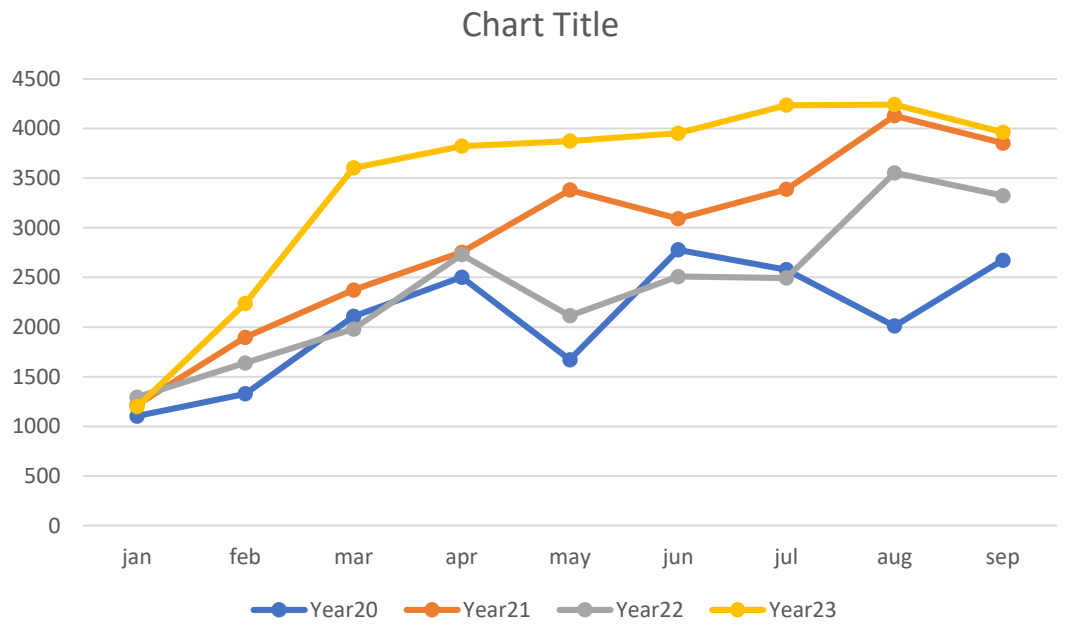
Cyber Security and Risk Management

Grey Rolling (GM) Model and Survival Analysis

GM(1,1) Model

- Professor Deng (1982) proposed the GM(1,1) model.
- GM(1,1) is a single-variable grey prediction model with a first-order difference equation.
- Its greatest feature is that GM(1,1)has only a dependent variable but no independent variables.
- It has been widely used to solve various prediction problems.



GM(1,1) Model (Contd…)

- To evaluate the overall effectiveness of the GM(1, 1, ⊗b) model, it is imperative to assess both its simulation and prediction capabilities concurrently.
- In this study, the initial nine data points serve as the basis for constructing the GM(1, 1, ⊗b) model, while the remaining three data points are reserved for testing the predictive accuracy of the model.
- The modeling of GM (1,1) is done using the following steps which are as follows:
- Assuming that $X^{(0)}$ is a nonnegative sequence, where
- $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5), x^{(0)}(6), x^{(0)}(7), x^{(0)}(8), x^{(0)}(9))$. *(1)*
- Then based on the above equation, the subsequent equation i.e. $X^{(1)}$ called as 1-AGO (Accumulating Generation Operator) sequence is computed as:
- $X^{(1)} = (((i), x^0(i+1, \ldots, x^0(k)))$ *(2)*
- and $z^{(1)}(k)$ called as the mean generation of consecutive neighbors sequence of $X^{(1)}$ which is computed as:
- $z^{(1)}(k) = 0.5*(x^{(1)}(k) + x^{(1)}(k-1))$; $k=2, 3, \ldots, n.$ *(3)*
- Constructing the Matrices Y and B to compute the parameters a and b using the least squares method. The modeling of GM (1,1) is done using the following steps which are as follows:

GM(1,1) Model (Contd…)

- Assuming that $X^{(0)}$ is a nonnegative sequence, where
- $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5), x^{(0)}(6), x^{(0)}(7), x^{(0)}(8), x^{(0)}(9))$. (*1)*
- Then based on the above equation, the subsequent equation i.e. $X^{(1)}$ called as 1-AGO (Accumulating Generation Operator) sequence is computed as:
- $X^{(1)} = (((i), x^0(i+1, \ldots, x^0(k)))$ (*2)*
- and $z^{(1)}(k)$ called as the mean generation of consecutive neighbors sequence of $X^{(1)}$ which is computed as:
- $z^{(1)}(k) = 0.5*(x^{(1)}(k) + x^{(1)}(k-1));$ $k=2, 3, \ldots, n.$ (*3)*
- Constructing the Matrices Y and B to compute the parameters a and b using the least squares method.

GM(1,1) Model (Contd…)

- The Matrices Y and B are constructed as follows:

- $Y = \begin{matrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{matrix}$ $B = \begin{matrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \\ -z^{(1)}(n) & 1 \end{matrix}$ *(4)*

- The value of a and b are computed using the given equations:

- $x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 2, 3, \ldots, n.$ *(5)*

- $\hat{a} = (B^T B)^{-1} B^T Y = \begin{matrix} a \\ b \end{matrix}$ *(6)*

GM(1,1) Model (Contd…)

- From equation *5*, the different values of grey action quantity can be calculated as:
- k=2 $b^2 = x^{(0)}(k) + az^{(1)}(2)$,
- k=3 $b^2 = x^{(0)}(k) + az^{(1)}(3)$,
- :
- :
- k=n $b^2 = x^{(0)}(k) + az^{(1)}(n)$, k = 2, 3, …, n. *(7)*
- The simulated data are obtained by the following equations:
- $\hat{x}^{(0)}_{min}(k) = (1 - e^{a})(x^{(0)}(1) - \frac{b_{min}}{a}) e^{-ak}$, k= 1,2,3,4……n. *(8)*
- $\hat{x}^{(0)}_{mid}(k) = (1 - e^{a})(x^{(0)}(1) - \frac{b_{mid}}{a}) e^{-ak}$, k= 1,2,3,4……n. *(9)*
- $\hat{x}^{(0)}_{max}(k) = (1 - e^{a})(x^{(0)}(1) - \frac{b_{max}}{a}) e^{-ak}$, k= 1,2,3,4……n. *(10)*

GM(1,1) Model (Contd…)

- After estimating the simulation and predicted values, the sigmoid curve (S-curve) was drawn to assess the performance of the vulnerabilities with respect to time.

- Further, the goodness of fit for the given variables were computed. These computations were done using MATLAB R2023b.

GM(1,1) Model (Contd…)

- **Empirical Result (Year 2023)**

- The original time series data can be written in the form of $X^{(0)}$ as:

$$X^{(0)} = (1199, 2238, 3603, 3822, 3874, 3953, 4235, 4242, 3963)$$

- Now generating the time series data $X^{(1)}$ from $X^{(0)}$ as:

$$X^{(1)} = (1199, 3437, 7040, 10862, 14736, 18689, 22924, 27166, 31129)$$

- The partial series data $z^{(1)}(k)$ from $X^{(1)}$ can be written as:

$$z^{(1)}(k) = (2318, 5238.5, 8951, 12799, 16712.5, 20806.5, 25045, 29147.5)$$

GM(1,1) Model (Contd…)

- Now, the matrices Y and B are formed as given:

Y =	2238	B =	-2318	1
	3603		-5238.5	1
	3822		-8951	1
	3874		-12799	1
	3953		-16713	1
	4235		-20807	1
	4242		-25045	1
	3963		-29148	1

- The value of a and b are computed using the equation *6*:

a = -0.048856

GM(1,1) Model (Contd…)

- The different values of b are computed using equation 7:

  $Bs = \{b_2, b_3, \ldots, b_n\}$

  $= \{2124.751, 3347.066, 3384.687, 3248.688, 3136.490, 3218.472, 3018.395, 2538.962\}$

- Then, the maximum, minimum and mid value of b are computed.

  $b_{max} = 70477.507$

  $b_{min} = 44688.859$

  $b_{mid} = 62648.449$

GM(1,1) Model (Contd…)

- The simulation and prediction data $\hat{x}^{(0)}_{min}$(k), $\hat{x}^{(0)}_{mid}$(k) and $\hat{x}^{(0)}_{max}$(k) were computed using the equations *8, 9 and 10*:

$\hat{x}^{(0)}_{min}$(k)	$\hat{x}^{(0)}_{mid}$(k)	$\hat{x}^{(0)}_{max}$(k)
$\hat{x}^{(0)}_{min}(2)$ = 2237.543	$\hat{x}^{(0)}_{mid}(2)$ = 3136.769	$\hat{x}^{(0)}_{max}(2)$ = 3528.765
$\hat{x}^{(0)}_{min}(3)$ = 2349.576	$\hat{x}^{(0)}_{mid}(3)$ = 3293.825	$\hat{x}^{(0)}_{max}(3)$ = 3705.448
$\hat{x}^{(0)}_{min}(4)$ = 2467.218	$\hat{x}^{(0)}_{mid}(4)$ = 3458.745	$\hat{x}^{(0)}_{max}(4)$ = 3890.978
$\hat{x}^{(0)}_{min}(5)$ = 2590.750	$\hat{x}^{(0)}_{mid}(5)$ = 3631.922	$\hat{x}^{(0)}_{max}(5)$ = 4085.797
$\hat{x}^{(0)}_{min}(6)$ = 2720.467	$\hat{x}^{(0)}_{mid}(6)$ = 3813.770	$\hat{x}^{(0)}_{max}(6)$ =4290.370
$\hat{x}^{(0)}_{min}(7)$ = 2856.679	$\hat{x}^{(0)}_{mid}(7)$ = 4004.724	$\hat{x}^{(0)}_{max}(7)$ = 4505.186
$\hat{x}^{(0)}_{min}(8)$ = 2999.711	$\hat{x}^{(0)}_{mid}(8)$ = 4205.238	$\hat{x}^{(0)}_{max}(8)$ = 4730.758
$\hat{x}^{(0)}_{min}(9)$ = 3149.905	$\hat{x}^{(0)}_{mid}(9)$ = 4415.791	$\hat{x}^{(0)}_{max}(9)$ = 4967.624

GM(1,1) Model (Contd…)

- Similarly, when k=10, 11, 12, the predicted data can be computed for $\hat{x}_{min}^{(0)}$(k), $\hat{x}_{mid}^{(0)}$(k) and $\hat{x}_{max}^{(0)}$(k) as follows:

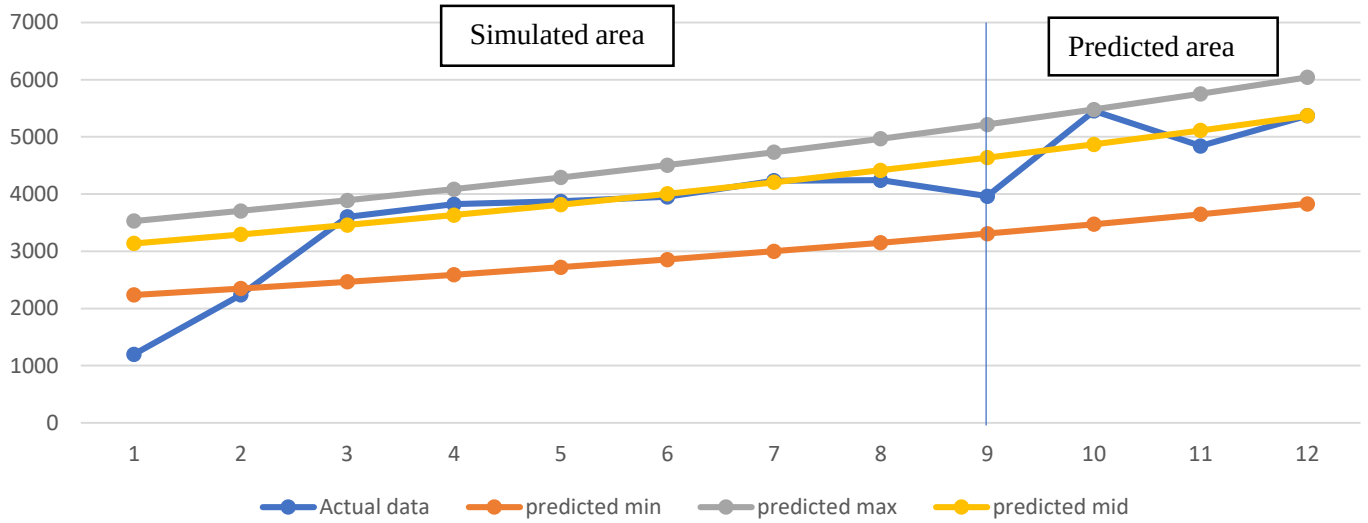
$\hat{\boldsymbol{x}}_{\boldsymbol{min}}^{(\mathbf{0})}$(k)	$\hat{\boldsymbol{x}}_{\boldsymbol{mid}}^{(\mathbf{0})}$(k)	$\hat{\boldsymbol{x}}_{\boldsymbol{max}}^{(\mathbf{0})}$(k)
$\hat{x}_{min}^{(0)}(10)$ = 3473.229	$\hat{x}_{mid}^{(0)}(10)$ = 4869.053	$\hat{x}_{max}^{(0)}(10)$ = 5477.530
$\hat{x}_{min}^{(0)}(11)$ = 3647.132	$\hat{x}_{mid}^{(0)}(11)$ = 5112.844	$\hat{x}_{max}^{(0)}(11)$ = 5751.786
$\hat{x}_{min}^{(0)}(12)$ = 3829.742	$\hat{x}_{mid}^{(0)}(12)$ = 5368.841	$\hat{x}_{max}^{(0)}(12)$ = 6039.775

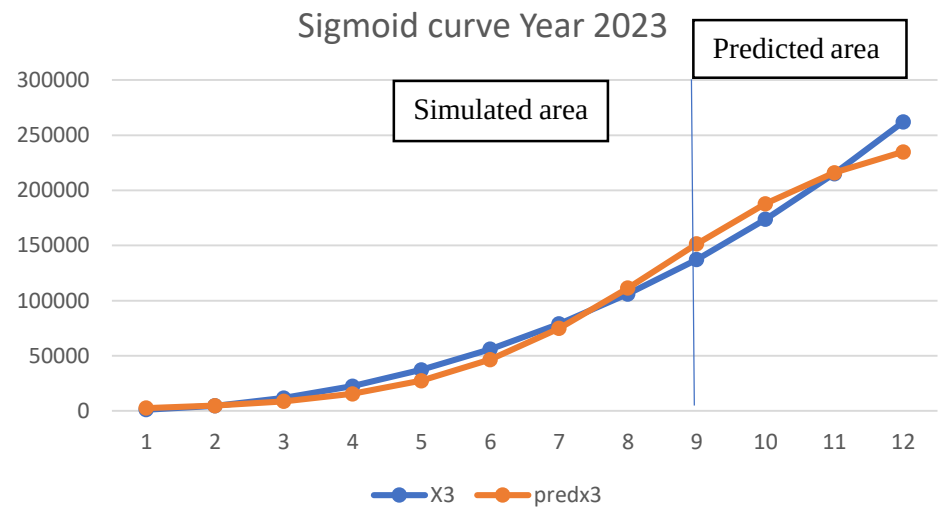
GM(1,1) Model (Contd…)

- The following conclusions can be drawn from the fig below:

a) The overall trend of vulnerabilities is increasing year by year but slight downfall can be seen from the month July to September and again increasing from October to December. The deviations in the curve can be seen from the actual data.

b) From the above fig, it can be seen that the actual value of the number of vulnerabilities is smaller than the predicted value of the upper bound but larger than the predicted value of lower bound. This shows that the GM (1,1, ⊗b) model is effective in predicting the range of vulnerabilities in the next three months and proves the rationality of the predicted results of the GM (1,1, ⊗b) model.





GM(1,1) Model (Contd…)

- **Sigmoid curve results**
- The sigmoid curve typically exhibits S-shaped pattern. At the beginning or starting phase, the curve is almost flat and then slowly rises, then in the middle phase, it accelerates (shows fast growth) and finally, it levels off.
- From the above figure, it can be seen that the vulnerabilities in the starting phase are showing almost constant growth, but finally a rapid growth can be seen in the number of vulnerabilities with respect to time.
- The Chi-square goodness of fit p-value was obtained <0.99 for the year 2023 which means that the observed data does not significantly deviate from the expected data. It suggests that the observed data closely matches the expected values as predicted by the model.

Year	Chi-sq (p-value)
2023	<0.99

Descriptive Statistics

The table given below represents the number of vulnerabilities/attacks per year (Microsoft)

	Year 2020	Year 2021	Year 2022	Year 2023
Total (Vulnerabilities)	393	287	232	390
Total (Killed count/year)	358	248	189	366
Maximum (Killed count/day)	53	27	21	42
Minimum (Killed count/day)	1	1	1	1

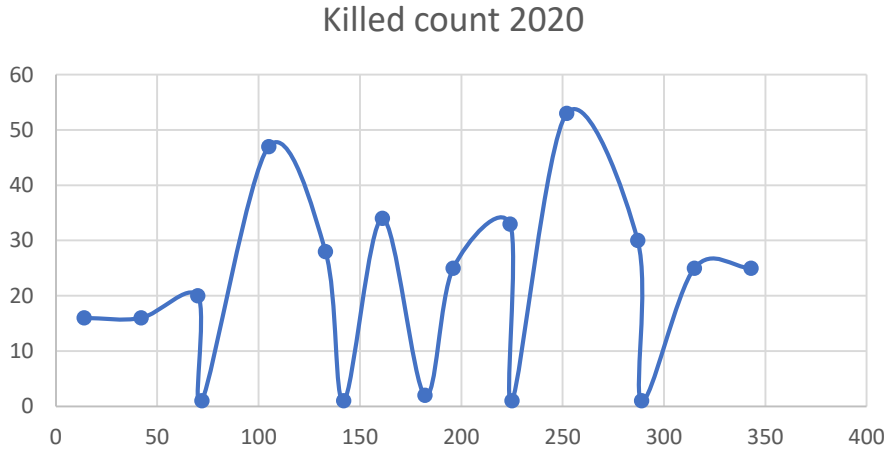
The table given below represents the number of vulnerabilities/attacks per week (Overall)

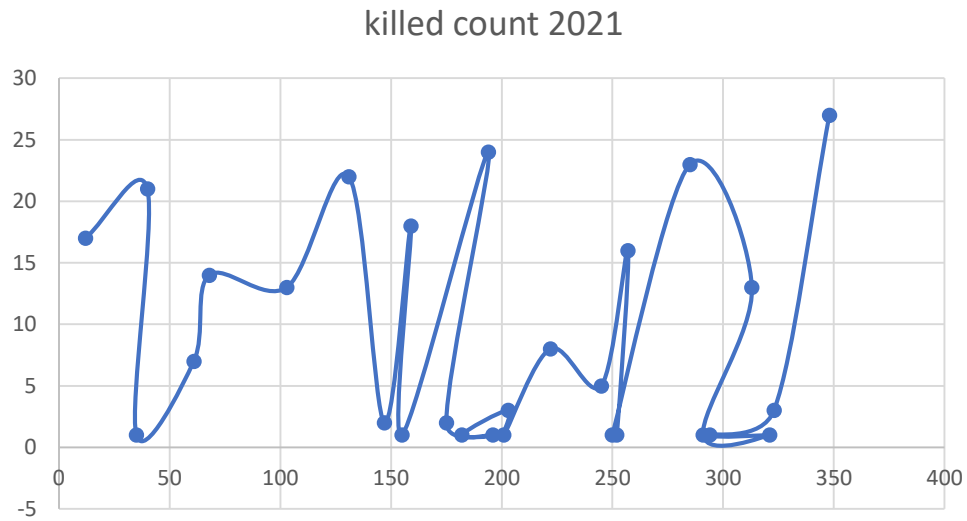
	Year 2020	Year 2021	Year 2022	Year 2023
Total (Vulnerabilities)	26595	30780	37461	46794
Average (Vulnerabilities/week)	502	581	707	883
Maximum (Vulnerabilities/week)	1143	1099	1377	1747
Minimum (Vulnerabilities/week)	14	6	1	60

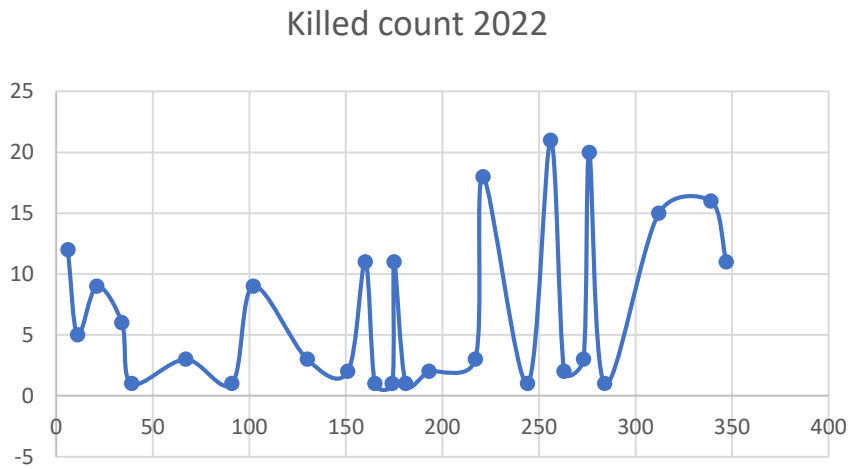
- Source: Table 1- https://msrc.microsoft.com/update-guide

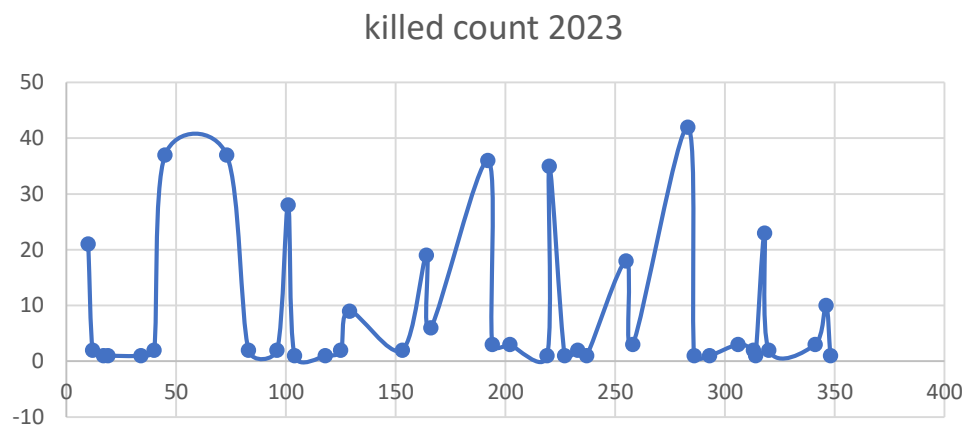
  Table 2- https://www.cve.org/Downloads

Graphical representation of killed counts









Survival Analysis

- The survival function, here is used to assess the lifetime of CVE’s.

- Survival function graphs visually depict the estimated likelihood that a subject will survive beyond a specific time point in survival analysis.

- These graphs play a crucial role in comprehending the survival journey of a population or a studied group.

- Below is an explanation of the elements and the interpretation of survival function graphs.

- The **survival function** provides the probability of surviving past a specific time point, while the **cumulative hazard function** indicates the cumulative risk or failure rate leading up to that time point.

- Both functions are crucial in survival analysis for understanding time-to-event data and evaluating associated risk factors.

Survival Analysis (Contd…)

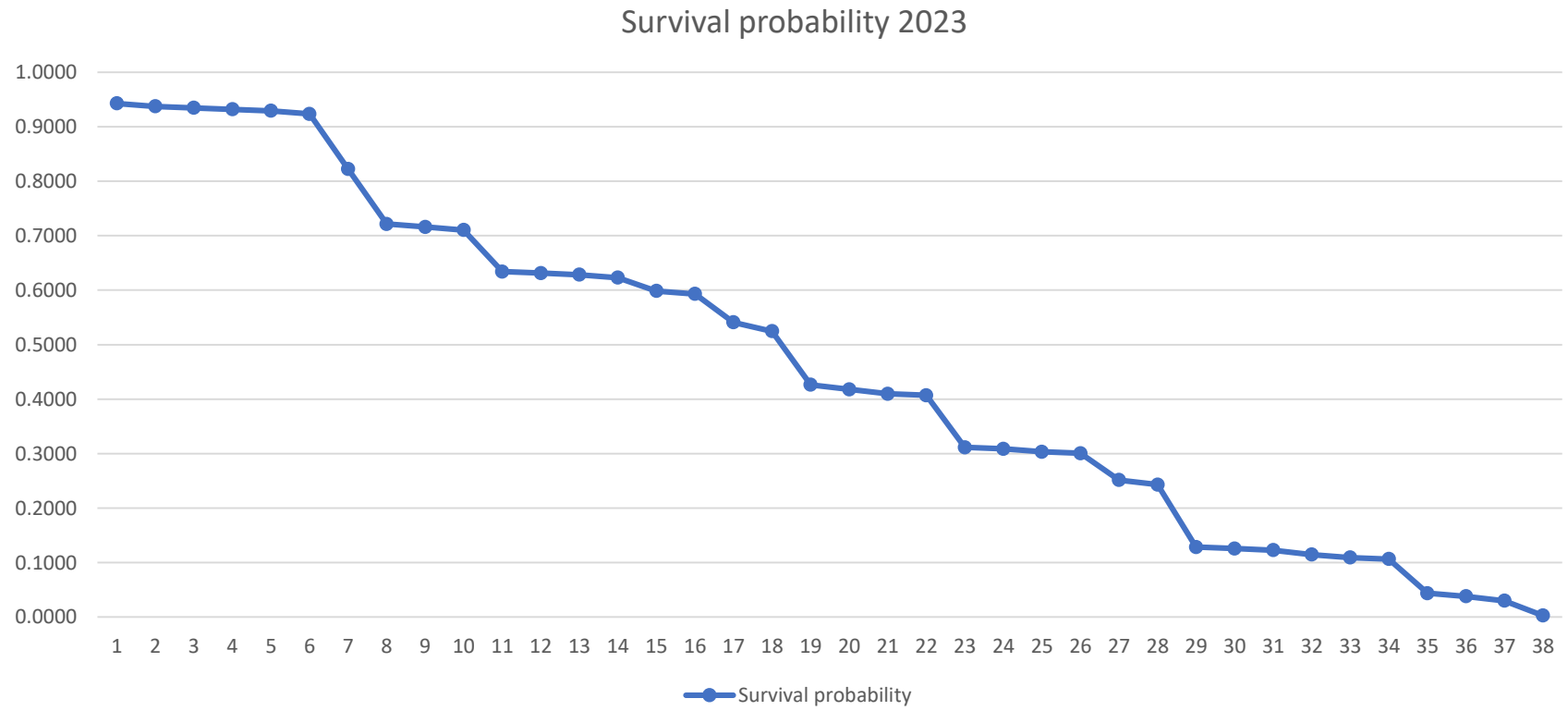
- **Empirical Results (Year 2023)**

- The survival analysis comprises of 393 vulnerabilities, 366 of whom were killed as their patch was made available. The number of censored vulnerabilities were 24 as their patch was not available. Our analysis observed a total of 365 days. (Table 1)

Time First seen (day)	First seen	CVE	Time Last seen (day)	Last seen	Microsoft patch CVE	Killed count	censored
11	Jan 11, 2023	CVE-2023-21554	101	Apr 11, 2023	CVE-2023-21554	28	0
31	Jan 31, 2023	CVE-2023-24934	104	Apr 14, 2023	CVE-2023-24934	1	0
31	Jan 31, 2023	CVE-2023-29334	118	Apr 28, 2023	CVE-2023-29334	1	0
31	Jan 31, 2023	CVE-2023-29350	125	May 5, 2023	CVE-2023-29350	2	0
32	Feb 1, 2023	CVE-2023-24881	129	May 9, 2023	CVE-2023-24881	9	0
32	Feb 1, 2023	CVE-2023-29345	153	Jun 2, 2023	CVE-2023-29345	2	0
72	Mar 13, 2023	CVE-2023-28310	164	Jun 13, 2023	CVE-2023-28310	19	0
72	Mar 13, 2023	CVE-2023-29349	166	Jun 15, 2023	CVE-2023-29349	6	0
72	Mar 13, 2023	CVE-2023-32033	192	Jul 11, 2023	CVE-2023-32033	36	0
72	Mar 13, 2023	CVE-2023-36883	194	Jul 13, 2023	CVE-2023-36883	3	0
94	Apr 4, 2023	CVE-2023-35392	202	Jul 21, 2023	CVE-2023-35392	3	0
94	Apr 4, 2023	CVE-2023-38157	219	Aug 7, 2023	CVE-2023-38157	1	0

Survival Analysis (Contd…)

• The plots for the survival probability, cumulative hazard function are represented in the fig. 1, and 2 respectively.



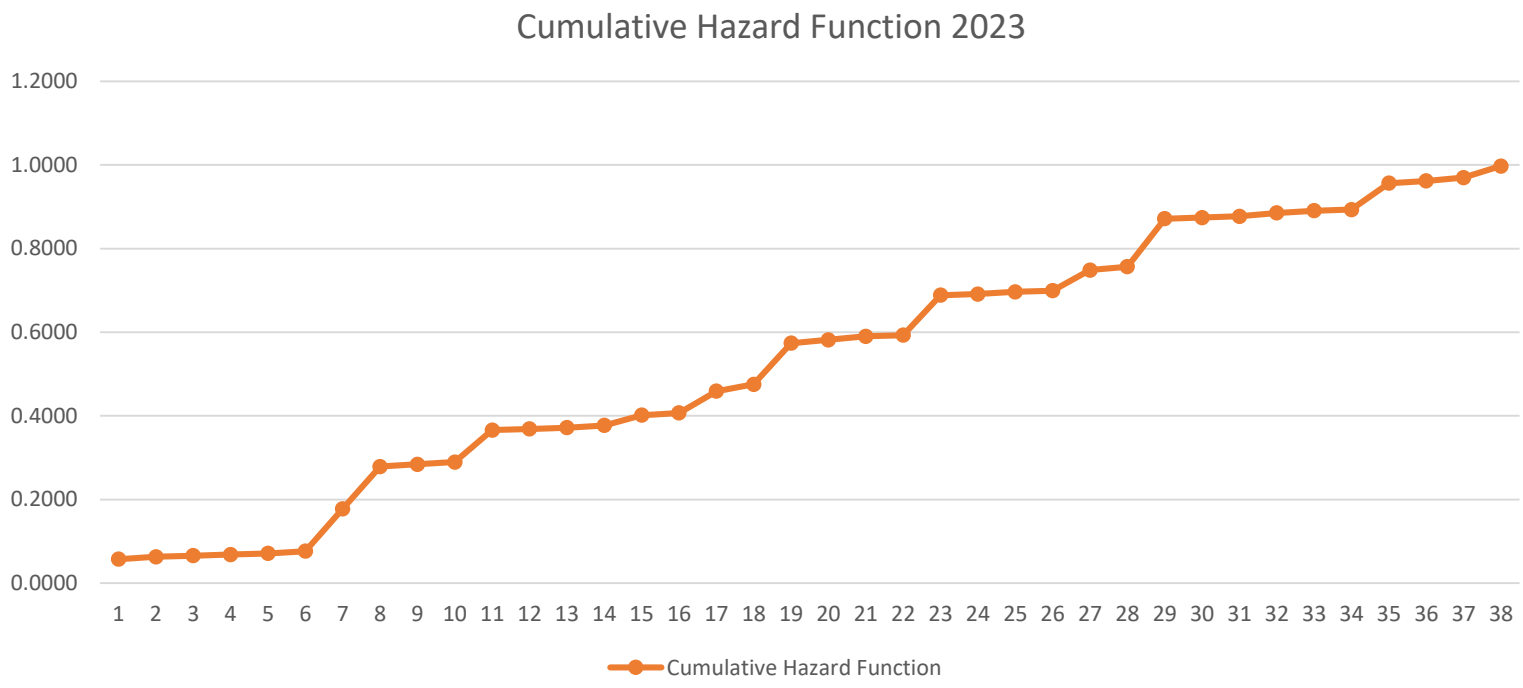
Survival Analysis (Contd…)

- From the above figure it can be said that the overall trend of vulnerabilities is decreasing with respect to time. We can see that 92% of the vulnerabilities are alive for about 40^{th} days, 63% of the vulnerabilities live for about 104^{th} days and 03% were alive for about 320^{th} days.
- As the period of time increases, the number of vulnerabilities decreases as the patch was made available for them.

Time	Survival probability
10	0.9426
40	0.9235
45	0.8224
104	0.6311
118	0.6284
283	0.1284
286	0.1257
320	0.0383
341	0.0301
346	0.0027

Survival Analysis (Contd…)

- The plots for the cumulative hazard function are represented in the fig. below.



Survival Analysis (Contd…)

- From the above figure it can be said that the overall trend of vulnerabilities is increasing with respect to time. We can see that 07% of the vulnerabilities were found to be killed at 40^{th} day, 36% of the vulnerabilities were found to be killed at 104^{th} day and 96% were found to be killed at 320^{th} day. As the period of time increases, the number of killed vulnerabilities increases as the patch was made available for them.

Time	Cumulative Hazard Function
10	0.0574
40	0.0765
101	0.3661
104	0.3689
306	0.8852
313	0.8907
314	0.8934
318	0.9563
320	0.9617

Thank You