



National Outbreaks

Michaela Danielle, Panfilo Marbibi, Tahseen Shaik

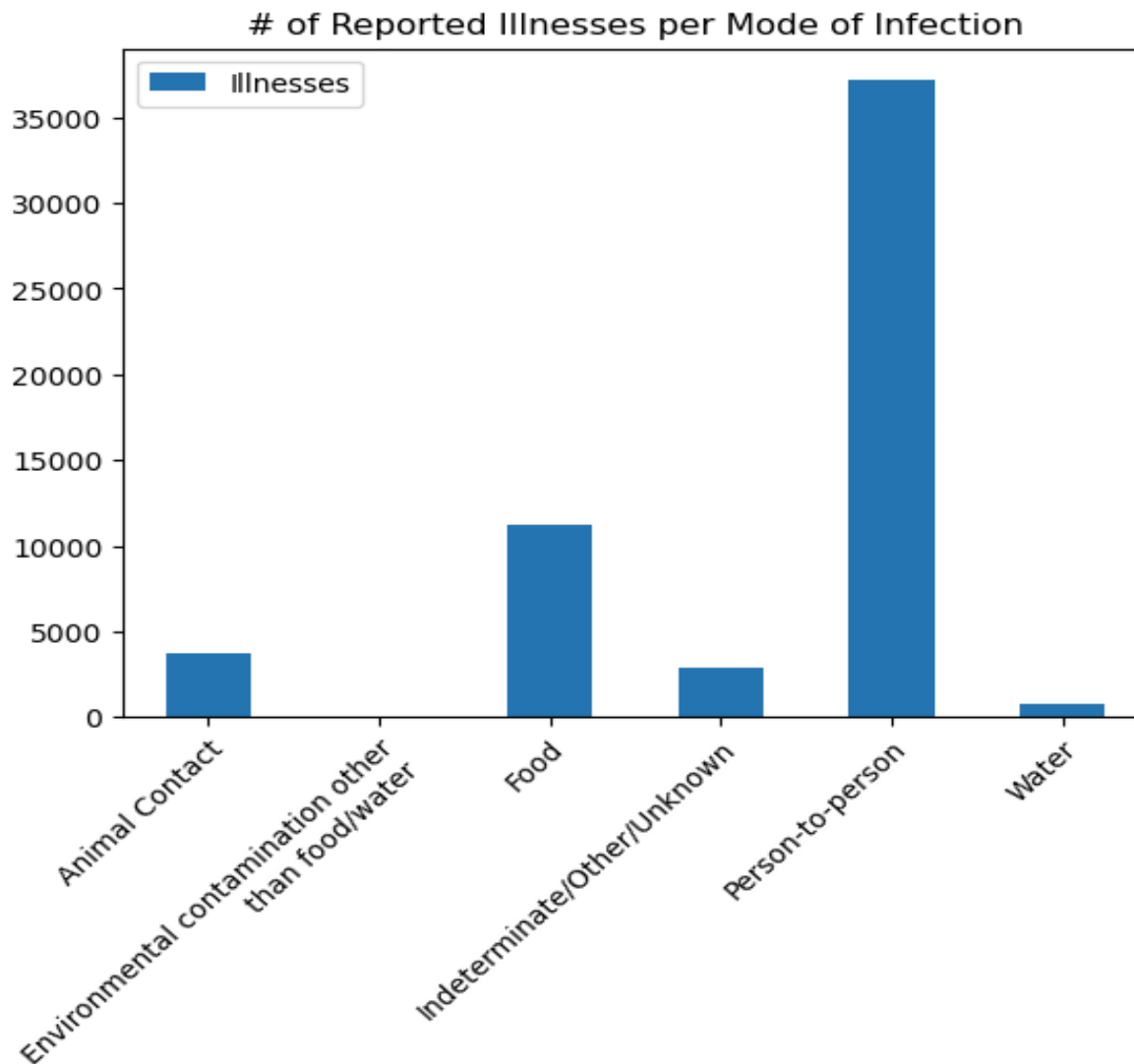
Data Overview

	Year	Month	State	Primary Mode	Etiology	Etiology Status	Setting	Illnesses	Hospitalizations	Info on Hospitalizations	Deaths	Info on Deaths
1	2020	1	Wisconsin	Person-to-person	Norovirus Genogroup II	Confirmed	Other healthcare facility	6	0.0	6.0	0.0	6
2	2020	1	Nebraska	Person-to-person	Norovirus Genogroup II	Confirmed	Long-term care/nursing home/assisted living fa...	20	1.0	20.0	0.0	20
3	2020	1	New York	Food	Clostridium perfringens	Suspected	Caterer (food prepared off-site from where ser...	7	0.0	7.0	0.0	7
4	2020	1	New Mexico	Person-to-person	Norovirus Genogroup II	Confirmed	Long-term care/nursing home/assisted living fa...	18	0.0	18.0	0.0	18

- National disease outbreak data from CDC
- Years 2020 - 2021
- Preprocessing of data



Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

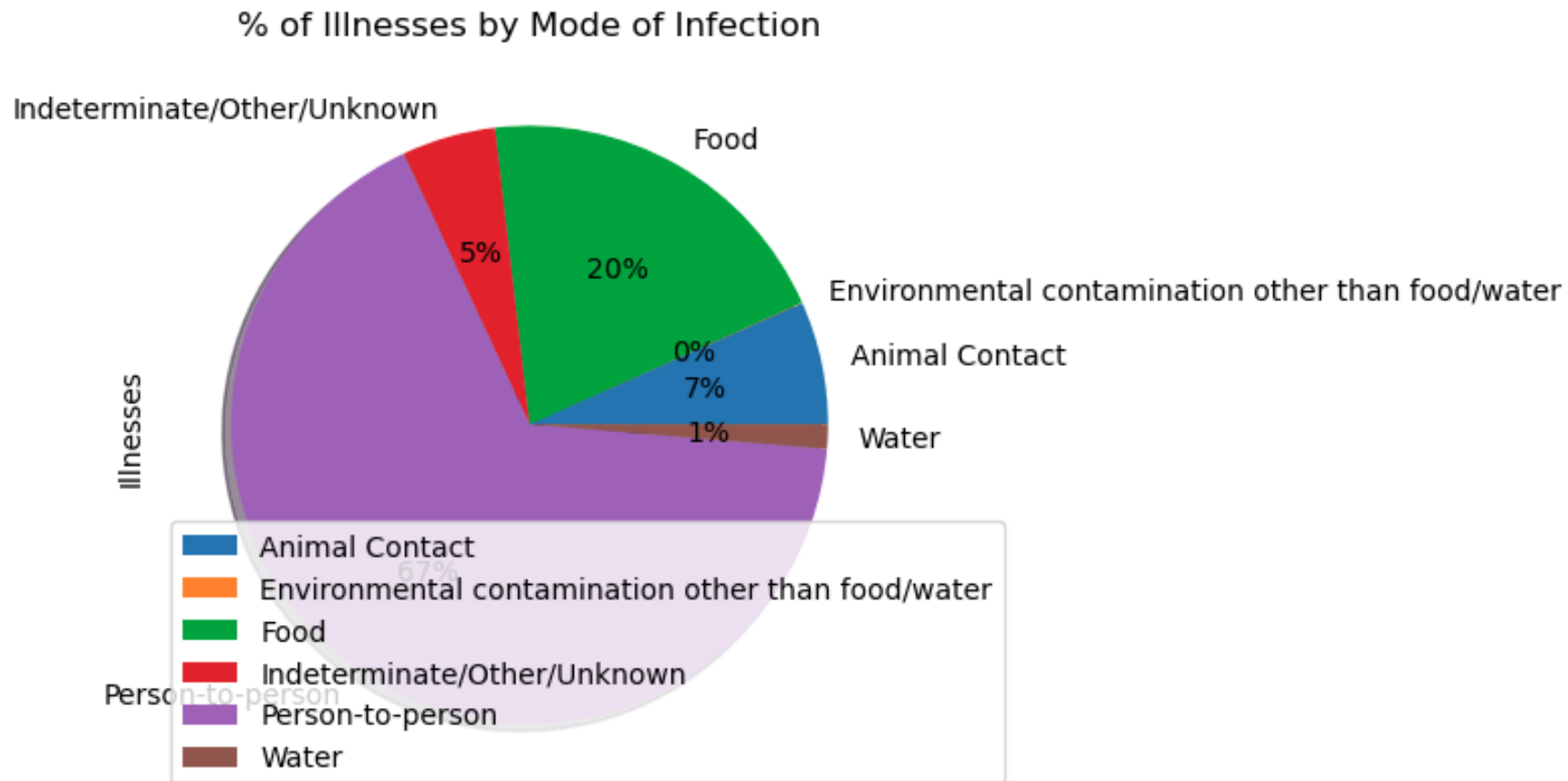


Illnesses	
Primary Mode	
Person-to-person	37147
Food	11223
Animal Contact	3703
Indeterminate/Other/Unknown	2829
Water	733
Environmental contamination other than food/water	19

What were the primary modes of outbreak by number of cases in the United States

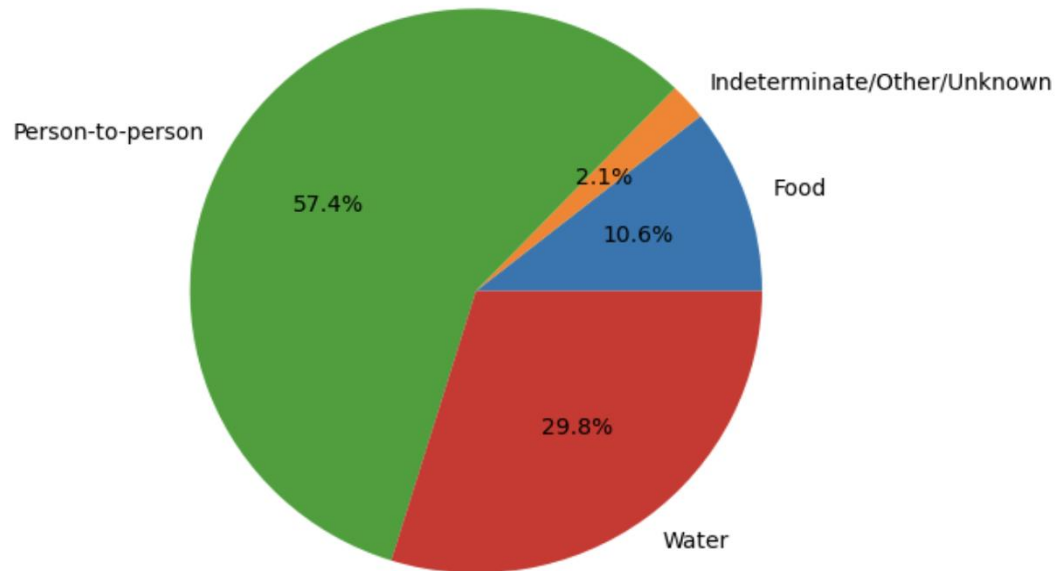
% of Illnesses by Mode of Outbreak (US overall)

Person-to-person contact is - by far - the largest mode of infection at 67%

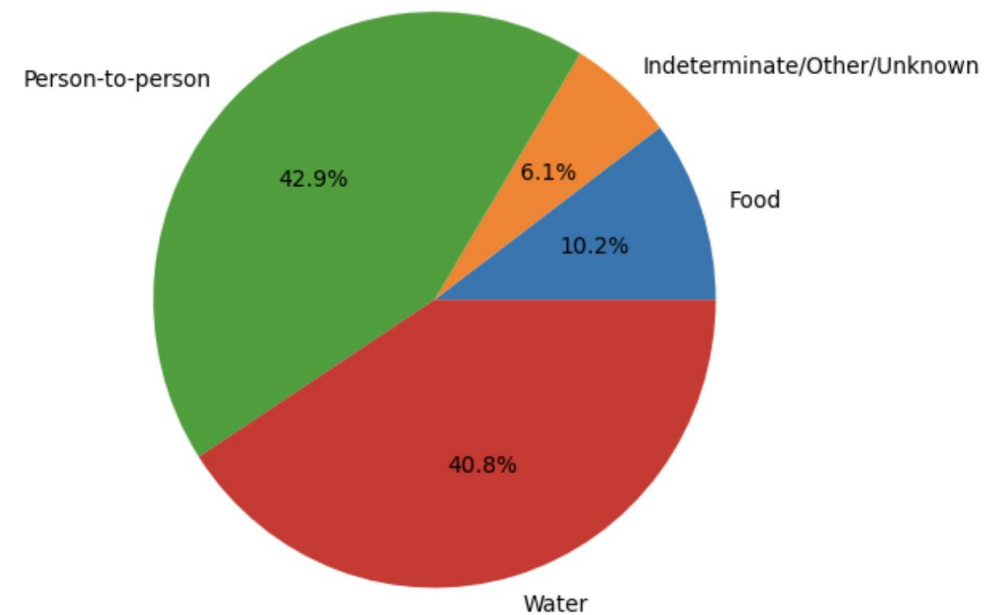


Findings on Primary Mode of Outbreak per State

Top Primary Mode by State in 2020



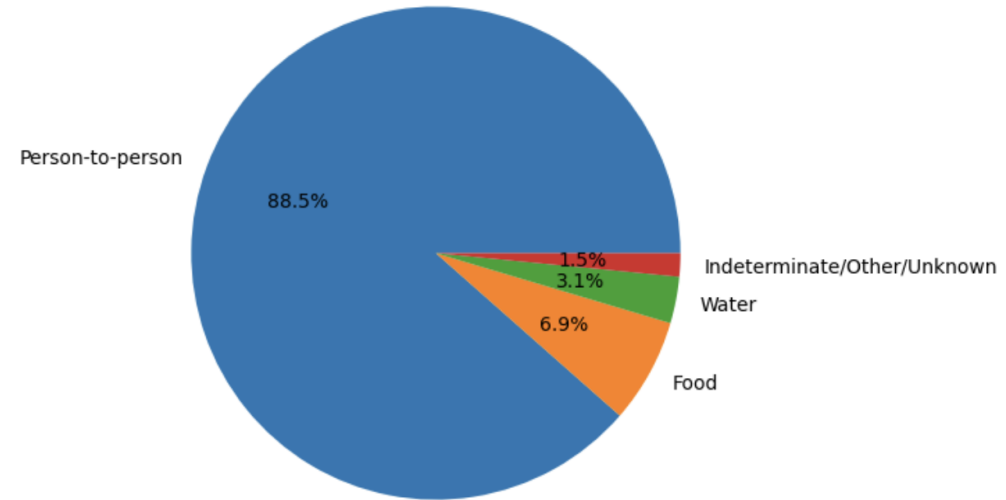
Top Primary Mode by State in 2021



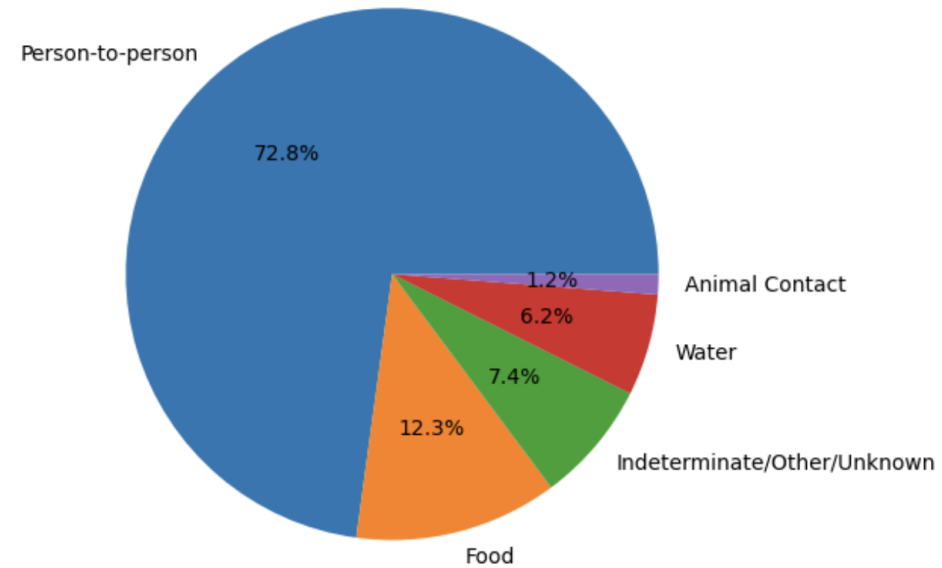
In 2020 and 2021 the top primary mode, per state, was person-to-person. In 2020 57.4% of the states had person-to-person as the top primary mode of outbreak. In 2021 42.9% of the states had person to person as their primary mode of outbreak .

Primary Modes in Wisconsin

Primary Modes in Wisconsin - 2020



Primary Modes in Wisconsin - 2021



Statistical Analysis

- Is there a difference in the number of person-to-person outbreaks per state in 2020 to 2021?
- P-value .88
- There isn't a significant difference between the years

```
In [160]: filtered_2021.head()
```

```
Out[160]:
```

	State	Primary Mode	Year
2	Alabama	Person-to-person	15
5	Arizona	Person-to-person	10
10	Colorado	Person-to-person	146
12	Connecticut	Person-to-person	6
20	Hawaii	Person-to-person	5

```
In [159]: filtered_2020 = data_2020[data_2020["Primary Mode"] == person]
filtered_2020.head()
```

```
Out[159]:
```

	State	Primary Mode	Year
1	Alabama	Person-to-person	2
4	Arizona	Person-to-person	2
9	Colorado	Person-to-person	7
11	Connecticut	Person-to-person	22
13	Florida	Person-to-person	1

```
In [162]: stats.ttest_ind(filtered_2020["Year"], filtered_2021["Year"])
```

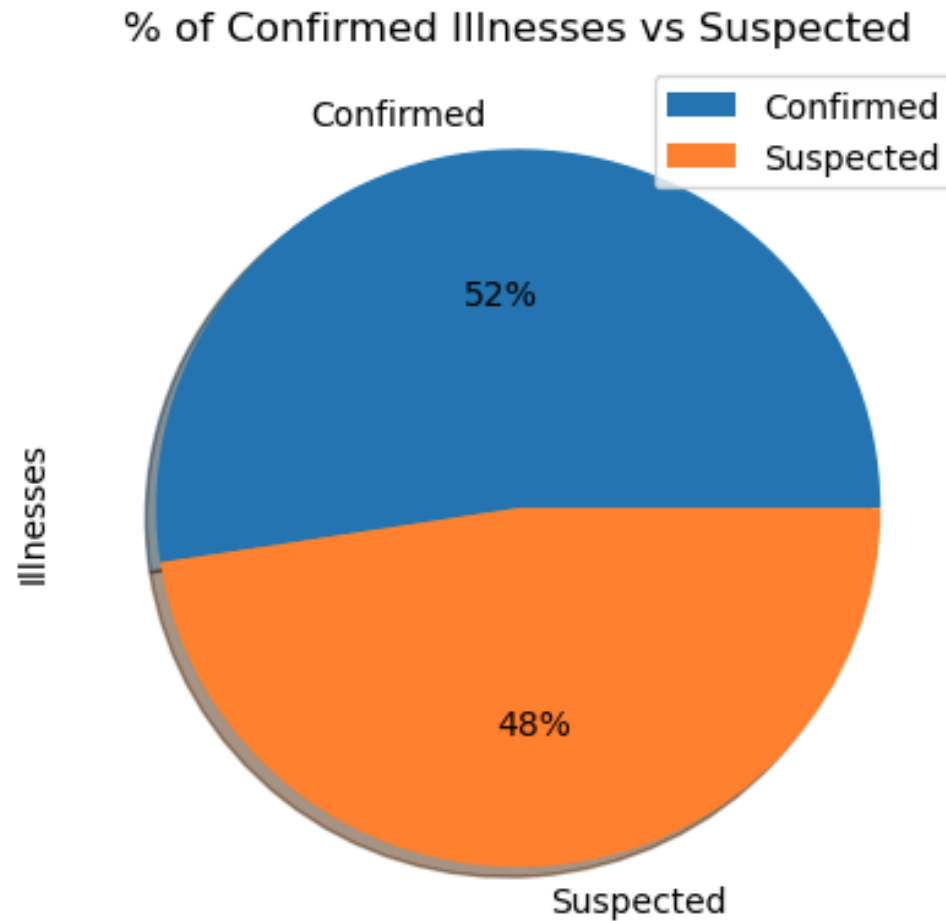
```
Out[162]: Ttest_indResult(statistic=-0.14934605539009238, pvalue=0.8816710497530617)
```

For both confirmed and suspected cases, what setting had the most outbreaks

The top 5 Settings by Illnesses (Case count)

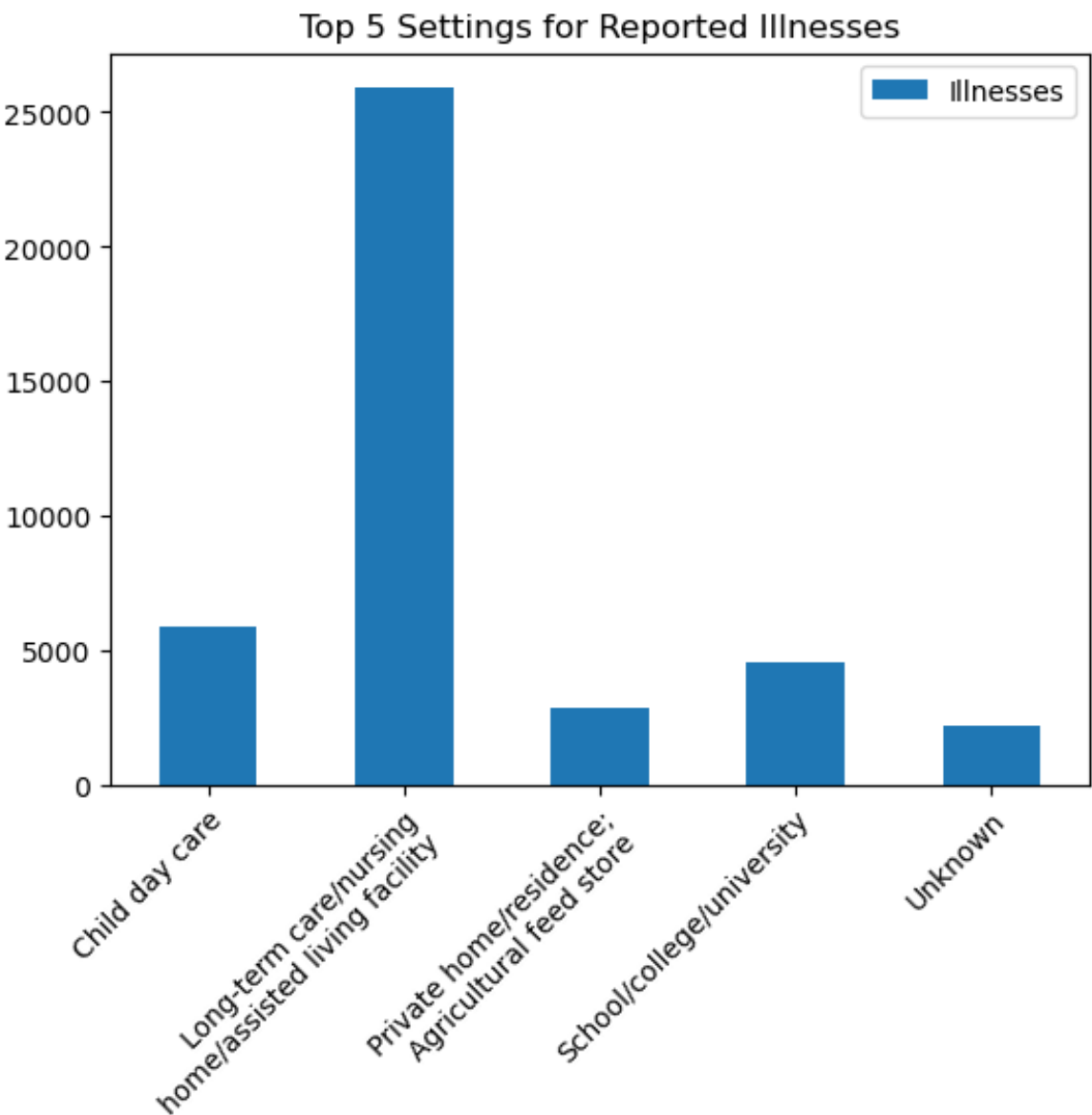
			Illnesses
Setting	Etiology Status		
Long-term care/nursing home/assisted living facility	Suspected		12867
	Confirmed		10928
Child day care	Suspected		4662
School/college/university	Suspected		3541
Private home/residence; Agricultural feed store	Confirmed		2859

% of Confirmed and Suspected cases



Etiology Status		Illnesses
0	Confirmed	26914
1	Suspected	24426

Settings by the numbers

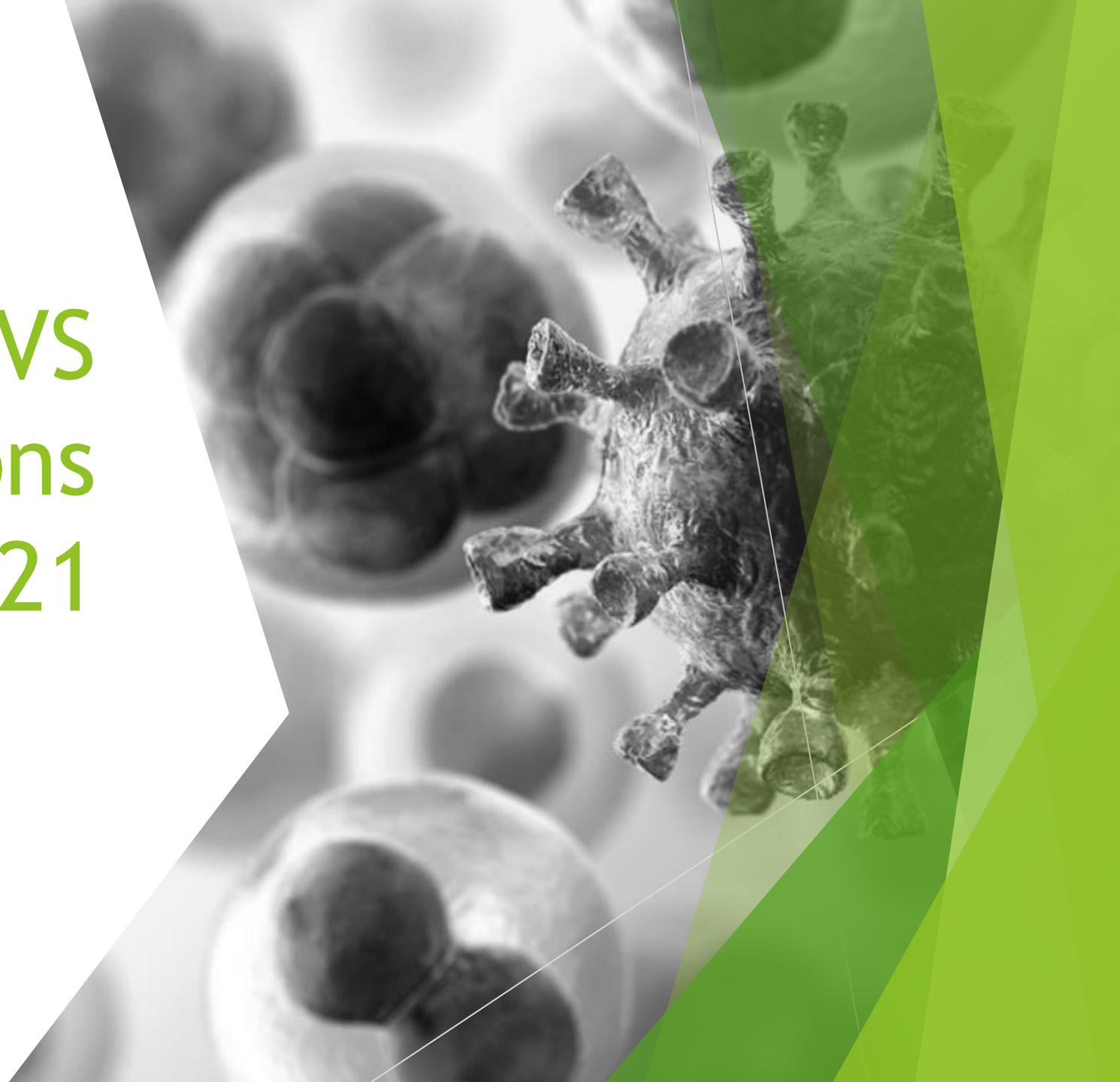


Viewing the top 5 Settings

Long-term care/assisted living facilities is the #1 setting for outbreaks

Illnesses	
Setting	
Long-term care/nursing home/assisted living facility	25890
Child day care	5882
School/college/university	4588
Private home/residence; Agricultural feed store	2859
Unknown	2242

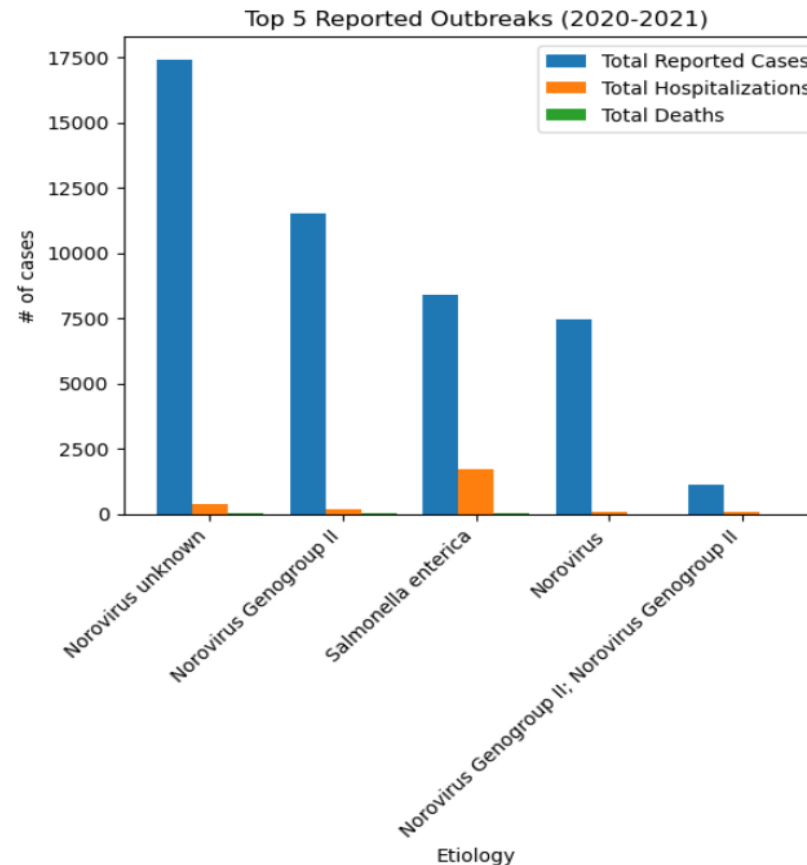
Deaths VS Hospitalizations in 2020-2021



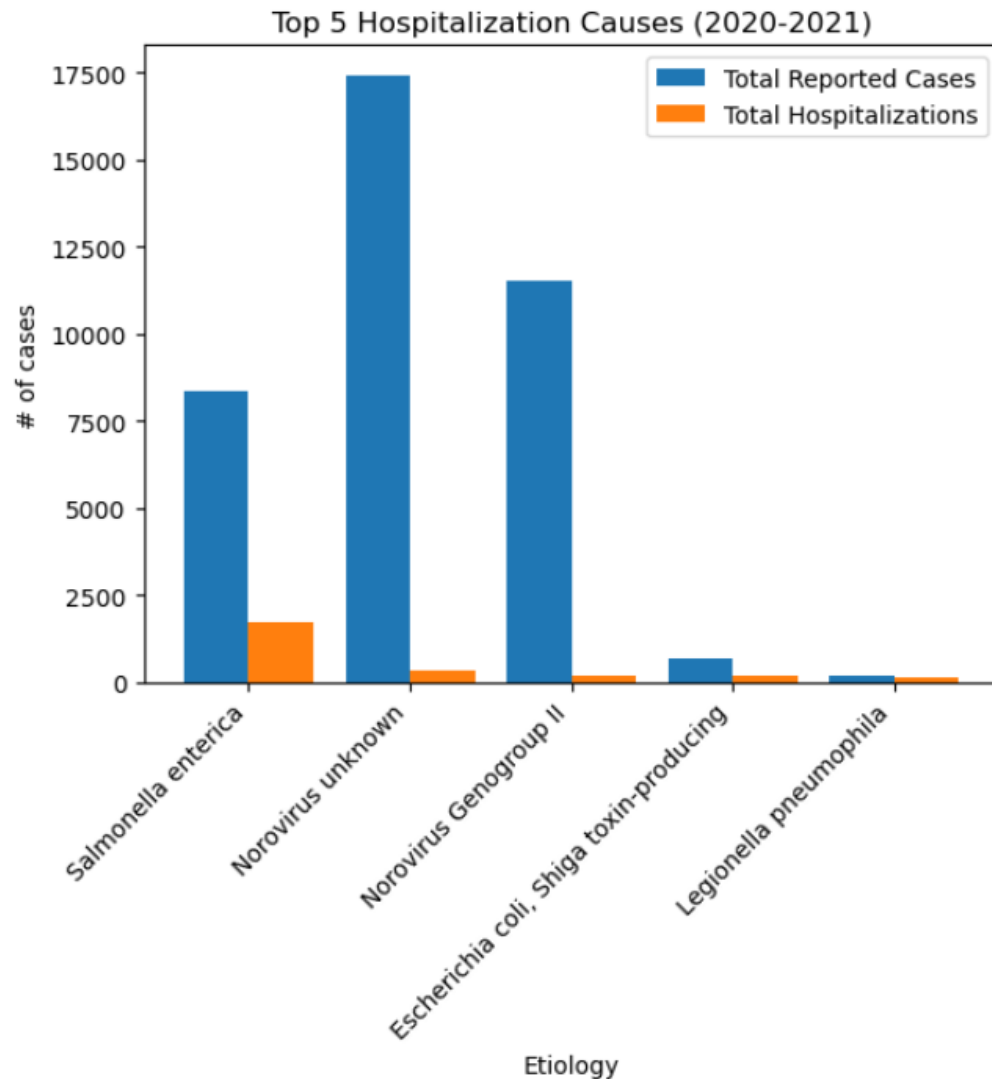
2020-2021 TOP 5 Reported Outbreaks by Case Count

► The highest reported number of cases for 2020-2021 was “Norovirus unknown” with 17,424 cases, 355 hospitalized cases, and 22 deaths. This was followed by “Norovirus Genogroup II” with 11,536 cases, 189 hospitalized, and 19 deaths.

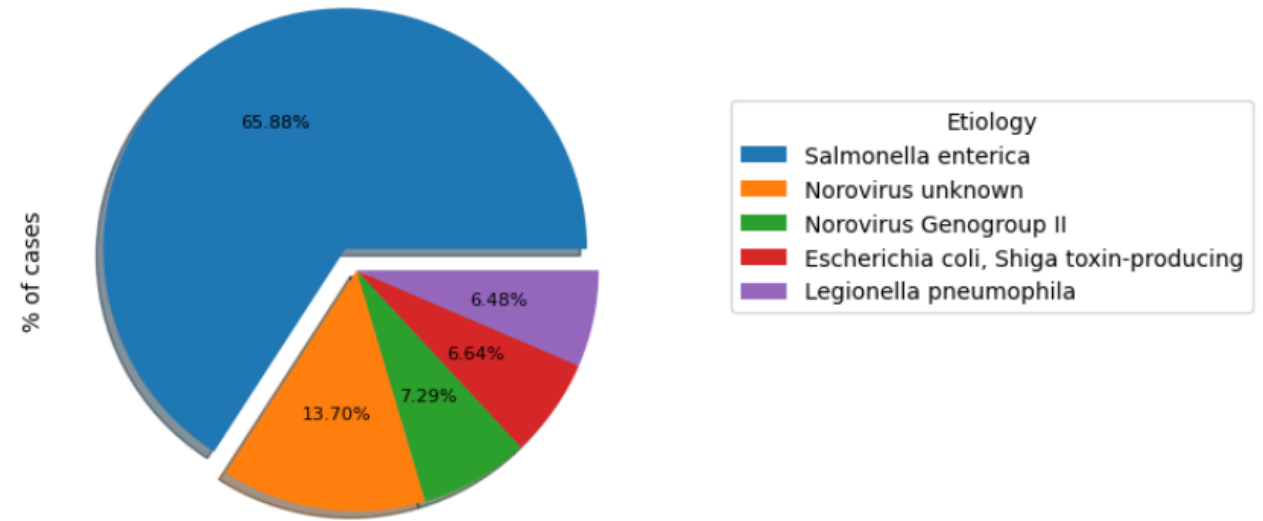
Etiology	Total Reported Cases	Total Hospitalizations	Total Deaths	Hospitalization %	Death %	% Death to Hospitalization
Norovirus unknown	17424	355.0	22.0	2.037420	0.126263	6.197183
Norovirus Genogroup II	11536	189.0	19.0	1.638350	0.164702	10.052910
Salmonella enterica	8382	1707.0	11.0	20.365068	0.131234	0.644405
Norovirus	7458	96.0	1.0	1.287208	0.013408	1.041667
Norovirus Genogroup II; Norovirus Genogroup II	1141	65.0	0.0	5.696757	0.000000	0.000000



2020-2021 Top 5 Outbreaks That Caused the Most Hospitalizations



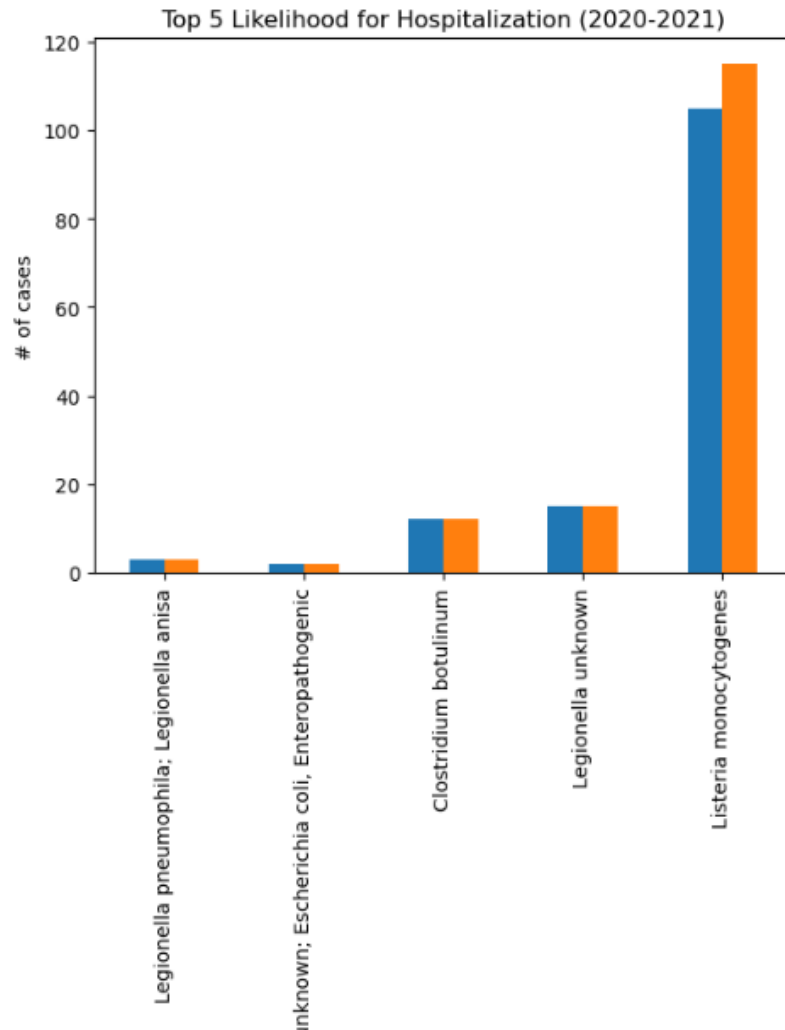
Top 5 Hospitalization Causes (2020-2021)



Etiology	Total Reported Cases	Total Hospitalizations	Total Deaths	Hospitalization %	Death %	% Death to Hospitalization
Salmonella enterica	8382	1707.0	11.0	20.365068	0.131234	0.644405
Norovirus unknown	17424	355.0	22.0	2.037420	0.126263	6.197183
Norovirus Genogroup II	11536	189.0	19.0	1.638350	0.164702	10.052910
Escherichia coli, Shiga toxin-producing	704	172.0	3.0	24.431818	0.426136	1.744186
Legionella pneumophila	210	168.0	14.0	80.000000	6.666667	8.333333

The highest outbreak that had the most hospitalized cases was “Salmonella enterica” with 1,707 cases out of 8,282 reported cases, yielding 11 deaths total in 2020-2021.

2020-2021 Top 5 Outbreaks That Have a High Rate of Hospitalization



Etiology

- Total Hospitalizations
- Total Reported Cases

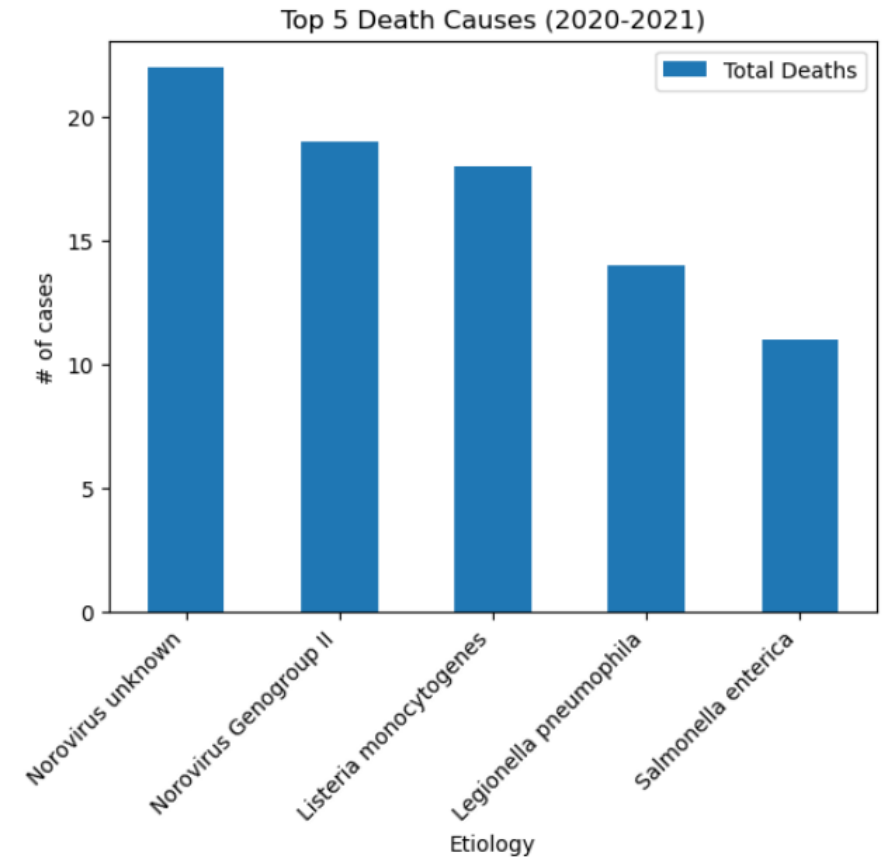
Etiology	Total Reported Cases	Total Hospitalizations	Total Deaths	Hospitalization %	Death %	% Death to Hospitalization
Legionella pneumophila; Legionella anisa	3	3.0	1.0	100.000000	33.333333	33.333333
Campylobacter coli; Escherichia coli, Enterotoxigenic; Salmonella enterica; Cryptosporidium unknown; Escherichia coli, Enteropathogenic	2	2.0	0.0	100.000000	0.000000	0.000000
Clostridium botulinum	12	12.0	2.0	100.000000	16.666667	16.666667
Legionella unknown	15	15.0	2.0	100.000000	13.333333	13.333333
Listeria monocytogenes	115	105.0	18.0	91.304348	15.652174	17.142857

There was a 4-way tie of outbreaks that caused 100% hospitalization; “Legionella pneumophila; Legionella anisa”, “Campylobacter coli; Escherichia coli, Enterotoxigenic; Salmonella enterica; Cryptosporidium unknown; Escherichia coli, Enteropathogenic”, “Clostridium botulinum”, and “Legionella unknown”. Followed by “Listeria monocytogenes” with 91.30% likelihood.

2020-2021 Top 5 Outbreaks that caused the most Deaths

Etiology	Total Reported Cases	Total Hospitalizations	Total Deaths	Hospitalization %	Death %	% Death to Hospitalization
Norovirus unknown	17424	355.0	22.0	2.037420	0.126263	6.197183
Norovirus Genogroup II	11536	189.0	19.0	1.638350	0.164702	10.052910
Listeria monocytogenes	115	105.0	18.0	91.304348	15.652174	17.142857
Legionella pneumophila	210	168.0	14.0	80.000000	6.666667	8.333333
Salmonella enterica	8382	1707.0	11.0	20.365068	0.131234	0.644405

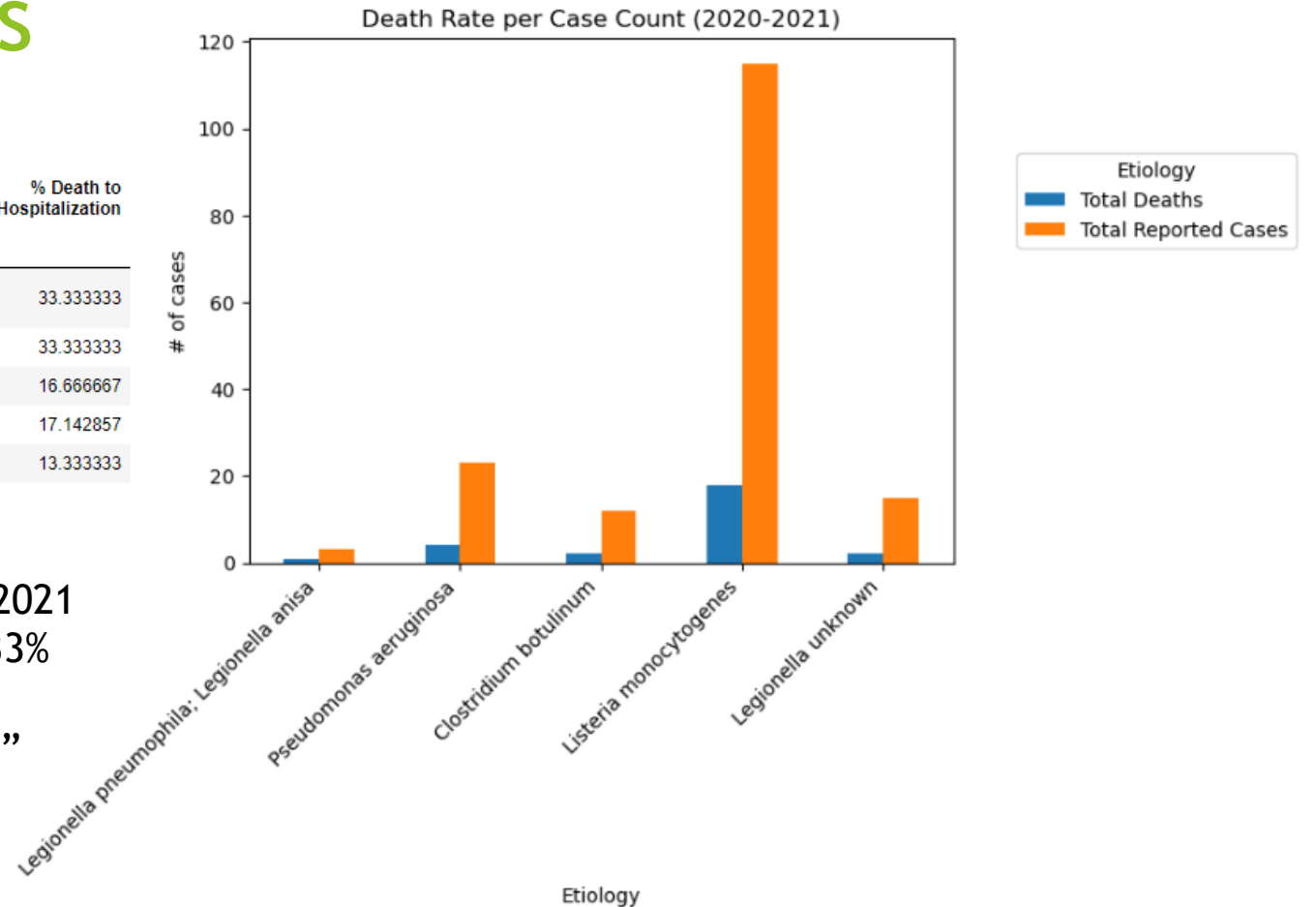
The outbreak that caused the most deaths in 2020-2021, was “Norovirus unknown” with 22 total deaths, 17,424 reported cases, and 355 hospitalizations. Followed by “Norovirus Genogroup II” with 19 deaths, 11,536 reported cases, and 189 hospitalizations. Respectively, they had 0.13% and 0.16% death rate.



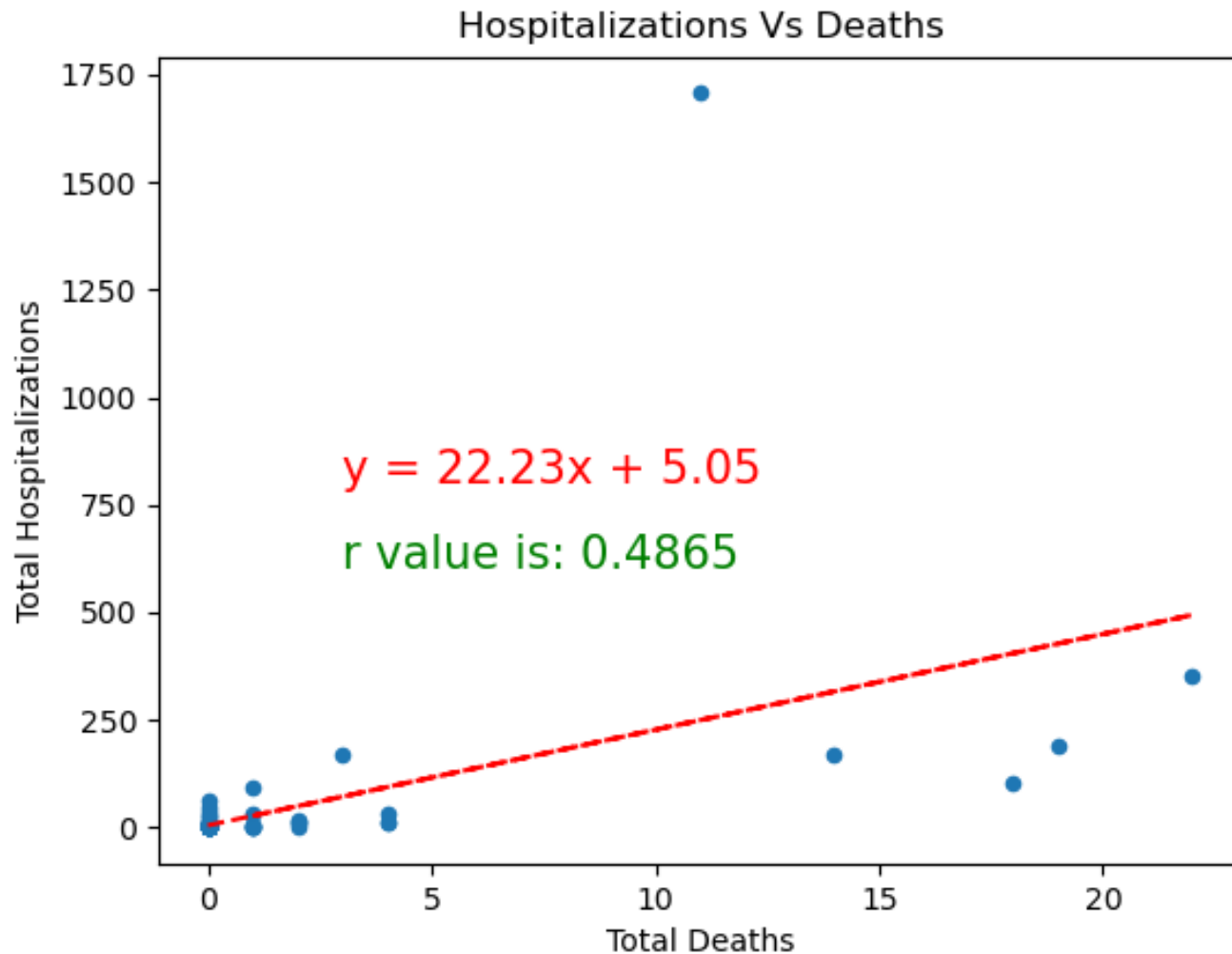
2020-2021 Top 5 Outbreaks That Had The Highest Death Rates

Etiology	Total Reported Cases	Total Hospitalizations	Total Deaths	Hospitalization %	Death %	% Death to Hospitalization
Legionella pneumophila; Legionella anisa	3	3.0	1.0	100.000000	33.333333	33.333333
Pseudomonas aeruginosa	23	12.0	4.0	52.173913	17.391304	33.333333
Clostridium botulinum	12	12.0	2.0	100.000000	16.666667	16.666667
Listeria monocytogenes	115	105.0	18.0	91.304348	15.652174	17.142857
Legionella unknown	15	15.0	2.0	100.000000	13.333333	13.333333

The outbreak that had the highest death rate in 2020-2021 was “Legionella pneumophila; Legionella anisa” with 33% death rate with 1 death for the 3 reported cases--3 hospitalization. Followed by “Pseudomonas aeruginosa” with 17.39% death rate with 4 death out of the 23 reported cases-12 hospitalizations.



Hospitalization VS Death Correlation



The linear regression equation from the data in the data set yielded $y=22.23x + 5.05$ with an r value of .4865. With this, we can say that the number of deaths has a weak correlation with the number of hospitalizations.



Final Data Analysis

- Person-to-person is the primary mode for outbreaks in the U.S. with over 37,000 reported cases in 2020-2021.
- Looking at the primary mode for outbreaks per state, over 50% of states had person-to-person as the highest cause for outbreaks in 2020-2021; 57.4% in 2020, and 47.9% in 2021.
- Looking specifically at Wisconsin, person-to-person was the primary mode of infection in 2020-2021—supporting the above conclusions.
- Statistically speaking, the year 2020 and 2021 did not have a significant difference between the primary mode of outbreaks with a P-value of 0.88.
- Long-term care/nursing home/assisted living facilities, as a setting, had the most confirmed and suspected outbreaks in 2020-2021.
- There was an even distribution of confirmed and suspected cases in 2020-2021—52% to 48% respectively.



Final Data Analysis (Continued)

- ▶ The highest reported number of cases for 2020-2021 was “Norovirus unknown” with 17,424 cases, 355 hospitalized cases, and 22 deaths. This was followed by “Norovirus Genogroup II” with 11,536 cases, 189 hospitalized, and 19 deaths.
- ▶ The highest outbreak that had the most hospitalized cases was “Salmonella enterica” with 1,707 cases out of 8,282 reported cases, yielding 11 deaths total in 2020-2021.
- ▶ There was a 4-way tie of outbreaks that caused 100% hospitalization; “Legionella pneumophila; Legionella anisa”, “Campylobacter coli; Escherichia coli, Enterotoxigenic; Salmonella enterica; Cryptosporidium unknown; Escherichia coli, Enteropathogenic”, “Clostridium botulinum”, and “Legionella unknown”. Followed by “Listeria monocytogenes” with 91.30% hospitalization rate.
- ▶ The outbreak that caused the most deaths in 2020-2021, was “Norovirus unknown” with 22 total deaths, 17,424 reported cases, and 355 hospitalizations. Followed by “Norovirus Genogroup II” with 19 deaths, 11,536 reported cases, and 189 hospitalizations. Respectively, they had 0.13% and 0.16% death rate.



Final Data Analysis (Continued)

- ▶ The outbreak that had the highest death rate in 2020-2021 was “*Legionella pneumophila*; *Legionella anisa*” with 33% death rate with 1 death for the 3 reported cases--3 hospitalization. Followed by “*Pseudomonas aeruginosa*” with 17.39% death rate with 4 death out of the 23 reported cases-12 hospitalizations.
- ▶ The number of deaths had a weak correlation with the number of hospitalizations—with an r-value of .04865

Additional Analysis and Findings

- Multistate (unknown grouped states) had the most cases (7,758), hospitalizations(1752), and deaths(31). This was followed closely by Wisconsin with 4,377 cases, 140 hospitalizations and 12 deaths.
- The least amount of reported cases of outbreaks in 2020-2021 per state, was Mississippi, with 10 reported cases, 0 hospitalizations, and 0 deaths.
- 17 states reported having "Norovirus unknown" as the leading cause for outbreaks, followed by "Norovirus Genogroup II" with 15 states.