Trends in U.S. Chemical Industry Accidents

Michael R. Elliott, Ph.D., Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine
Paul Kleindorfer, Ph.D., The Wharton School, University of Pennsylvania
Yanlin Wang, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine
Isadore Rosenthal, Ph.D., The Wharton School, University of Pennsylvania

For Correspondence:
Paul R. Kleindorfer
Risk Management and Decision Processes Center
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6366
(215) 898-5830 (phone)
(215) 573-2130 (fax)
kleindorfer@wharton.upenn.edu

September 9, 2004

ABSTRACT

This paper presents the results of an analysis of the accident history data reported under Section 112(r) of the Clean Air Act Amendments. This data provides a fairly complete record of the consequences of reportable accidental releases occurring during the time frame 1995-1999 in the U. S. chemical industry and covering 77 toxic and 63 flammable substances subject to the provisions of 112(r). Our analysis shows that the frequencies of reported accidents climbed from approximately one per 67 facilities in 1994 to one in 35 facilities in 1997 and 1998, then declined again to approximately one per sixty facilities in 2000. While this change was highly statistically significant ($p<.0001$), some or all of this trend may be due to spurious data collection effects. No statistically significant differences in injury rates over time were observed. Future rounds of RMP*Info reporting may allow trend analyses under weaker assumptions than required here.

KEY WORDS: RMP*Info, Risk Management Plans, Hazardous Chemicals, Clean Air Act

Introduction

Section 112(r) of the 1990 Clean Air Act Amendments set forth a series of requirements aimed at preventing and minimizing the consequences associated with accidental releases of chemicals at U.S. manufacturing facilities. Its implementation in EPA regulation, 40 CFR 68, required all facilities storing on-site at any of 77 toxic or 63 flammable substances above a threshold quantity (ranging from 250 to 20,000 lbs)) to develop a risk management program (RMP)[1]. These RMPs include assessments of hazards, details on accident histories during the past 5 years, worst-case accident release scenarios, and prevention and emergency response programs. A total of 15,219 unique facilities reported to this database. All facilities were to report accidental releases of covered chemicals or processes that resulted in deaths, injuries, significant property damage, evacuations, sheltering in place, or environmental damage. For this analysis we disregard much the extensive detail in this database to focus on overall accident rates over time among reporting facilities. We consider trends in overall accidents, injuries, and property damage during the five preceding years for each facility.

The regulation implementing 112(r) set a preliminary deadline of June 21, 1999 and a final deadline of June 21, 2001, for the filing of the RMP data, which then covered the preceding five-year period. A total of 15,430 facilities filed by the final deadline. As there were some special reasons for filings beyond the calendar year 2000 time frame (see Kleindorfer et al. (2003) for details), we will be using the snapshot of the RMP data taken on December 11, 2000, comprising 15, 219 facilities. Since filing under 112(r) is mandated by law, and carries significant penalties for non-compliance, the resulting data might be viewed as a reasonably complete and exhaustive picture of the hazards represented by the US chemical industry during 1994-2000.
For trend analysis, however, there is a significant failing in the available data: not all facilities may have been in operation of the five-year reporting period, or if in operation, may not have been using chemicals above threshold levels; and no data are available to determine the length of (reporting-relevant) operation. Since facilities that were in operation earlier in the five-year period but not at the time of filing are not observed, accident rates earlier in the reporting period are downwardly biased by an unknown degree. Consequently we report not only the observed trend data but consider a sensitivity analysis to determine how trends would differ assuming different "appearance rates" for facilities.

Methods

Sources of Data

The data underlying the analysis include the 15,219 facilities that had files by December 11, 2000. Five facilities are dropped from this analysis because that lacked "postmark

[1] There were certain exceptions: e.g., farmers using ammonia as an agricultural nutrient. See http://www.epa.gov/swercepp/pubs/potw/part6899.pdf for details on the chemicals regulated under 112(r).

dates" identifying the end of their five-year follow-up history. The information contained in RMP*Info™ database is extensive and includes details about on-site chemicals and processes; regulatory program coverage; geographic location; and number of full-time employees (FTE). The accident-related information includes date and time of accident; number of associated injuries or deaths among workers, public responders, or the public at large; and off-site consequences such as property damage (on-site, offsite), evacuations, confinement indoors, and environmental damage. For a more complete description of the RMP*Info database, see Kleindorfer et al. (2003).

We consider three main outcomes of interest: accidents, total injuries resulting from reported accidents (among workers, public responders, or the public at large), and total property damage (both on-site and off-site) resulting from reported accidents. Postmark dates were not the same for all 15,214 facilities, and ranged from January 3, 1999, to December 7, 2000. Consequently a complete history for the 2,135 day period ranging from January 3, 1994, though December 7, 2000 was constructed, with the total number of accidents, total number of injuries, and total property damage reported on each day over that period determined, along with the total number of facilities "at risk" each day based on the five-year follow-up period (see Figure 1).

More than 97% of the RMP*Info™ filings are submitted electronically, permitting consistency and range checks during the submission process. The data obtained were screened for accuracy and consistency by the research team via interviews with plant-level and corporate managers responsible for submitting RMP data and via examination for outliers and internal inconsistencies in the data (Kleindorfer et al. 2003). Managers generally exhibited a clear understanding of the RMP process and devoted considerable effort toward its completion, suggesting data quality was likely to be high. The data were also subject to detailed reviews by both EPA staff and by facilities before their release; a review of the data by the researchers revealed no major remaining outliers or inconsistencies.

Data Analysis

We computed the total number of accidents, injuries, and property damage for each of the seven calendar years from 1994 through 2000. To determine if the underlying rates of accidents differed over time, we used generalized linear regression models (McCullagh and Nelder 1989) to relate the risk of an accident or injury to the time period of observation. (Because of difficulties in modeling property damage statistically, we restricted our analysis to accidents and injuries.) We assume that the number of accidents reported on a given day *d* is a binomially-distributed random variable with n_d "trials" given by the number of facilities reporting on day *d*, and with probability of accident that is assumed to be constant in any given calendar year. We assume that the count of injuries on any given day follows a negative binomial distribution, with an underlying injury risk parameter that is assumed to be constant in any given calendar year. To account for the fact that these accident and injury counts are not independent, since they come from the same facilities are hence may be autocorrelated, generalized estimating equations (Liang and Zeger 1986) have been used to compute empirical covariance

matricies under a working independence assumption. These empirical covariance matricies have been used to construct confidence intervals for these annual risk parameters, and to compute Wald statistics to test the hypothesis that the annual risk parameters were the same throughout the seven-year period considered.

The preliminary analysis assumed that each facility was "at risk" throughout the entire five-year period preceding the postmark date. Because some facilities may have come "on-line" during this period and thus could not have contributed to the accident or injury total, we considered a second analysis that downweights the number of facilities "at risk" in the earlier years in the following fashion. Each at-risk facility on a given day was downweighted by $(1-p)^{x/365}$, where x is the number of days from the day to interest to the postmark date, and p is a sensitivity parameter that is the annual rate at which the facilities in the database became eligible to report accidents to RMP*Info; thus if p=.05, we assume that only 95% of the facilities were at risk one year before their postmark data, 90% two years before their postmark data, and so forth through 77% five years before their postmark data, at the beginning of the reporting period. We consider p=.03, .05, and .10.

Yet another issue is that facilities that reported early may have had different risk profiles than firms that reported later. To deal with this, we conduct a second analysis with was restricted to the dates (December 7, 1995-January 3, 1999) in which data were available for all 15,214 firms.

Results

Table 1 shows the annual number of reported accidents, injuries, and property damage for the January 3, 1994 – December 7, 2000 period considered. A total of 56 accidents for the 15,214 facilities considered were dropped, as they were reported outside of the 5-year filing "window" considered; a total of 1,902 reported accidents remained, resulting in 2,048 injuries and $984 million in property damage.

The peak year for accidents and injuries was 1997, with 439 accidents and 440 injuries reported. The peak year for property damage was the partial reporting year of 1994, where a single $219 million accident accounted for most of the $310 million total. The annual rate of facility accidents appeared to follow a U-shaped trend, climbing from .015 per facility (95% CI=.012,.018) in 1994 to .028 (95% CI=.026,.032) in 1997 and 1998, then declining to .016 (95% CI=.009,.029) in 2000 ($p<.0001$). The injury rate, however, followed no clear trend, ranging from .026 to .029 per facility per year from 1994 through 1998, climbing to .033 in 1999, then dropping to .022 in 2000 (p=.98). Recall that a rate of .01 incidents per facility per year represents approximately 150 facilities per year across all 15,219.

As previously discussed, however, some or all of this trend may be spurious due to the fact that we are treating all facilities as being present throughout the five year period, when some may not have reported accidents in their early years due to the fact that they were not in operation at that time. Thus we consider scenarios under the assumption: a)

the all facilities operated throughout the five-year reporting period, or b) facilities came "on line" at the rate of 3% per year, c) 5% per year, or d) 10% per year. Figure 2 gives the number of facilities by year of operation under these various scenarios. Table 2 shows the annual accident and injury rates that would have been observed, assuming that the facilities coming on-line were replacing facilities with a similar accident rate, under the three "replacement rate" scenarios. Clearly the U-shaped trend for accidents remains a consistent feature under a range of plausible scenarios for facility replacement rates, with rates rising through 1997 and then declining. For injuries, assuming an annual "replacement rate" for the facilities indicates that the injury rate may have actually been declining over the five year time period; no statistically significant trend could be ascertained, however, even assuming a 10% annual replacement rate.

Finally, to account for the possibly differing risk profiles among early and late-replying facilities, an analysis was conducted restricted to the dates in which all firms were part of the risk set (December 7, 1995-January 3, 1999). After considering possibly different annual rates, as in the previous analyses, as well as linear or quadratic daily trends, no statistically significant trend for either reported accidents or injuries could be discerned over this period.

Discussion

Our analysis shows that both the frequency and the underlying rate of reported accidents climbed from approximately one per 67 facilities in 1994 to one in 35 facilities in 1997 and 1998, then declined again to approximately one per sixty facilities in 2000. This change in rate was highly statistically significant ($p<.0001$). The underlying injury rate, however, did not show a stable trend, ranging from one injury per 30 facilities per year to one injury per 45 facilities per year over the 1994-2000 period.

However, some of this trend may be due to spurious data collection effects. It is unlikely that all of the 15,214 facilities operated throughout the five-year reporting period. A sensitivity analysis under a variety of assumptions about the likely "replacement rate" among the reporting facilities indicated that the "U-shaped" trend in the observed accident data was likely real. For injuries, very large replacement rates were required to reject the hypothesis that no change in injury rates occurred over the 1994-2000 period. Also, an analysis restricted to the period in which all firms were reporting (late 1995 through 1998) indicated no statistically significant trend in either accidents or injuries. Thus the observed U-shaped trend in accidents may be due to differing risk profiles among firms that reported early or late to RMP*Info.

<u>Limitations</u>

It was originally estimated by the U.S. Office of Management and Budget (OMB) that over 64,000 facilities would be required to submit RMPs for the 1995-1999 time period; however, only 15,219 ultimately did so. While much of this discrepancy is likely due to 1999 legislation that exempted propane dealers, it is possible that some facilities meeting the RMP*Info reporting requirements may have simply refused to respond. Any

accidents occurring at these non-responding facilities would obviously not contribute to the frequencies reported in this manuscript.

A further limitation is that the ability to "refile" near the close-out time may have allowed some facilities to avoid reporting accidents either immediately preceding or following the five-year reporting period, by choosing to refile if doing so would allow a five-year-old accident to be dropped, or by choosing not to refile if a recent accident had occurred. Such behavior could also produce at least part of the trend observed in Figure 1.

Finally, another limitation is that facilities that might have been required to report accidents under RMP*Info during the 1994-1999 time fames may have been missed because, by the end of the five-year period, they were no longer in operation, or had adjusted their processes in a fashion that allowed them to reduce inventories of toxic or hazardous chemical below the RMP*Info reporting thresholds.

Conclusion

This paper presents evidence that accident rates at US chemical facilities increased from 1994 through 1997 and declined from 1998 though 2000, while injury rates were roughly stable over this time period. However, spurious data collection issues such as not being in the risk set in the early part of the time period considered, or differences in faciltity types among early and late responders, may have caused part or all of this trend in accidents -- or masked a negative trend among injury rates. A second round of RMP*Info reporting is underway as we write, and data from this second round, covering the previous five years, should be available by the end of 2004. At this point, we hope to identify facilities that appeared in both the 1994-2000 and 1999-2004 reporting periods. Among this subset of facilities, at least, we should be able to obtain trend estimates under considerably weaker assumptions about facility reporting responsibilities.

Acknowledgements

Research for this paper was partially supported under a Cooperative Agreement with the Office of Emergency Preparedness, Prevention and Response (OEPPR) of the U. S. Environmental Protection Agency, supporting research at the Wharton Risk Management and Decision Processes Center of the University of Pennsylvania. The authors acknowledge helpful comments on earlier drafts from Robert Lowe, MD, MSCE, Oregon Health & Science University.

References

Kleindorfer, P.R., Belke, J.C., Elliott, M.R., Lee, K., Lowe, R.A., Feldman, H.I. (2003). Accident Epidemiology and the US Chemical Industry; Accident History and Worst-Case Data from RMP*Info. *Risk Analysis*, 23:865-881.

Liang, K-Y , Zeger, S.L. (1986), Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13-22.

McCullagh P, Nelder J. *Generalized linear models.* Chapman and Hall: London, 1989.

Year	Number of Accidents	Accident Rate (per facility)	Number of Injuries	Injury Rate (per facility)	Property Damage ($ millions)
1994 (partial)	101	.015 (.012,.018)	198	.026 (.014,.048)	309.8
1995 (partial)	324	.022 (.020,.025)	427	.029 (.022,.040)	68.3
1996	396	.026 (.023,.029)	393	.026 (.020,.034)	139.2
1997	439	.028 (.026,.032)	440	.029 (.022,.037)	218.4
1998	436	.028 (.026,.032)	399	.026 (.018,.038)	93.6
1999 (partial)	194	.023 (.020,.027)	209	.033 (.015,.075)	154.2
2000 (partial)	12	.016 (.009,.029)	17	.022 (.010,.049)	.1
Total	1902	.025 (.024,.026)	2083	.028 (.023,.033)	983.5

Table 1: Total number of accidents, injuries, and property damage among 15,214 RMP*Info facilities (excluding 5 missing postmark date). Table excludes 56 reported accidents that occurred outside of the five-year reporting window preceding postmark date. Accident and injury rates computed from binomial and negative binomial generalized linear models respectively, using daily accidents and injury counts and total number of facilities reporting for a given date; annualized rates are reported. 95% confidence interval in parentheses. P-value for annual differences in accident rate=<.0001; p-value for annual differences in injury rate=.98. Yellow highlight years for which all facilites were reporting.

Year	Accident Rate per facility (observed)	Accident Rate per facility (1)	Accident Rate per facility (2)	Accident Rate per facility (3)	Injury Rate per facility (observed)	Injury Rate per facility (1)	Injury Rate per facility (2)	Injury Rate per facility (3)
1994	.015 (.012,.018)	.017 (.014,.021)	.019 (.015,.023)	.024 (.020,.029)	.026 (.014,.048)	.031 (.017,.056)	.034 (.018,.061)	.043 (.024,.079)
1995	.022 (.020,.025)	.025 (.023,.028)	.028 (.025,.031)	.034 (.031,.038)	.029 (.022,.040)	.033 (.025,.045)	.036 (.027,.049)	.045 (.033,.061)
1996	.026 (.023,.029)	.028 (.026,.032)	.030 (.027,.034)	.036 (.032,.040)	.026 (.020,.034)	.028 (.022,.037)	.030 (023,.039)	.036 (.027,.046)
1997	.028 (.026,.032)	.030 (.028,.034)	.032 (.029,.035)	.036 (.033,.040)	.029 (.022,.037)	.031 (.024,.040)	.032 (.025,.042)	.036 (.028,.047)
1998	.028 (.026,.032)	.029 (.027,.033)	.030 (.027,.034)	.032 (.029,.036)	.026 (.018,.038)	.027 (.019,.039)	.028 (.019,.040)	.029 (.020,.042)
1999	.023 (.020,.027)	.023 (.020,.027)	.024 (.020,.028)	.024 (.021,.028)	.033 (.015,.075)	.033 (.015,.075)	.034 (.015,.076)	.034 (.015,.078)
2000	.016 (.009,.029)	.016 (.009,.029)	.016 (.009,.030)	.017 (.009,.030)	.022 (.010,.049)	.022 (.010,.049)	.022 (.010,.050)	.022 (.010,.050)
Total	.025 (.024,.026)	.027 (.026,.028)	.028 (.027,.030)	.032 (.031.034)	.028 (.025,.031)	.030 (.025,.036)	.032 (.027,.038)	.036 (.030,.043)
p	<.001	<.001	<.001	<.001	.98	.95	.90	.54

Table 2:Accidents and injury rates per year: (1) assuming that 3% of facilities were not at risk for each year prior to the postmark date; (2) assuming that 5% of facilities were not at risk for each year prior to the postmark date; (3) assuming that 10% of facilities were not at risk for each year prior to the postmark date. Accident and injury rates computed from binomial and negative binomial generalized linear models respectively, using daily accidents and injury counts and total number of facilities reporting for a given date; annualized rates are reported. 95% confidence interval in parentheses.

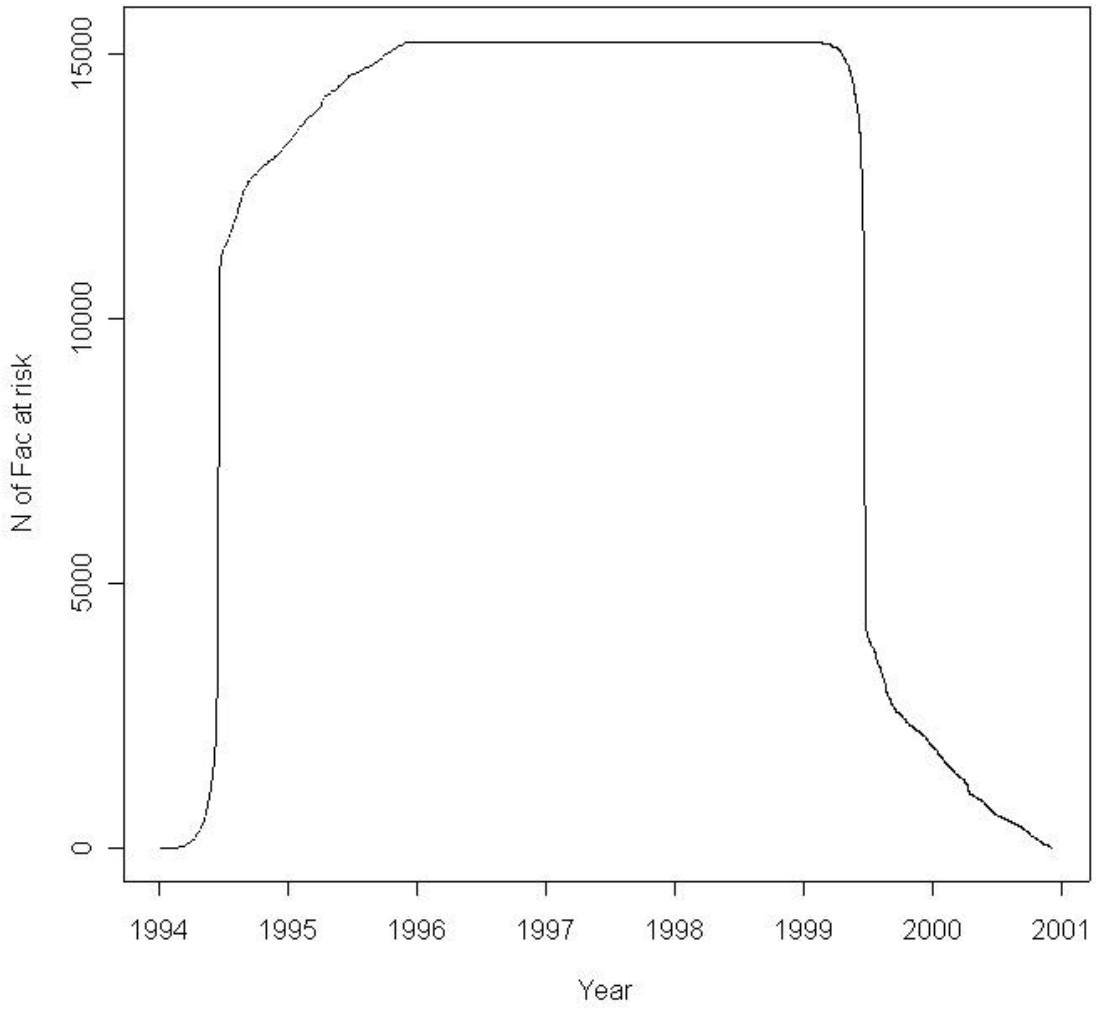

Figure 1: Number of Facilities At Risk by Year.

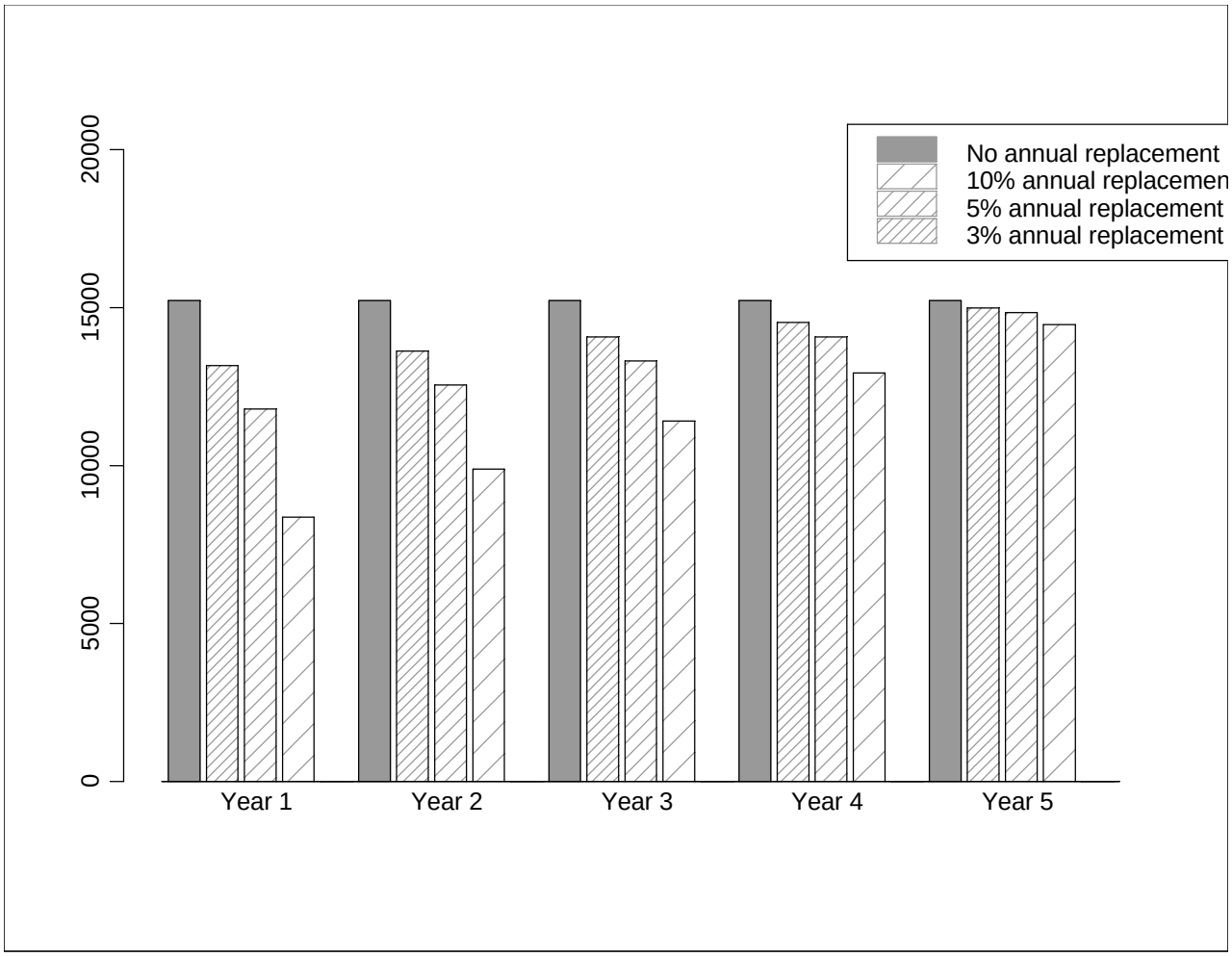

Figure 2: Number of facilities by year of operation, assuming a) full operation over 5 year reporting period, b) facilities turn over at the rate of 3% per year, c) facilities turn over at the rate of 5% per year, d) facilities turn over at the rate of 10% per year.