Analysis of GFCI Data

Michael A. Greene
Division of Hazard Analysis
U. S. Consumer Product Safety Commission

Summary

This report reanalyzes data from a survey of Ground-Fault Current Interrupters (GFCIs), conducted by the National Electrical Manufacturers Association (NEMA).[1] In that study, participating home inspectors in a number of cities added testing of receptacle and circuit breaker GFCIs to their normal home inspections. The home inspectors reported this data to NEMA. Also as part of the study, homeowners with GFCIs that were found to be non operational were given instructions for free device removal. Removed GFCIs were sent to Underwriters Laboratories. UL conducted further analyses on these GFCIs.

The survey report showed that overall, 14% of circuit breaker GFCIs and 8% of receptacle GFCIs were not operational. Most of the report contains tabulations of the effect of one or more variables at a time on GFCI operability. Tabulations compare the proportion operable in cities with high and low average temperature, high and low humidity and high and low number of lightning strikes and some other variables. The report did not analyze the effect of age on operability, except for a small number of failed devices. The report also contains a single table from a multivariate analysis (ANOVA), showing the simultaneous effects of humidity, lightning and temperature on operability.

Using the survey data that was provided to us by NEMA, we were able to analyze the data for receptacle GFCIs in different ways than presented in the report. Brief objectives of our analysis and our conclusions were as follows:

- Objective: To introduce a crude measure of age to assess if GFCI age or model year affects operability.

  Findings: We used property age as a proxy for GFCI age, cutting off age at 20 years. This variable is likely to overestimate GFCI age, and therefore underestimate the relationship between age and failure probability. Even so, we found a statistically valid relationship between GFCI age and operability.

- Objective: To compare the results from logistic regression with ANOVA. ANOVA was used in the NEMA analysis, but logistic regression is more commonly used in statistical analyses to estimate probabilities (in this case, the probability that a GFCI is not operable).

---

[1]National Electrical Manufacturers Association (2001), “GFCI Field Test Survey Report,” Arlington, VA, hereafter NEMA (2001).

Findings: We found that the results were comparable. Like the original analysis, the logistic regression analysis pointed to humidity as the most important environmental factor in GFCI operability. There were no practical differences between the results of the two analyses.

- Objective: To determine if the proxy for GFCI age and climate effects (i.e. humidity, temperature and lightning) were correlated. Correlation could hide the effects of the climatic conditions, or alternatively, correlation could make a particular climatic condition look stronger than it really is.

  Findings: We found no appreciable correlation between any of the variables. It does not appear that one variable may be masking the effect of any other variable.

- Objective: To use specialized software that was designed for analysis of survey data, to determine if the results are sensitive to clustering. Since there may be more than one GFCI sampled in a given property, it seems possible that operability of a particular GFCI may be correlated with the other GFCIs in the property as all are exposed to the same environmental factors.

  Findings: We found that clustering had no practical effect on the analysis.

In conclusion, different ways of analyzing the data produced similar results. Also, we believe that using GFCI age is important.

We understand that NEMA's unobtrusive survey design with home inspectors, was effective in gaining access to people's homes, in contrast to other designs that would be subject to other biases from lack of access, and additional costs. But we suspect that properties, that are primarily inspected usually at the time of sale, represent those in better than average condition. We wonder if a general survey of properties might have found a higher proportion of non operable GFCIs. A survey design that included some non resale properties might have been better to satisfy NEMA's objective of obtaining "statistically valid data to identify, define and quantify long term operation of GFCIs in the installed infrastructure."[2]

Introduction

The Ground-Fault Personnel Protection Section of the National Electrical Manufacturers Association (NEMA) conducted a survey of Ground-Fault Circuit Interrupters (GFCIs) during 1999 and 2000. The survey was conducted in 10 cities and surrounding areas. The study authors chose the cities to provide variability in lightning strikes, temperature and humidity; factors that were believed to be associated with GFCI

[2] NEMA (2001) page 1.

failure. The cities were arrayed into seven permutations that represented all but one of the combinations of possible high and low value of these three conditions.[3]

GFCIs were tested by home inspectors during the course of home inspections, that are often associated with sale or resale. Receptacle and circuit breaker (panel board) GFCIs were tested. The home inspector completed a data form with information about the property (single or multi-family), urban or suburban location, and its approximate age. The locations of the GFCIs in the building were noted. The home inspector determined the operability of the GFCI from whether it tripped and/or reset itself during the test. For receptacle GFCIs, an indicating lamp/load was used to help determine miswiring.

When non operating GFCIs were found, the home inspector was required to leave a packet with the property owner containing information about how the device could be replaced by a licensed electrician free of charge to the homeowner. Devices that were replaced were sent to Underwriters Laboratories for further operability analysis. UL then sent non operable GFCIs to the manufacturers for analysis and dating.

Study results are described below.

NEMA Study Results

A total of 2840 GFCIs were tested. GFCI data were excluded from the analysis in tests in homes six months old or newer, or where the test result had not been transcribed by the home inspector.[4] The remaining data included 2527 receptacle type GFCIs and 153 circuit breaker type GFCIs.

Overall 8.3% (211 of 2527) of receptacle GFCIs and 14.4% (22 of 153) of circuit breaker GFCIs were found to be non operational by the home inspectors.[5] The sample

---

[3] As follows:

High Lightning, Warm and Humid:	Birmingham, AL and Tampa, FL;
High Lightning, Warm and Dry:	Austin, TX;
High Lightning, Cool and Humid:	Washington, DC
High Lightning, Cool and Dry:	Kansas City, MO
Low Lightning, Warm and Humid:	<None>
Low Lightning, Warm and Dry:	Los Angeles, CA
Low Lightning , Cool and Humid:	Portland, ME and Seattle, WA
Low Lightning, Cool and Dry:	Denver, CO and Minneapolis, MN

[4] Exclusion of these new devices is reasonable because they did not have much exposure to the climatic conditions. See NEMA (2001), page 10, note 1.

[5] See NEMA (2001) page 11. The NEMA analysis and our reanalysis uses the home inspector determination of the GFCI operability.

size was too small for a multivariate analysis of the circuit breaker GFCIs. The remainder of the discussion in this paper is about the receptacle GFCIs only.

According to the report,

- In high lightning areas, 8.7% of receptacle GFCIs were not operational in contrast to 8.0% in low lightning areas.
- In warm areas, 8.5% of receptacle GFCIs were not operational in contrast to 8.2% in cooler areas.
- In high humidity areas 10.5% of receptacle GFCIs were not operational in contrast to 7.3% in dry areas.[6]

Tabulating these results by permutation showed the effect of humidity to be stronger than any other effect. Permutations are listed in table 1 below in decreasing order of the proportion non operational.

Table 1
Percent of Non Operable Receptacle GFCI by Permutation

Permutation	Lightning	Temperature	Humidity	Number Tested	Percent Non Operational
7	Low	Cool	Humid	287	11.1
1	High	Warm	Humid	274	10.9
3	High	Cool	Humid	256	9.4
2	High	Warm	Dry	342	8.5
8	Low	Cool	Dry	618	7.3
4	High	Cool	Dry	393	6.9
6	Low	Warm	Dry	357	6.7
All				2,527	8.3

Source: NEMA (2001), page 12, rearranged.

The report also examined the relationship between GFCI location and operability. This is shown in table 2 below.

[6] NEMA (2001) pages 13-14.

Table 2
Percent of Non Operable GFCIs by Location

Location	Number Tested	Percent Non Operational
Unknown	8	25.0
Outdoor	197	20.3
Bathroom	1071	9.5
Garage	418	7.4
Basement	122	7.4
Kitchen	706	3.8
Other	5	0.0
All	2,527	8.3

Source: NEMA (2001), page 19.

Of the 211 non operational GFCIs, 90 were removed by an electrician and sent to UL for testing.[7] Among these recovered GFCIs, UL found 56% operated correctly in the laboratory after being correctly wired for line and load. The recovered GFCIs also included some that were miswired (13 of 90 returned or 14%) as determined by the electrician. Some of the other GFCIs that operated correctly at UL may have been miswired without being detected by the electrician.[8]

The report also presented the age distribution of the 90 recovered GFCIs. There was no particular pattern.

Assessment

The NEMA report did not draw any conclusions about the effect of climate or GFCI location on GFCI operability rates. Instead, the report produced summary tables showing the percent operable by one or two variables, or by the entire permutation. There was no attempt made to account for the effect of GFCI location within the property or GFCI age, nor to determine the statistical strength of the relationship between humidity and GFCI operability as shown in table 1 above.

[7] This required the homeowner to contact the electrician for a free GFCI replacement after the home inspection. In the majority of cases, the electrician was not contacted.

[8] NEMA (2001) pages 20,21. Note that the determination about wiring was only made when the electrician removed the GFCI.

The purpose of this reanalysis was to determine if the relationship between climatic conditions (humidity, temperature and lightning) and GFCI operability was statistically significant, while taking into account the effect of GFCI location and age in a multivariate analysis. As described above, analyzing one variable at a time could lead one variable to mask another. For example, if most of the outdoor receptacle GFCIs were in high humidity areas, this could appear to produce a humidity effect rather than a location effect or mixed location-humidity effect.

The report presented a multivariate analysis of the climate main effects (lightning, temperature and humidity) and all possible interactions using analysis of variance (ANOVA). ANOVA is not an appropriate technique for binary response variables like GFCI operability, which is recorded as either 0 (non operating) or 1 (operating). ANOVA is inappropriate because (1) the residuals do not have a normal distribution and (2) the estimated conditional probabilities could be less than zero or greater than one. The standard approach is logistic regression.[9]

The reanalysis is subject to some of the same cautions as the original analysis. These include the possibility of incorrect results from the home inspector and/or electrician, and the inability to obtain good estimates of GFCI age. This last problem would lead to age appearing as less of a factor in non operability than it would in reality. Another caution is that the sample might be biased toward better than typical GFCIs because homes may be in better than typical condition when inspected.

Data Preparation

NEMA provided CPSC with an Excel spread sheet containing the study data from Appendix F.[10] The spread sheet was translated into a SAS® dataset.[11] To make sure that the data and analyses agreed with the NEMA report, we were able to reproduce all of their tabulations and data summaries.

We created a GFCI age variable from the reported age of the building or the installation date of the circuit breaker GFCI.[12] Because the ages of some buildings predated the widespread installation of GFCIs, we truncated age to 20 years for buildings

[9] See D. W. Hosmer and S. Lemeshow (1989), *Applied Logistic Regression.* John Wiley and Sons, NY, pages 5-7.

[10] NEMA (2001) pages F-2 to F-53.

[11] SAS is a statistical software system. See SAS Institute Inc. (1990), *SAS® Language: Reference, Version 6, First Edition.* Cary, NC: SAS Institute, Inc.

[12] Installation dates of circuit breaker GFCIs were not often found in the data. Most of the time we used the reported age. See NEMA (2001), Appendix F.

20 years old and older. Aside from the unusual building with older devices, this would generally result in overestimating the age of GFCIs.[13]

Like the tabulations in the original report, we also used the definition of GFCI operability as found by the home inspector, not as later determined by UL.[14] The analysis was restricted to receptacle type GFCIs because the sample size for the circuit breaker devices was very small.

To estimate separate effects of the different variables, we estimated a number of logistic regressions. This method allows estimates of odds ratios, which are typically interpreted as relative risks.[15] Confidence intervals are typically used for odds ratios, where the odds ratio is considered to be significant when the confidence interval does not include one.

Results

Table 3 presents results from logistic regression on main effects only. Higher order interactions were not significant.

---

[13] For example if a homeowner in a 15 year old property installed GFCIs 5 years ago, then the age of the GFCIs would be actually 5 years, while we would represent it as 15 years.

[14] As noted above, some of the GFCIs returned to UL were found to be operable. We did not correct the analysis because only about half the GFCIs that were found by home inspectors to be inoperable were returned to UL.

[15] For example, in this study, the relative risk of failure attributable to lightning is the probability that a GFCI in a high lightning area failed divided by the probability that a GFCI in a low lightning area failed.

Table 3
Odds Ratios for Logistic Regression Predicting the Probability of Non Operational Receptacle GFCIs

Variable	Odds Ratio	95% Confidence Interval	p-value
Lightning Density High vs. Low	1.08	0.79 , 1.47	0.6320
Temperature Warm vs. Cool	1.09	0.80 , 1.49	0.5770
Humidity Humid vs. Dry	1.57	1.15 , 2.14	0.0043
Age per year	1.04	1.02 , 1.07	0.0002
Location (vs. Bathroom)			
Outdoor	2.44	1.62 , 3.68	0.0001
Unknown	1.65	0.36 , 7.64	0.5221
Basement	0.91	0.44 , 1.86	0.7896
Garage	0.84	0.55 , 1.30	0.4402
Kitchen	0.41	0.26 , 0.64	0.0001

Data provided by NEMA. The regression used Proc Probit in SAS® (with the logistic option). P-values less than 0.05 are considered significant.

In table 3, humidity, age and location of the GFCI are shown to be significant predictors of the probability that a GFCI is not operational. The table shows that a GFCI in a humid area is 1.57 times as likely to be non operable as one in a dry area. Where the GFCI is located also affects the operability rate. An outdoor GFCI is non operable almost 2.5 times as often as a GFCI in a bathroom, while GFCIs in kitchens are less than half as likely to be non operable as in bathrooms. The other locations do not have significantly different rates than bathroom locations.

These effects are very close to the univariate results shown earlier. For example, as shown above, 10.5% of receptacle GFCIs were non operational in humid areas, in contrast to 7.3% in drier areas. This would give a univariate odds ratio of 1.49, which is close to the regression result of 1.57 as shown above for humidity.[16] Also, the location variables generally follow the same pattern as shown in the univariate analysis in table 2.

[16] The univariate odds ratio is computed as (0.105/0.895) /(0.073/0.927) = 1.49.

That the univariate percentages and the multivariate regression results produced similar results suggests that there is little, if any, confounding or correlation among the design variables (temperature, humidity and lightning). There is also no apparent correlation between GFCI age, location and any of the climate variables.

The results also were similar to the NEMA multivariate analysis that used ANOVA. Table 6.1, (page 9) shows the largest coefficient associated with "climate dry/humid" at 0.0131, as compared with the other main effects at 0.0002 for "cool/warm" and 0.0027 for high and low lightning.

Using the crude measure of age described above, the age odds ratio of 1.04 (95% confidence interval 1.02-1.07) suggests that there is a statistically significant relationship between non operability of the receptacle GFCIs and age. Table 4 below shows the percent operable by age, as defined above.

Table 4
Percent Non Operable Receptacle GFCIs by GFCI Age

Age (years)	Percent Non Operable	Sample Size
-2	2.9	137
2-4	5.7	296
4-6	5.7	279
6-8	6.8	191
8-10	5.3	151
10-12	8.9	214
12-14	9.0	156
14-16	7.8	116
16-18	6.4	63
18-20	11.8	51
20+	13.5	607
Unknown	7.1	266
All Ages	8.3	2,527

Data provided by NEMA. Right endpoint not included in the interval, i.e. property two years old is found in the cell 2-4 years.

As discussed above the crude age measure of using property age is likely to overestimate GFCI age. In turn, if age and non operability are truly related, then our measured age/non operability relationship will underestimate the way that GFCIs become non operational with age. For example, some of the GFCIs that we labelled as 10-12 years old above (from the property age), may actually have been installed more recently. This means that we are underestimating the age/non operability relationship of the 10-12 year old GFCIs because they are actually a mixture of 10-12 year old GFCIs and newer devices. But given these difficulties in measuring ages, both the univariate relationship in table 4 and the multivariate logistic regressions show an increasing trend of non operability with increasing age.

We ran the same logistic regression using SUDAAN®, which produced very similar values for odds ratios, confidence intervals and p-values.[17] Technically, the survey is a stratified cluster sample, where the eight climate permutations are the strata and the property is the primary sampling unit (PSU). It is reasonable to expect that all the GFCIs within a property are likely to have similar operability rates because they are exposed to the identical climate, household power consumption, and other similar factors. The SUDAAN software corrects for this clustering in estimation of the coefficients, odds ratios and confidence intervals. Differences between SUDAAN odds ratios and SAS odds ratios were at the second decimal place (i.e. hundreds), which were not of any practical importance. This suggests that that the GFCI operability rates seem to be independent within a property.

Conclusions

As a result of this analysis, it appears that GFCIs in humid areas of the country and those located outdoors are more likely to be non operable as compared with those in dry climates and indoors. GFCIs in older properties (which are likely to be older GFCIs) are also more prone to be non operable, although it is not clear whether this effect is due to environmental damage over time (such as humidity or rain water damage) or the possibility that older GFCIs are not designed as well as the newer devices. This result also has to be tempered by the crude way that GFCI age was measured in our analysis.

We do not know the effect on the sample of GFCIs from using home inspectors. Because home inspections are typically associated with sales, there is the possibility that the properties and the GFCIs tested in the study were in better condition than in typical homes.

If the study is conducted again, we would recommend that some GFCIs be sampled from non resale properties. This would provide a better idea of the conditions of GFCIs nationwide, not just in resale properties. We would also recommend that the age of at least some operable GFCIs be determined so that the relationship between age and operability could be analyzed.

---

[17] Shah, BV, Barnwell, BG and Bieler, GS (1997), *SUDAAN User's Manual, Release 7.5*. Research Triange Institute, Research Triangle Park, NC.