Accident Analysis and Prevention 96 (2016) 88–100

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

How rare are large, multiple-fatality work-related incidents?

Brooks Pierce

U.S. Bureau of Labor Statistics, Washington DC, United States

ARTICLE INFO

Article history:
Received 20 May 2015
Received in revised form 13 April 2016
Accepted 10 July 2016

Keywords:
Multiple-fatality incident
Work-related fatality
Census of fatal occupational injuries
Size distribution of fatal incidents

ABSTRACT

Despite their salience, the prevalence of incidents that result in many work-related deaths is not well-documented. This study estimated probabilities of observing large scale work-related fatal incidents using 1995–2010 records from the Census of Fatal Occupational Injuries. A range of model estimates suggest approximately a one-in-four annual chance of observing an incident resulting in 20 or more work-related fatalities. The most likely contributors are aircraft incidents, and fires and explosions. The probability that a large scale incident occurs has declined in recent years due to a general decline in the number of fatal incidents, and due to a compositional shift away from those types of incidents more likely to result in large scale outcomes.

Published by Elsevier Ltd.

1. Introduction

Thousands of U.S. workers die annually in work-related incidents. The great majority of these incidents involve a single worker, but a non-negligible fraction of incidents result in multiple fatalities. Rarely, a dozen or more individuals will perish in a single incident. Such incidents are often portrayed as exceptional, especially relative to the historical record for similar events. For example, on June 30, 2013, a wildfire near Yarnell, Arizona overran the position of a brigade of firefighters, causing 19 deaths (Santos, 2013). The Yarnell blaze was noted to have caused the greatest number of wilderness firefighter casualties in a single incident in 80 years (Dolan, 2013). Unfortunately, there are other examples, such as the 2010 Upper Big Branch mine explosion that killed 29 mine workers, the worst coal mining incident in 40 years (Gabriel, 2014).

Severity as measured by the number of fatalities is an important incident attribute; as such it is typically recorded as part of the public health documentation of incidents, and is a point of emphasis for prevention and investigation priorities (Drudi and Zak, 2004; Iannacchione et al., 2008; U.S. Department of Health and Human Services, 2009). Studies of affected populations or sectors note the importance of multiple-decedent incidents, and how such incidents can influence summary statistics, for example by inducing greater year-to-year variability in measured fatality rates, and by making incident-based fatality rates a useful auxiliary measure of situational risk (Estes et al., 2011; Pegula, 2004; Rice and Janocha, 2008; Tao et al., 2011). Furthermore, because multi-decedent incidents are extraordinarily tragic they can affect public and legislative focus on safety (Rice and Janocha, 2008; Levine, 2008).

Despite this, there is little systematic documentation of the prevalence in the U.S. of such extreme events.[1] Are they anomalies? Have the apparent improvements in workplace and transport safety extended to these events as well? And, based on the historical record, what is the likelihood that a large magnitude event will occur sometime in the near future? This paper analyzed work-related fatalities over the 1995–2010 period to document the prevalence of multi-decedent fatal incidents and to ascertain the likelihood of very severe incidents.

2. Materials and methods

2.1. Data

This study used records on fatal work-related incidents occurring within the U.S., as enumerated in the Bureau of Labor Statistics' Census of Fatal Occupational Injuries (U.S. Department of Labor, undated). This census has broad scope, for example including work-related fatalities among volunteer workers, the self-employed, and resident military. Information on each fatality is based on multiple source documents identifying characteristics of the incident and decedent. This census identified the individual fatalities occurring in a common incident, beginning in 1993. However, incident link-

[1] Important exceptions are Biddle and Hartley (2002) and Drudi and Zak (2004). CFOI data analyses accompanying annual press releases typically note numbers of incidents and fatalities by event group.

E-mail address: pierce.brooks@bls.gov

http://dx.doi.org/10.1016/j.aap.2016.07.014
0001-4575/Published by Elsevier Ltd.

age codes in the early years of the data are subject to linkage errors, and so I follow Drudi and Zak (2004) in taking 1995 as the start of the analysis window, and in reviewing all linkages for coding errors. Individual incidents were determined by cross-referencing incident dates and locations across fatalities, and further by a manual review of each fatality's narrative; an Appendix further describes linkage definitions. I analyzed all incidents between 1995 and 2010, excluding the incidents of the September 11, 2001 terrorist attacks. The September 11 incidents were true outliers involving thousands of workers and are typically excluded from analyses of work-related fatalities (U.S. Department of Labor, 2002).

The likelihood that multiple fatalities arise from any single incident depends on the circumstances surrounding the incident and accordingly this study conditioned on case characteristics. CFOI data are coded with several classifying fields describing case circumstances. These include the event or exposure leading to injury, the incident location, the nature of the injury occurring to the worker, the part of body affected, and the source of injury (meaning the object or substance directly producing the injury) (U.S. Department of Labor, 1992). The event or exposure classifies what happened to cause injury and is especially relevant for documenting large scale incidents. Event frequencies among non-fatal work-related injuries and illnesses from the Survey of Occupational Injuries and Illnesses (SOII) were also tabulated and reported. The coding structure for event and other case characteristics changed in 2011, so for most calculations I restricted analysis to the 1995–2010 period.

2.2. Statistical analysis

My primary goal was to estimate the probability of observing a fatal incident involving many decedents during a representative year. I derived these estimates in three steps. First, initial descriptive analyses examined the overall prevalence of multi-decedent incidents, and identified case characteristics associated with the most severe incidents. Based on this analysis I stratified estimation, with stratifying groups defined by certain combinations of case characteristics. Second, I fit probability distributions for incident severity for each stratifying group. The parameters of the fitted models were used to estimate the probability that any given incident would involve 20 or more decedents. Third, I applied these probabilities to the number of incidents expected to occur within a representative year, to obtain estimates for the probability of observing an extreme event over the course of a year. Trend regression analysis identified differential trends in incident counts for the separate stratifying groups, and provided estimates for expected incident counts for 2016.

2.2.1. Distributional assumptions and tests

Estimating the probability of an extremely large incident is complicated by the fact that the empirical distribution is highly variable in the area of interest: because extreme incidents are rare, a realized empirical distribution can give an inaccurate measure of the true probabilities. The typical solution to this difficulty involves fitting distributional models on the realized empirical distributions in question. The distributional model, if appropriate, effectively brings in information that is used to generate better (less variable) estimates. Although it may seem coarse to treat human catastrophes as a proper subject for statistical modeling, doing so is helpful as an intermediate step toward understanding their true prevalence. This approach has been used to model extreme outcomes in a variety of settings, including wages or earnings (Armour et al., 2014), severity of conflicts (Friedman, 2015), severity of terrorist attacks (Clauset and Woodard, 2013), financial returns (Gabaix et al., 2003), and earthquake magnitude (Gutenberg and Richter, 1944). Newman (2005) and Clauset et al. (2009) give useful insight into these methods.

To fix ideas, let x refer to the severity of a fatal incident as measured by the number of decedents. The probability of an incident of severity x for x at or above some minimum value $x_{\min}$ may be written as $\Pr(x|\theta, x_{\min})$, where the θ are distributional parameters to be estimated. I fit these probabilities by maximum likelihood, conditional on the distributional form for $\Pr(x|\theta, x_{\min})$ and the cutoff value $x_{\min}$ above which the distribution is assumed to hold.

Testing indicated that $\Pr(x|\theta, x_{\min})$ was reasonably described with a discrete Pareto, or power-law, probability density,

$$\Pr(x|\theta, x_{\min}) = Cx^{-\theta} \quad (1)$$

for x taking on integer values greater than or equal to $x_{\min}$, and where C is a constant of integration. This functional form is a standard benchmark for problems involving heavy-tailed phenomena. It implies a particular relationship between relative probabilities at different severity levels, embodied in the single parameter θ. The distributional model uses relative frequencies throughout the entire range of x above $x_{\min}$ to estimate probabilities relevant to an extreme subset of outcomes. If the model is passably correct then this incorporates extra information, because we typically have much more data available at lower severity levels and the modeling uses that data to form estimates of extreme outcome probabilities. For example, in the current application we have very good data on relative frequencies of incidents involving 2–4, and even 7 or 8 decedents, and we can usefully combine that information with relative frequencies of incidents involving 20, 21, or more decedents, to obtain a sensible estimate for θ. That parameter can then be used to estimate probabilities of very large incidents, which are rare enough so that simple frequency-based estimates are problematic. Alternative distributions to (1) would also impose some common functional form to the data above $x_{\min}$ and thereby potentially provide less variable estimates for extreme probabilities.

I fit models of Eq. (1) in subsamples defined by the circumstances under which the fatalities occurred. This allowed me to distinguish between the severity of incidents in, say, aircraft incidents and fires. That is, letting subscripts j denote incident subgroups, I fit the probabilities

$$\Pr(x|\theta_j, x_{min,j}) = C_j x^{-\theta_j} \quad (2)$$

where $x_{\min,j}$ is a group-specific cutoff value and θ_j is a group-specific distributional parameter.

I used various diagnostic checks to determine if the particular distributional form in Eq. (1) was sensible. Researchers in other contexts have noted a sensitivity of results to choices for the cutoff $x_{\min}$ and the form assumed for $\Pr(x|\theta, x_{\min})$. I restricted values of $x_{\min}$ to be 2 or greater, so that estimation occurred only on the set of multi-fatality incidents. To gauge robustness I fit data for different $x_{\min}$ values and alternative functional forms for $\Pr(x|\theta, x_{\min})$. I used evaluation tests suggested by Clauset et al. (2009) and Clauset and Woodard (2013), which are designed to gauge whether the Pareto distribution reasonably describes the empirical distribution, and further, whether the Pareto distribution provides a superior fit than particular alternative distributions such as the Poisson or exponential distributions. Appendix B gives details on estimation and testing.

2.2.2. The probability an incident is extremely severe

Values for θ and $x_{\min}$ in Equation (1) completely describe estimated probabilities that an incident involves x decedents, conditional on $x \geq x_{\min}$. The probabilities in turn imply a fitted cumulative distribution $F(x|\hat{\theta}, \hat{x}_{min})$ defined over $x \geq x_{\min}$, and an estimated probability that an incident involves more than x decedents, conditional on the incident involving at least $x_{\min}$ decedents,