

# Household transmission

November 22, 2021

# Table of Contents

- A. Steps in creating HH dataset and analysis variable
- B. Issues and Limitations
- C. Findings

## A. Steps in creating HH dataset and analysis variable



# Steps

1. Access and curate dphe144 data CEDRS\_view
2. Fix selected data errors to create CEDRS\_fix
3. Filter cases
  - a. Reported Date in Sept 1 - Nov 1 for 2020 and 2021
  - b. County NOT "INTERNATIONAL"
  - c. Live in Institution NOT "YES"
4. Define Household based on
  - a. Street address
  - b. City
  - c. County
5. Must also have State to filter cases to CO

# Completeness of address data

address1	address_city	address_state	address_zipcode	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing data	Missing data	Missing data	Missing data	263	0.14	263	0.14
Missing data	Missing data	Missing data	Has data	5	0.00	268	0.15
Missing data	Missing data	Has data	Missing data	131	0.07	399	0.22
Missing data	Missing data	Has data	Has data	8	0.00	407	0.22
Missing data	Has data	Missing data	Missing data	1	0.00	408	0.22
Missing data	Has data	Missing data	Has data	7	0.00	415	0.23
Missing data	Has data	Has data	Missing data	61	0.03	476	0.26
Missing data	Has data	Has data	Has data	300	0.17	776	0.43
Has data	Missing data	Missing data	Missing data	6	0.00	782	0.43
Has data	Missing data	Missing data	Has data	1	0.00	783	0.43
Has data	Missing data	Has data	Missing data	6	0.00	789	0.43
Has data	Missing data	Has data	Has data	42	0.02	831	0.46
Has data	Has data	Missing data	Missing data	7	0.00	838	0.46
Has data	Has data	Missing data	Has data	2702	1.49	3540	1.95
Has data	Has data	Has data	Missing data	171	0.09	3711	2.04
Has data	Has data	Has data	Has data	178052	97.96	181763	100.00

# Steps, continued

6. Filter cases with complete address (and Age data):

- a. State=CO
- b. Address1, City, and County NOT missing
- c. Age at Reported NOT missing

7. Define Eligible Households

- a. Same street address (for given County-City)
- b. Number of cases in HH restricted to 2-10

8. Define Clusters in Households

- a. Time between case and previous case NOT >30 days

# Number of filtered cases per eligible HH

COUNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	17570	71.66	17570	71.66
3	4503	18.37	22073	90.02
4	1483	6.05	23556	96.07
5	537	2.19	24093	98.26
6	224	0.91	24317	99.18
7	94	0.38	24411	99.56
8	50	0.20	24461	99.76
9	29	0.12	24490	99.88
10	29	0.12	24519	100.00

# Steps, continued

## 9. Create analysis variable based on Age Groups

I or i = “infant” (0-4 years old)

K or k = kid (5-11 years old)

T or t = Teen (12-17 years old)

A or a = Adult (18-115 years old)

Lower case for Fall 2020 and upper case for Fall 2021



dress1

514 E 18TH PL  
515 EAST 18TH PLACE  
54 IOLA ST  
54 IOLA ST

56  
56 CIMARRON ST  
56 FLORENCE ST  
56 FLORENCE ST

593 E 17TH PL  
64 CHESTER ST  
64 CHESTER ST  
64 FLORENCE ST

67 KENTON ST  
76 FRASER CT  
76 FRASER CT  
82 KINGSTON ST

91 BEELER ST  
91 BEELER ST  
07 PYNE WAY LOT 94  
25 ESPANA WAY

25 ESPANA WAY  
68 CIMARRON  
68 CIMARRON ST  
70 CARSON ST

70 CARSON ST  
829 E CARMEL CIR  
903 E 16TH AVE  
903 E 16TH AVE

903 EAST 16TH AVENUE  
954 E 17TH AVE  
954 E CARMEL DR  
00 FULTON ST

05 BLACKHAWK ST APT 107  
05 JAMAICA ST  
07 CATHAY CT  
07 CATHAY CT

09 LIMA ST

Cluster

AG

Fall20\_  
AGFall21\_  
AG

1

aa

aa

1

at

at

1

â

â

2

â

â

1

tt

tt

1

aa

aa

1

â

â

2

âââ

âââ

1

aa

aa

1

â

â

2

âK

âK

1

Kââ

Kââ

1

aa

aa

1

â

â

2

TKI

TKI

1

ia

ia

1

a

a

2

t

t

1

ta

ta

1

tta

tta

2

â

â

1

ta

ta

1

aa

aa

1

â

â

2

âK

âK

1

ai

ai

1

â

â

2

â

â

1

TK

TK

1

atk

atk

1

ki

ki

1

aa

aa

1

aa

aa

1

aa

aa

1

â

â

2

â

â

1

aaa

aaa

Two adult cases in this  
HH during Fall 2020

1 case in a kid followed  
by two adult cases in  
Fall 2021

Two clusters in this HH  
in different seasons  
(2020 and 2021)

Two clusters in this HH  
in same season

# Number of clusters per eligible HH

COUNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	18142	73.99	18142	73.99
2	6287	25.64	24429	99.63
3	88	0.36	24517	99.99
4	2	0.01	24519	100.00

# Number of observations\*

- **CEDRS cases:** 811,813
- **Filtered cases:** 181,763  
Sept 1-Nov1 in 2020 and 2021; County ^= International; Live in Institution ^= Yes.
- **Filtered cases with address:** 178,093  
Address is "County, City, Address1"
- **Filtered cases in eligible HH's:** 62,364
- **Eligible HH's:** 24,519
- **Clusters of filtered cases in Eligible HH's:** 30,988

\*as of 11/22/21

## B. Issues and Limitations



# Issues and Limitations

- I used *ReportedDate* instead of *CollectionDate* since the former has no missing values.
- Fall season went up to and including November 1st
- Over 2700 records were excluded because State = MISSING, even though city and zipcode was not.
- Cases were linked on Address1 field. This field contains numerous errors and very little effort was made in cleaning the data.

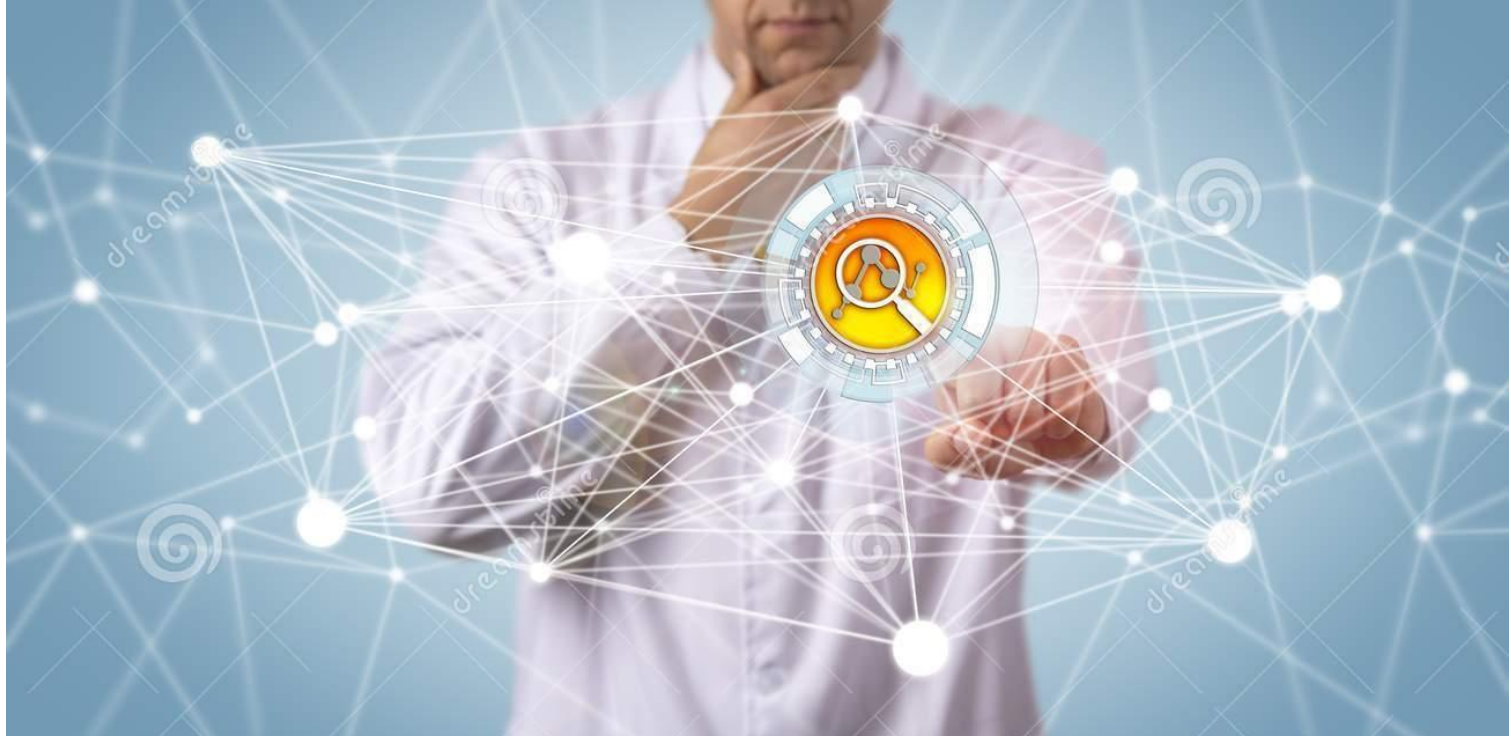
# Data errors in Address1

- Some records contain 'NO ADDRESS PROVIDED', 'N/A', 'UNK', 'UNKNOWN', '0'. These were set to missing.
- Contain phone number with address.
- Some records have address2 data in Address1 field, e.g. Unit #.  
Location of unit # may differ even within same street address.
- Street addresses that are clearly the same differ in format, e.g. ST vs Street.
- Address for an institution even though LiveInInstitution is not YES.

## Issues and Limitations, continued

- Records with the same street address may belong to an apartment complex. If unit # was in address1 field, then each apartment was considered a HH. If unit # was in address2 field (where it belongs) then the apartment complex as a whole was defined as the HH. No effort was made to clean this data so definition of HH for apartments was inconsistent.

## C. Data results and findings





# Number of HH with 1 or more cases by time period

	Number of cases	Number of clusters
Fall 2020	8232	8451
Fall 2021	17694	23151
Total	24519	30988

# Number of clusters per HH for Fall 2020

COUNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	8013	97.34	8013	97.34
2	219	2.66	8232	100.00

# Number of clusters per HH for Fall 2021

COUNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	11706	68.40	11706	68.40
2	5393	31.51	17099	99.91
3	15	0.09	17114	100.00

# Number of clusters per HH over all

COUNT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	18142	73.99	18142	73.99
2	6287	25.64	24429	99.63
3	88	0.36	24517	99.99
4	2	0.01	24519	100.00

# Age Group that began the HH cluster by time period

Fall 2020

Fall20_AG	Frequency	Percent
a	7261	85.92
i	234	2.77
k	403	4.77
t	553	6.54

Fall 2021

Fall21_AG	Frequency	Percent
A	16480	71.18
I	1082	4.67
K	3399	14.68
T	2190	9.46

**There was a large increase in the percent of HH clusters that began with kids in Fall 2021 compared to Fall 2020.**