

# Using Natural Language Processing to increase identification of child maltreatment in electronic health records

Frances Lynch, PhD, MSPH

Center for Health Research, Kaiser Permanente Northwest & OCHIN Research

April 13, 2022, HCSRN Annual Meeting Pasadena, CA

\*\*Funded by the National Institute of Health MHRN Pilot Fund

# Study Team

- Sonya Negriff, PhD,  
Kaiser Permanente Southern California
- Roy Pardee, JD, MA, Kaiser Permanente  
Washington
- David Cronkite, PhD, Kaiser Permanente  
Washington
- Rob Penfold, PhD, Kaiser Permanente  
Washington

# Short and Long-Term Impact of Child Maltreatment

- Consistent links between child maltreatment and mental health symptoms in childhood and adulthood
- Formerly maltreated individuals report higher rates of physical health problems such as
  - obesity (childhood & adulthood)
  - cardiovascular disease
  - cancer
- The Centers for Disease Control & Prevention estimate total lifetime economic burden from child maltreatment in the U.S. at **\$585 billion**
  - Per victim cost of \$830,928

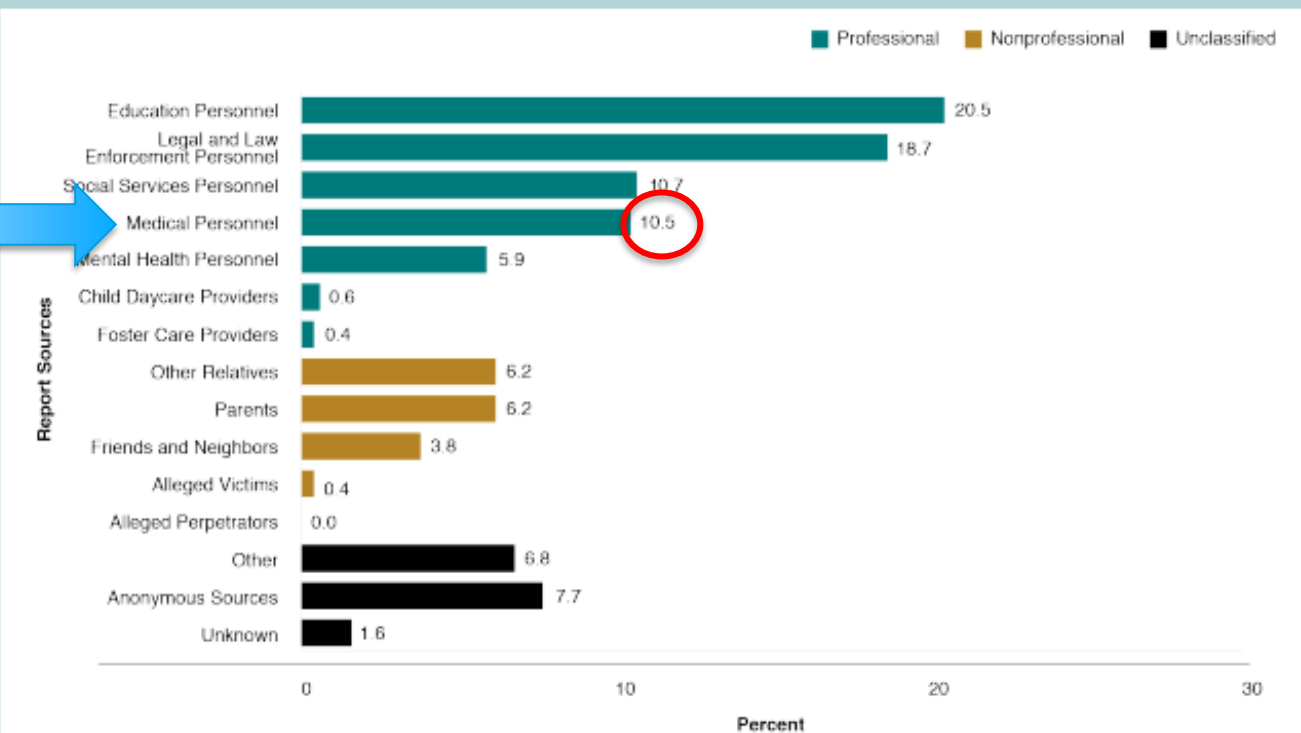
# The Prevalence of Child Maltreatment

- In 2018, 674,000 children were confirmed victims of maltreatment (child welfare records)
  - 1,770 fatalities
  - 74.9% were neglected
- Cumulative prevalence rates
  - 12.5% of US children experience maltreatment by their 18<sup>th</sup> birthday
  - 5.8% have confirmed maltreatment by age 5

- Child maltreatment is a critical public health issue and health care systems play an important role in identifying and treating children who experience maltreatment

### Exhibit 2–D Report Sources, 2018

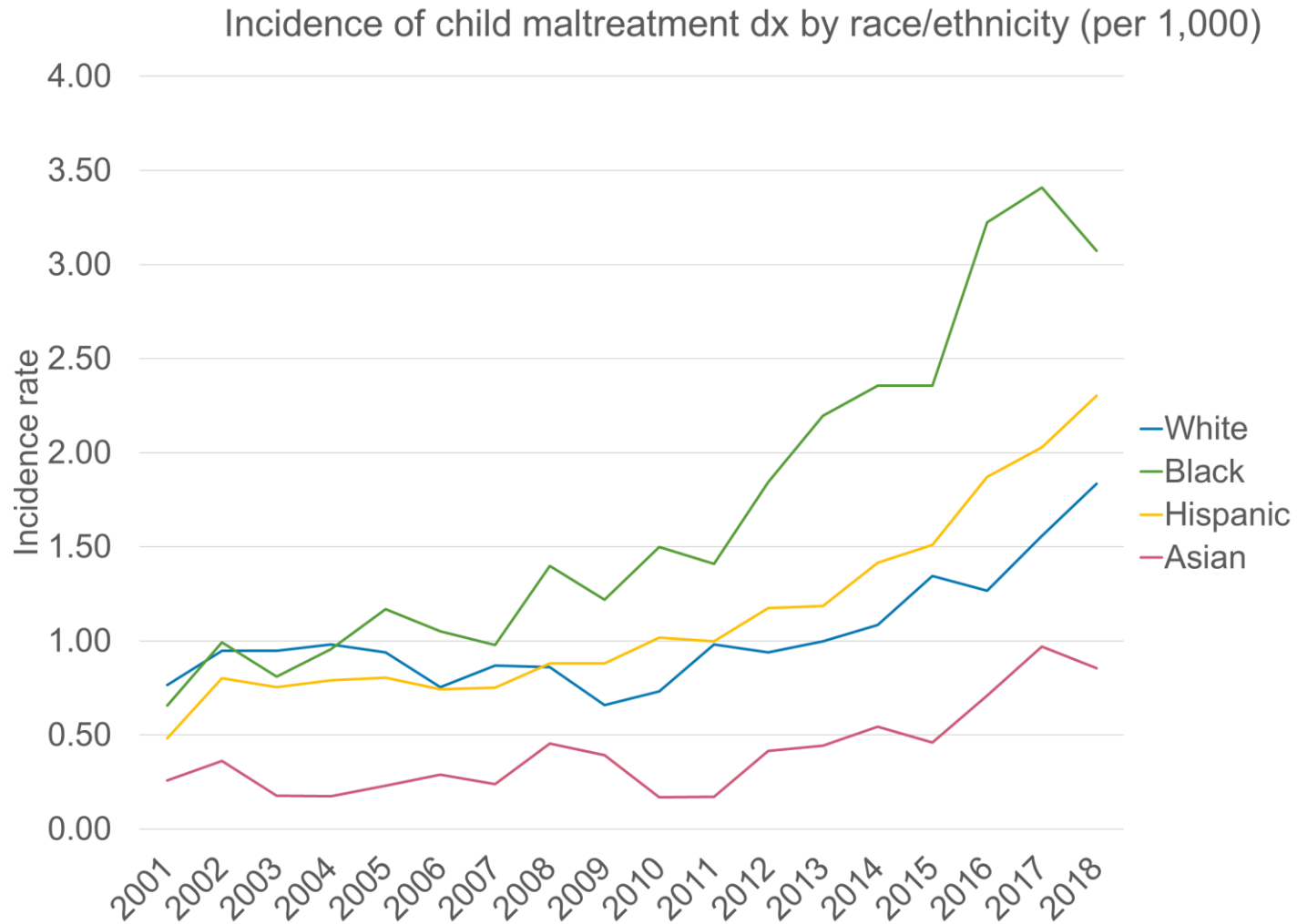
Professionals submit the majority of screened-in referrals (reports) that received an investigation or alternative response



Based on data from 49 states. Data are from the Child File. States were excluded from this analysis if more than 25.0 percent had an "other" or unknown report source. Supporting data not shown.

- Few studies of child maltreatment have used data from large health systems to understand how systems identify and manage youth who experience maltreatment
- Studies of incidence rates of maltreatment in EHR as well as the most recent MHRN quarterly descriptive analyses (2018) indicate *a significant under-reporting of child maltreatment in health systems*
  - MHRN data show average population rate of 0.12% to 0.58%

# Incidence of Child Maltreatment in KPSC



- **Reasons for potential under-reporting:**
  - Concerns about lack of resources to deal with probable maltreatment
  - Lack of training on suspicious injuries
  - “grey zone” of probable maltreatment (may be discussed in chart notes but not documented using ICD codes)
  - Provider may be concerned with referral to CPS rather than going back to document in the chart
  - Concerns that labeling as “maltreatment” will do more harm than good

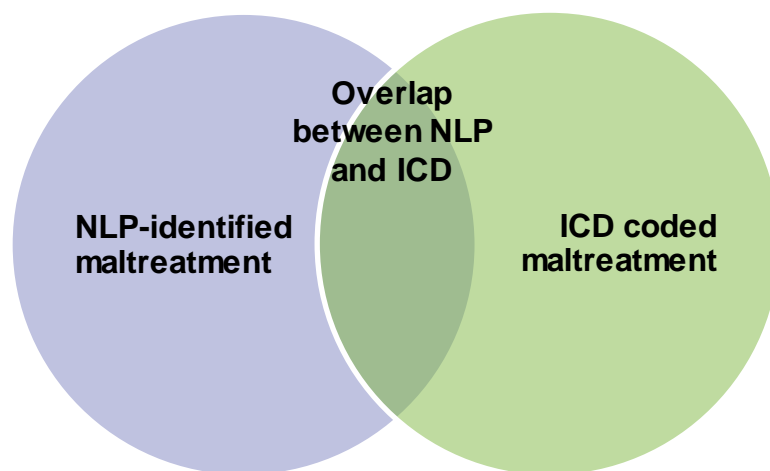


- It is critical to identify children who have been maltreated as they are at higher risk for multiple negative lifetime outcomes
- Consistent documentation could help medical providers identify children at risk, facilitate referrals, and prevent future abuse
- More consistent reporting could improve equity of identification and referrals of youth across racial/ethnic

- Natural Language Processing (NLP) may help to identify additional youth with maltreatment
- If NLP could identify cases that are not documented through ICD codes, this could indicate the need for health system efforts to develop new ways of consistently document child maltreatment
- NLP might also help to identify any groups (e.g., age, gender, race/ethnicity) that may be more/less likely to have ICD coded child maltreatment

# Research Questions

- Does NLP allow us to obtain estimates the number of children who experience maltreatment more comparable to national epidemiologic data?
- Does NLP of chart notes identify new cases of child maltreatment not already documented with ICD codes?
- What is the overlap between the two methods?
- Are there differences by age group, gender, or race-ethnicity?



# Methods

- Identified cohort a of children aged 0-18 at KPWA between 2016-2020
- Developed Bag of Words → concept unique identifiers (CUI)
- Manual review of text snips for each CUI, flag confirmed cases
- Retrained NLP and each note flagged as Y/N

# NLP Results

| Number and rate of notes with reference of child maltreatment |          |         |            |
|---|----------|---------|------------|
| Cohort  | Cohort n | NLP + n | Crude_Rate |
| 2016  | 98299    | 1526    | 1.55%      |
| 2017  | 86824    | 1746    | 2.01%      |
| 2018  | 94289    | 1853    | 1.97%      |
| 2019  | 99540    | 2353    | 2.36%      |
| 2020  | 96213    | 2183    | 2.27%      |
|   |          | 9661    |            |

some youth in more than one cohort  
7442 unique youth

# NLP Sample Description

| AGE       |           |         |
|-----------|-----------|---------|
| Age group | Frequency | Percent |
| <1        | 577       | 12.47   |
| 1-4       | 928       | 44.52   |
| 5-12      | 2624      | 35.26   |
| 13-18     | 3313      | 7.75    |

| SEX    |           |         |
|--------|-----------|---------|
|        | Frequency | Percent |
| Female | 4485      | 60.27   |
| Male   | 2956      | 39.72   |

| RACE            |           |         |
|-----------------|-----------|---------|
|                 | Frequency | Percent |
| Asian           | 592       | 7.95    |
| Black           | 892       | 11.99   |
| Hispanic        | 177       | 2.38    |
| American Indian | 255       | 3.43    |
| Other           | 322       | 4.33    |
| Unknown         | 550       | 7.39    |
| White           | 4654      | 62.54   |

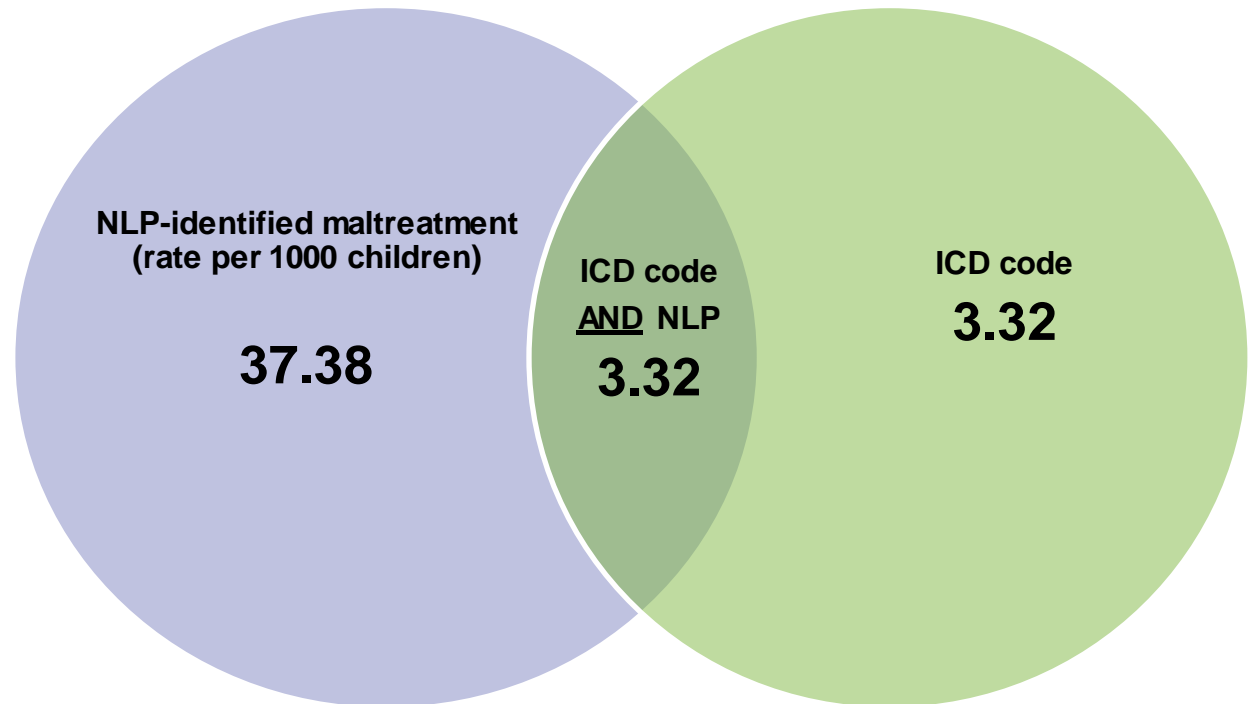
| Hispanic ethnicity |           |         |
|--------------------|-----------|---------|
|                    | Frequency | Percent |
| Yes                | 765       | 10.28   |
| No                 | 6044      | 81.21   |
| Unknown            | 633       | 8.51    |

| Commercial insurance |           |         |
|----------------------|-----------|---------|
|                      | Frequency | Percent |
| Yes                  | 4072      | 64.83   |
| No                   | 416       | 6.62    |
| Unknown              | 1793      | 28.55   |

| Medicaid coverage |           |         |
|-------------------|-----------|---------|
|                   | Frequency | Percent |
| No                | 4439      | 70.67   |
| Yes               | 1842      | 29.33   |

# ICD Dx code versus NLP identification

- CM ICD dx code
  - 3.38 per 1000
- NLP CM note
  - 37.38 per 1000



- 100% of children with a maltreatment dx code also had positive note via NLP

# ICD Dx code versus NLP identification

- Identification of CM using NLP increased most for:
  - ages 13-18
  - Females
  - Native Am. & Black youth
  - Medicaid members

|                           | All children<br>N | ICD code<br>per 1000 | NLP<br>per 1000 |
|---------------------------|-------------------|----------------------|-----------------|
| <b>Age Group</b>          |                   |                      |                 |
| <1                        | 21241             | 1.98                 | 27.16           |
| 1-4                       | 33524             | 3.01                 | 27.68           |
| 5-12                      | 70292             | 4.03                 | 37.33           |
| 13-18                     | 74028             | <b>3.16</b>          | <b>44.75</b>    |
| <b>Sex</b>                |                   |                      |                 |
| Female                    | 98279             | <b>4.49</b>          | <b>45.64</b>    |
| Male                      | 100792            | 2.16                 | 29.33           |
| <b>Race</b>               |                   |                      |                 |
| Asian                     | 20101             | 2.24                 | 29.45           |
| Black                     | 13481             | <b>6.45</b>          | <b>66.17</b>    |
| Hispanic                  | 3603              | 4.72                 | 49.13           |
| Am. Ind., AK Nat.         | 2791              | <b>9.67</b>          | <b>91.37</b>    |
| Other                     | 6852              | 6.13                 | 46.99           |
| Unknown                   | 69900             | 0.80                 | 7.87            |
| White                     | 82357             | 4.69                 | 56.51           |
| <b>Hispanic Ethnicity</b> |                   |                      |                 |
| Hispanic Ethnicity        | 13140             | 6.39                 | 58.22           |
| <b>Insurance</b>          |                   |                      |                 |
| Commercial                | 115587            | 2.81                 | 35.23           |
| Medicaid                  | 22933             | <b>8.07</b>          | <b>80.32</b>    |
| Other                     | 60565             | 2.48                 | 25.23           |
| <b>Total</b>              |                   |                      |                 |
| Total                     | 199085            | 3.32                 | 37.38           |



# Limitations

- Proof of concept using small pilot funding from NIMH funded MHRN
- Very basic NLP analysis
- Validation was limited, not full chart review
- Unable to look at maltreatment types
- Unclear whether the lack of dx code was due to type of injury or abuse that did not meet the criteria for mandated report (i.e., historical account of abuse)

# Conclusions & Future Directions

- Using ***NLP identified 10 times more children*** with child maltreatment than just using the diagnosis code
- Teenagers, girls, Native American, Black youth, and those with Medicaid insurance had the most increase in the identification when using NLP
  - This may be due to differences in reporting and/or documentation
- We don't know if children with maltreatment identified in progress notes (but no ICD code) received assessment or treatment related to their experiences
  - This has implications for physical and mental health