("\n", " ")
```

```
    cur_title = cur_title + " Bootstrap Iterations"

    cur_data = truck_result#pd.read_csv(cur_dir / "csv" / "full_car_weighted.csv")
    #cur_data.BS = cur_data.BS.apply(fix_arr)

    #fig, ax = plt.subplots()
    #do_violin(cur_data, fig, ax)
    fig = do_violin(cur_data)
    plt.title(cur_title + " Truck Results")

    #sns.violinplot(x="Age", y="BS", data=cur_data)
    plt.show()
    #print(cur_data.head())
    fig.savefig(fig_out / f"{fname}, Full Truck Violin.png")
    plt.close()

    fig = do_box(cur_data)
    plt.title(cur_title + " Truck Results")
    fig.savefig(fig_out / f"{fname}, Full Truck Box.png")
    plt.close()

    for jitter in jitters:
        for tick_mod in tick_mods:
            for msize in msizes:
                for malpha in malphas:
                    fig = do_stripplot(cur_data,
                                       tick_mods=tick_mod,
                                       msize=msize,
                                       malpha=malpha,
                                       jitter=jitter)
                    plt.title(cur_title + " Truck Results")
                    fig.savefig(fig_out / f"{fname}, Full Truck j{jitter} t{tick_mo
d}.png")
                    plt.close()


if 'MOTORCYCLE' in vehicle_types:

    cur_title = str(image_outputs).split(' ', 1)[1]

    cur_title = cur_title.replace("_", "\n")
    fname = cur_title.replace("\n", " ")
    cur_title = cur_title + " Bootstrap Iterations"

    cur_data = mcycl_result

    fig = do_violin(cur_data)
    plt.title(cur_title + " Motorcycle Results")

    #sns.violinplot(x="Age", y="BS", data=cur_data)
```

```
    plt.show()
    #print(cur_data.head())
    fig.savefig(fig_out / f"{fname}, Full Motorcycle Violin.png")
    plt.close()

    fig = do_box(cur_data)
    plt.title(cur_title + " Motorcycle Results")
    fig.savefig(fig_out / f"{fname}, Full Motorcycle Box.png")
    plt.close()

    for jitter in jitters:
        for tick_mod in tick_mods:
            for msize in msizes:
                for malpha in malphas:
                    fig = do_stripplot(cur_data,
                                       tick_mods=tick_mod,
                                       msize=msize,
                                       malpha=malpha,
                                       jitter=jitter)
                    plt.title(cur_title + " Motorcycle Results")
                    fig.savefig(fig_out / f"{fname}, Full Motorcycle j{jitter} t{ti
ck_mod}.png")
                    plt.close()

cur_data = final_result

fig = do_violin(cur_data)
plt.title(cur_title + " All Results")
fig.savefig(fig_out / f"{fname}, All Results Violin.png")
plt.close()

fig = do_box(cur_data)
plt.title(cur_title + " All Results")
fig.savefig(fig_out / f"{fname}, All Results Box.png")
plt.close()

for jitter in jitters:
    for tick_mod in tick_mods:
        for msize in msizes:
            for malpha in malphas:
                fig = do_stripplot(cur_data,
                                   tick_mods=tick_mod,
                                   msize=msize,
                                   malpha=malpha,
                                   jitter=jitter)
                plt.title(cur_title + " All Results")
                fig.savefig(fig_out / f"{fname}, All Results j{jitter} t{tick_mod}.
png")
                plt.close()
```

### *Clean up*

Due to a bug in the `nbclient` library, as of 3/31/2020, need to clean up large data from memory, or it will persist after the run.

In [ ]:

```
del(data)
```

In [ ]:

```
%reset -f
```