Analysis of the Genealogy Process in Investigative Genetic Genealogy

Lawrence M. Wein

Mine Su Erturk

Graduate School of Business, Stanford University

1

PREVIOUS CRIME-SOLVING RESEARCH

US-VISIT Program
- *PNAS* (2005) and Congressional testimony (2005)
- Led to switch from 2 to 10 fingers

UIDAI (Aadhaar) Program in India
- *PLOS ONE* (2014) and India Supreme Court briefing (2014)

Ballistic imaging
- *Journal of Forensic Sciences* 2014, 2017
- *AFTE Journal* 2018 (Stockton CA PD)

Sexual assault kits
- Testing the backlog: *Journal of Forensic Sciences* 2018
  cnn.com 2018
- How to test the backlog: *PNAS* 2020 (SFPD)

2

ACKNOWLEDGMENTS

Colleen Fitzpatrick
IdentiFinders

Margaret Press
DNA Doe Project

3

OUTLINE

Introduction: research questions and data

Main results

Parameter estimation

Modeling the two-stage genealogy process

Proposed strategy

Limitations

Conclusions

4

STEPS IN THE IGG PROCESS

1. Obtain DNA sample from crime scene or unidentified remains and perform genotyping (SNPs)

5

STEPS IN THE IGG PROCESS

1. Obtain DNA sample from crime scene or unidentified remains and perform genotyping (SNPs)

2. Upload SNP data to third-party service to obtain relatives of target

GED match FamilyTreeDNA

6

STEPS IN THE IGG PROCESS

1. Obtain DNA sample from crime scene or unidentified remains and perform genotyping (SNPs)
2. Upload SNP data to third-party service to obtain relatives of target GEDmatch FamilyTreeDNA
3. ➡ Perform genealogy research to identify target

7

STEPS IN THE IGG PROCESS

1. Obtain DNA sample from crime scene or unidentified remains and perform genotyping (SNPs)
2. Upload SNP data to third-party service to obtain relatives of target GEDmatch FamilyTreeDNA
3. ➡ Perform genealogy research to identify target
4. Obtain confirmatory DNA sample from identified target or family member

8

RESEARCH QUESTIONS

First mathematical analysis of the backend of the IGG process

(i.e., GEDmatch/FTDNA output → identify target)

9

RESEARCH QUESTIONS

First mathematical analysis of the backend of the IGG process

(i.e., GEDmatch/FTDNA output → identify target)

Performance analysis:

Given the GEDmatch/FTDNA output, compute:

Probability of identifying target

Expected workload (= size of final family tree)

RESEARCH QUESTIONS

First mathematical analysis of the backend of the IGG process

(i.e., GEDmatch/FTDNA output → identify target)

Performance analysis:

Given the GEDmatch/FTDNA output, compute:

Probability of identifying target

Expected workload (= size of final family tree)

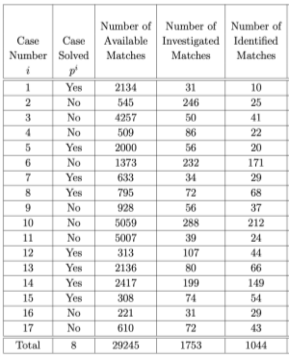
Optimization:

How many and which matches to investigate?

When/if to descend from (possible) MRCAs

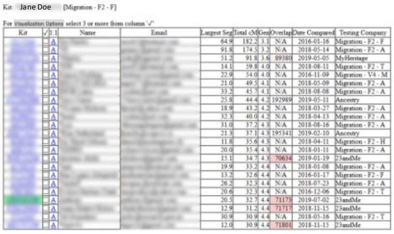
(MRCA = Most Recent Common Ancestor)

17 CASES FROM DNA DOE PROJECT



Case Number i	Case Solved p^i	Number of Available Matches	Number of Investigated Matches	Number of Identified Matches
1	Yes	2134	31	10
2	No	545	246	25
3	No	4257	50	41
4	No	509	86	22
5	Yes	2000	56	20
6	No	1373	232	171
7	Yes	633	34	29
8	Yes	795	72	68
9	No	928	56	37
10	No	5059	288	212
11	No	5007	39	24
12	Yes	313	107	44
13	Yes	2136	80	66
14	Yes	2417	199	149
15	Yes	308	74	54
16	No	221	31	29
17	No	610	72	43
Total	8	29245	1753	1044

GEDmatch OUTPUT

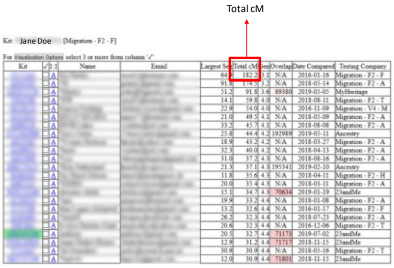


Kit: Jane Doe [Migration - F2 - F]

For Visualization Options select 3 or more from column '√'

Kit	√	1:1	Name	Email	Largest Seg	Total cM	Gen	Overlap	Date Compared	Testing Company
[redacted]	□	A	[redacted]	[redacted]	64.9	182.2	3.1	N/A	2016-01-16	Migration - F2 - F
[redacted]	□	A	[redacted]	[redacted]	91.8	174.5	3.2	N/A	2018-05-14	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	51.2	91.8	3.6	89380	2019-05-05	MyHeritage
[redacted]	□	A	[redacted]	[redacted]	14.1	59.8	4.0	N/A	2018-08-11	Migration - F2 - T
[redacted]	□	A	[redacted]	[redacted]	22.9	54.0	4.0	N/A	2016-11-09	Migration - V4 - M
[redacted]	□	A	[redacted]	[redacted]	21.0	49.5	4.1	N/A	2018-05-09	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	33.2	45.7	4.1	N/A	2018-08-08	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	25.8	44.4	4.2	192989	2019-05-11	Ancestry
[redacted]	□	A	[redacted]	[redacted]	18.9	43.2	4.2	N/A	2018-03-27	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	32.3	40.0	4.2	N/A	2018-04-13	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	31.0	37.2	4.3	N/A	2018-08-16	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	21.3	37.1	4.3	195341	2019-02-10	Ancestry
[redacted]	□	A	[redacted]	[redacted]	11.8	35.6	4.3	N/A	2018-04-11	Migration - F2 - H
[redacted]	□	A	[redacted]	[redacted]	20.0	35.4	4.3	N/A	2018-01-11	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	15.1	34.7	4.3	70634	2019-01-19	23andMe
[redacted]	□	A	[redacted]	[redacted]	19.9	33.2	4.4	N/A	2018-01-08	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	13.2	32.6	4.4	N/A	2016-01-17	Migration - F2 - F
[redacted]	□	A	[redacted]	[redacted]	26.2	32.3	4.4	N/A	2018-07-23	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	20.6	32.3	4.4	N/A	2016-12-06	Migration - F2 - T
[redacted]	□	A	[redacted]	[redacted]	20.5	32.7	4.4	71173	2019-07-02	23andMe
[redacted]	□	A	[redacted]	[redacted]	12.9	31.2	4.4	71717	2018-11-15	23andMe
[redacted]	□	A	[redacted]	[redacted]	30.9	30.9	4.4	N/A	2018-05-16	Migration - F2 - T
[redacted]	□	A	[redacted]	[redacted]	12.0	30.9	4.4	71801	2018-11-15	23andMe

13



Total cM

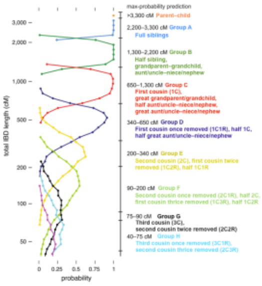
Kit: Jane Doe [Migration - F2 - F]

For Visualization Options select 3 or more from column '√'

Kit	√	1:1	Name	Email	Largest Seg	Total cM	Gen	Overlap	Date Compared	Testing Company
[redacted]	□	A	[redacted]	[redacted]	64.9	182.2	3.1	N/A	2016-01-16	Migration - F2 - F
[redacted]	□	A	[redacted]	[redacted]	91.8	174.5	3.2	N/A	2018-05-14	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	51.2	91.8	3.6	89380	2019-05-05	MyHeritage
[redacted]	□	A	[redacted]	[redacted]	14.1	59.8	4.0	N/A	2018-08-11	Migration - F2 - T
[redacted]	□	A	[redacted]	[redacted]	22.9	54.0	4.0	N/A	2016-11-09	Migration - V4 - M
[redacted]	□	A	[redacted]	[redacted]	21.0	49.5	4.1	N/A	2018-05-09	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	33.2	45.7	4.1	N/A	2018-08-08	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	25.8	44.4	4.2	192989	2019-05-11	Ancestry
[redacted]	□	A	[redacted]	[redacted]	18.9	43.2	4.2	N/A	2018-03-27	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	32.3	40.0	4.2	N/A	2018-04-13	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	31.0	37.2	4.3	N/A	2018-08-16	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	21.3	37.1	4.3	195341	2019-02-10	Ancestry
[redacted]	□	A	[redacted]	[redacted]	11.8	35.6	4.3	N/A	2018-04-11	Migration - F2 - H
[redacted]	□	A	[redacted]	[redacted]	20.0	35.4	4.3	N/A	2018-01-11	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	15.1	34.7	4.3	70634	2019-01-19	23andMe
[redacted]	□	A	[redacted]	[redacted]	19.9	33.2	4.4	N/A	2018-01-08	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	13.2	32.6	4.4	N/A	2016-01-17	Migration - F2 - F
[redacted]	□	A	[redacted]	[redacted]	26.2	32.3	4.4	N/A	2018-07-23	Migration - F2 - A
[redacted]	□	A	[redacted]	[redacted]	20.6	32.3	4.4	N/A	2016-12-06	Migration - F2 - T
[redacted]	□	A	[redacted]	[redacted]	20.5	32.7	4.4	71173	2019-07-02	23andMe
[redacted]	□	A	[redacted]	[redacted]	12.9	31.2	4.4	71717	2018-11-15	23andMe
[redacted]	□	A	[redacted]	[redacted]	30.9	30.9	4.4	N/A	2018-05-16	Migration - F2 - T
[redacted]	□	A	[redacted]	[redacted]	12.0	30.9	4.4	71801	2018-11-15	23andMe

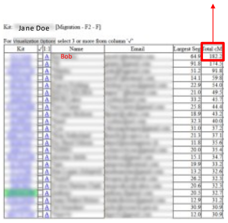
14

PROBABILISTIC MAPPING FROM TOTAL cM TO RELATIONSHIPS



https://thednageek.com/science-the-heck-out-of-your-dna-part-3/
Ball et al. (2016), Ancestry DNA Matching White Paper.

15

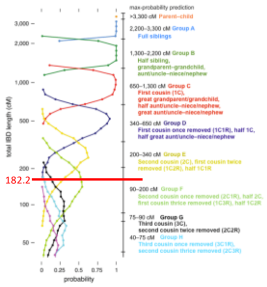


Total cM

Kit: Jane Doe [Migration - F2 - F]

For Visualization Options select 3 or more from column '√'

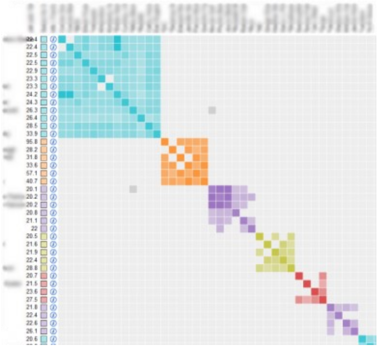
Kit	√	1:1	Name	Email	Largest Seg	Total cM
[redacted]	□	A	Bob	[redacted]	64.9	182.2
[redacted]	□	A	[redacted]	[redacted]	91.8	174.5
[redacted]	□	A	[redacted]	[redacted]	31.2	91.8
[redacted]	□	A	[redacted]	[redacted]	14.1	59.8
[redacted]	□	A	[redacted]	[redacted]	22.9	54.0
[redacted]	□	A	[redacted]	[redacted]	21.0	49.5
[redacted]	□	A	[redacted]	[redacted]	33.2	45.7
[redacted]	□	A	[redacted]	[redacted]	25.8	44.4
[redacted]	□	A	[redacted]	[redacted]	18.9	43.2
[redacted]	□	A	[redacted]	[redacted]	32.3	40.0
[redacted]	□	A	[redacted]	[redacted]	31.0	37.2
[redacted]	□	A	[redacted]	[redacted]	21.3	37.1
[redacted]	□	A	[redacted]	[redacted]	11.8	35.6
[redacted]	□	A	[redacted]	[redacted]	20.0	35.4
[redacted]	□	A	[redacted]	[redacted]	15.1	34.7
[redacted]	□	A	[redacted]	[redacted]	19.9	33.2
[redacted]	□	A	[redacted]	[redacted]	13.2	32.6
[redacted]	□	A	[redacted]	[redacted]	28.2	32.3
[redacted]	□	A	[redacted]	[redacted]	20.6	32.3
[redacted]	□	A	[redacted]	[redacted]	20.3	32.2
[redacted]	□	A	[redacted]	[redacted]	12.9	31.2
[redacted]	□	A	[redacted]	[redacted]	30.9	30.9
[redacted]	□	A	[redacted]	[redacted]	12.0	30.9



https://thednageek.com/science-the-heck-out-of-your-dna-part-3/
Ball et al. (2016), Ancestry DNA Matching White Paper.

16

AUTO CLUSTER TOOL IN GEDmatch



17

OUTLINE

Introduction: research questions and data

➡ Main results

Parameter estimation

Modeling the two-stage genealogy process

Proposed strategy

Limitations

Conclusions

18

BENCHMARK STRATEGY

Search for MRCAs between two matches,
and immediately descend from these MRCAs

19

BENCHMARK STRATEGY

Search for MRCAs between two matches,
and immediately descend from these MRCAs

Investigate n matches prioritized by highest total cM

20

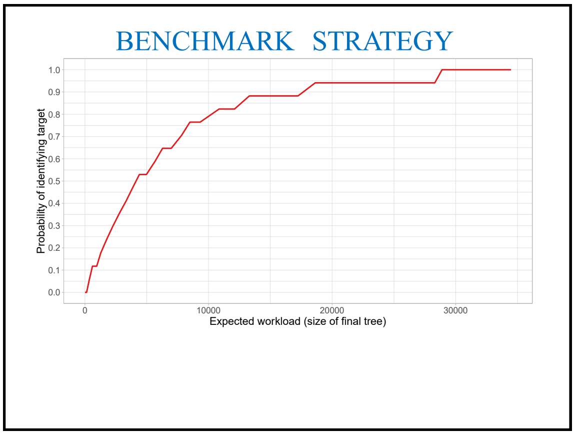
BENCHMARK STRATEGY

Search for MRCAs between two matches,
and immediately descend from these MRCAs

Investigate n matches prioritized by highest total cM

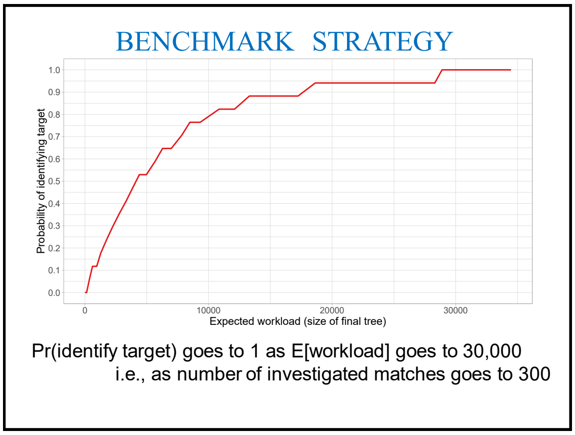
Vary n to generate Pr(identify target) vs. E[workload] curve

21



BENCHMARK STRATEGY

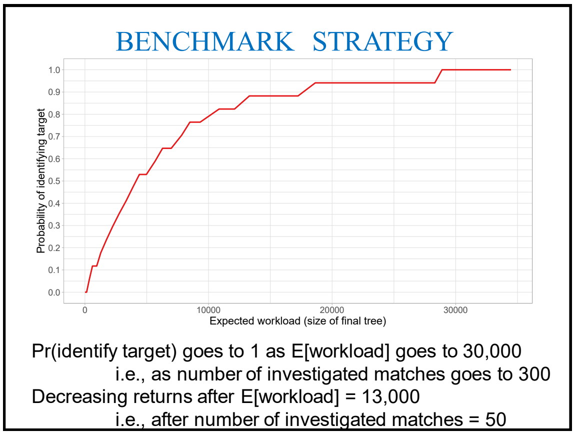
22



BENCHMARK STRATEGY

Pr(identify target) goes to 1 as E[workload] goes to 30,000
i.e., as number of investigated matches goes to 300

23

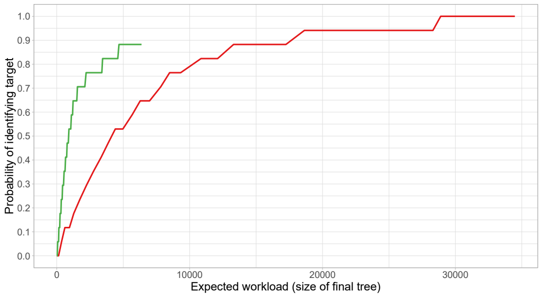


BENCHMARK STRATEGY

Pr(identify target) goes to 1 as E[workload] goes to 30,000
i.e., as number of investigated matches goes to 300
Decreasing returns after E[workload] = 13,000
i.e., after number of investigated matches = 50

24

PROPOSED vs. BENCHMARK STRATEGY



It solves cases much more quickly: at E[workload] = 2000, it solves 71% of cases vs. 27% for Benchmark

25

OUTLINE

- Introduction: research questions and data
- Main results
- ➡ Parameter estimation
- Modeling the two-stage genealogy process
- Proposed strategy
- Limitations
- Conclusions

26

PARAMETERS

Pr(can identify a match) = 0.59

27

PARAMETERS

Pr(can identify a match) = 0.59

Pr(can identify someone’s spouse) = 1

Pr(can identify someone’s child) = 0.98 (also considered 0.90)

Pr(can identify someone’s parents) = 0.60 (by simulation)

28

PARAMETERS

Pr(can identify a match) = 0.59

Pr(can identify someone’s spouse) = 1

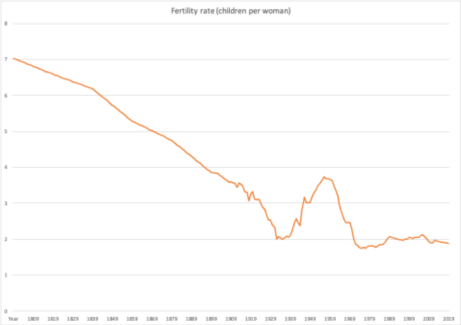
Pr(can identify someone’s child) = 0.98 (also considered 0.90)

Pr(can identify someone’s parents) = 0.60 (by simulation)

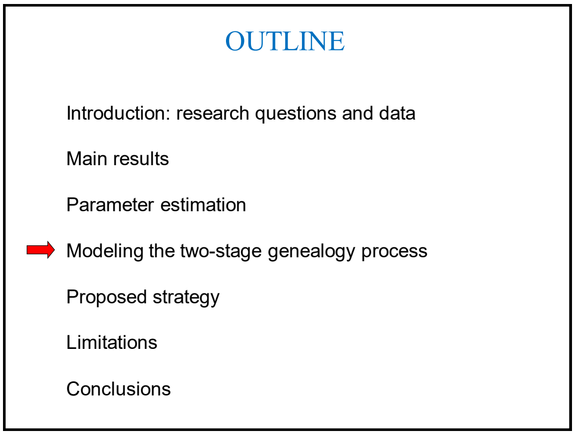
Number of children per couple

29

NUMBER OF CHILDREN PER COUPLE



30



OUTLINE

Introduction: research questions and data

Main results

Parameter estimation

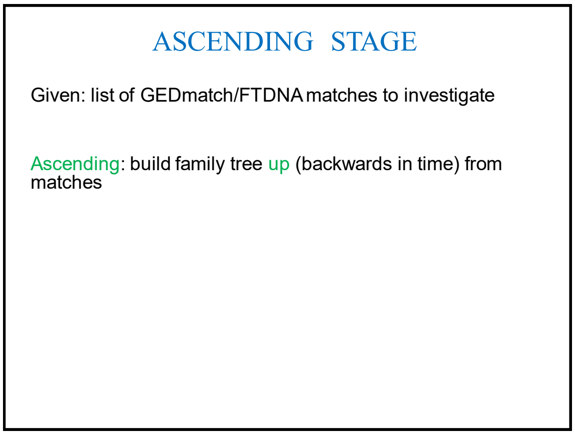
➡ Modeling the two-stage genealogy process

Proposed strategy

Limitations

Conclusions

31

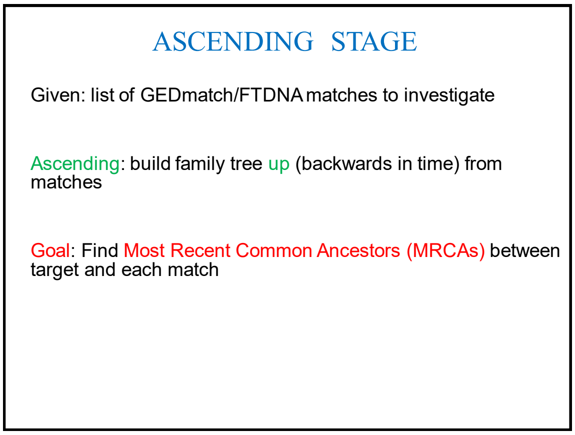


ASCENDING STAGE

Given: list of GEDmatch/FTDNA matches to investigate

Ascending: build family tree up (backwards in time) from matches

32



ASCENDING STAGE

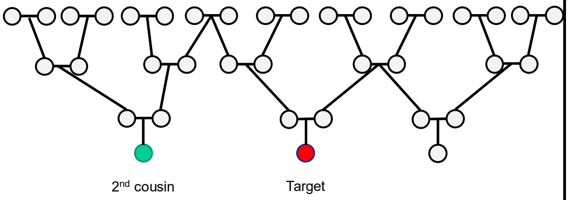
Given: list of GEDmatch/FTDNA matches to investigate

Ascending: build family tree up (backwards in time) from matches

Goal: Find Most Recent Common Ancestors (MRCAs) between target and each match

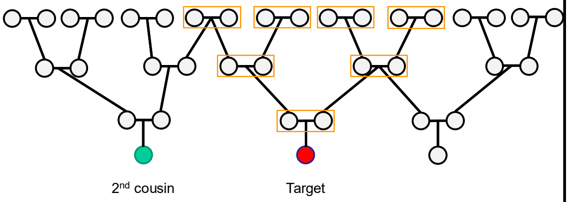
33

MOST RECENT COMMON ANCESTORS



34

MOST RECENT COMMON ANCESTORS



Cluster = ancestral couple of target
7 clusters in this example

35

ASCENDING STAGE

Given: list of GEDmatch/FTDNA matches to investigate

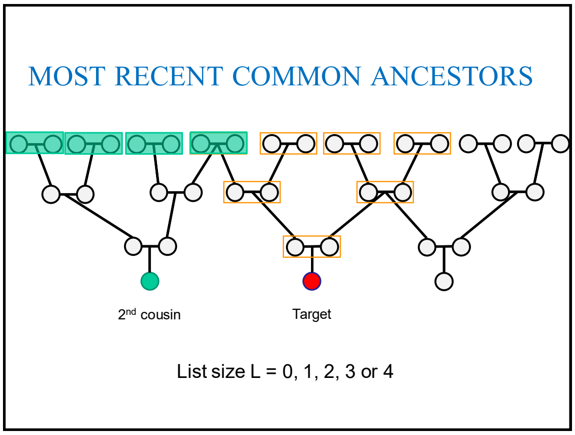
Ascending: build family tree up (backwards in time) from matches

Goal: Find Most Recent Common Ancestors (MRCAs) between target and each match

State of system during ascending stage:

For each generation g and cluster c = 1,…,2^{g-1}
$L_{g,c}$ = number of possible MRCAs identified
$P_{g,c}$ = Pr(one of the $L_{g,c}$ MRCAs is the correct MRCA)

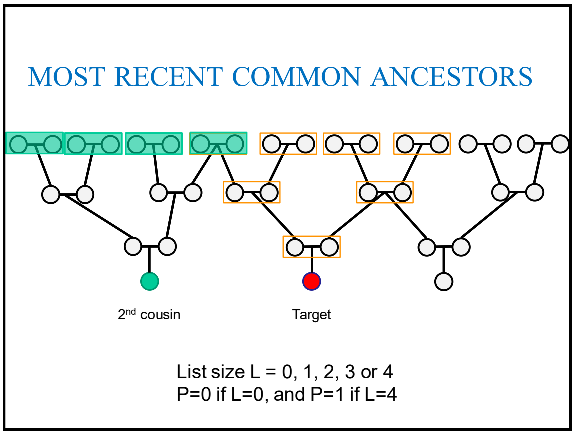
36



MOST RECENT COMMON ANCESTORS

List size L = 0, 1, 2, 3 or 4

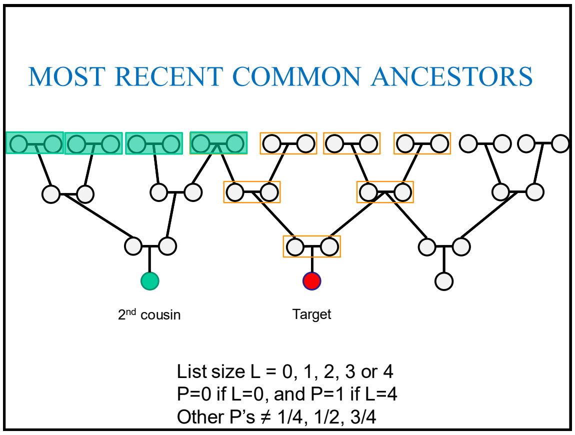
37



MOST RECENT COMMON ANCESTORS

List size L = 0, 1, 2, 3 or 4
P=0 if L=0, and P=1 if L=4

38

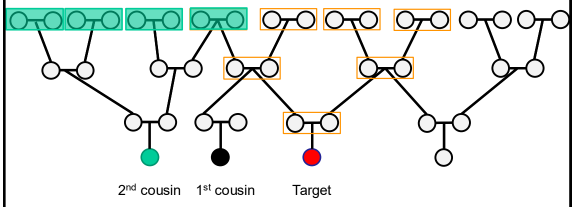


MOST RECENT COMMON ANCESTORS

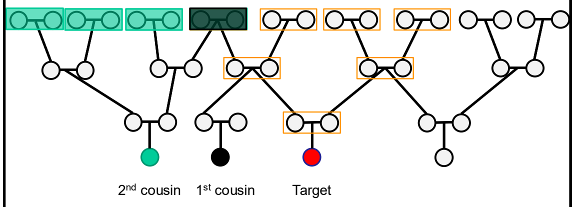
List size L = 0, 1, 2, 3 or 4
P=0 if L=0, and P=1 if L=4
Other P’s ≠ 1/4, 1/2, 3/4

39

MOST RECENT COMMON ANCESTORS



MOST RECENT COMMON ANCESTORS



Now state changes from (L=4,P=1) to (L=1,P=1)

DESCENDING STAGE

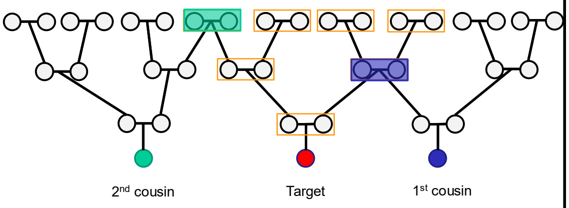
Given: State of system during ascending stage:

For each generation g and cluster c = 1,…,2^{g-1}
$L_{g,c}$ = number of possible MRCAs identified
$P_{g,c}$ = Pr(one of the $L_{g,c}$ MRCAs is the correct MRCA)

Descending: build family tree down (forwards in time) from possible MRCAs between target and match

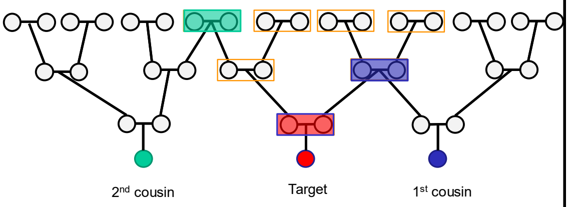
Goal: Find intersection of (i.e., marriage between) maternal and paternal family trees

FIND INTERSECTION OF FAMILY TREES



43

FIND INTERSECTION OF FAMILY TREES



44

DESCENDING STAGE

Given: State of system at end of ascending stage:

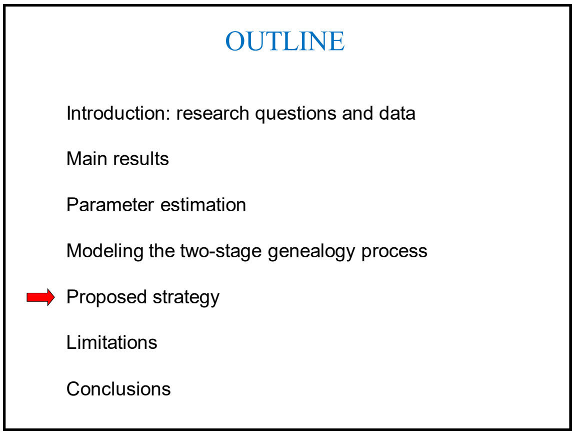
For each generation g and cluster c = 1,…,2^{g-1}
$L_{g,c}$ = number of possible MRCAs identified
$P_{g,c}$ = Pr(one of the $L_{g,c}$ MRCAs is the correct MRCA)

Descending: build family tree down (forwards in time) from possible MRCAs between target and match

Goal: Find intersection of (i.e., marriage between) maternal and paternal family trees

Compute: 1) probability of finding intersection of family trees
2) expected workload

45



OUTLINE

Introduction: research questions and data

Main results

Parameter estimation

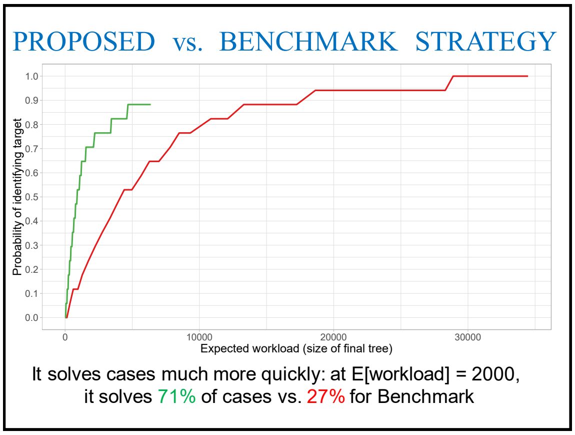
Modeling the two-stage genealogy process

➡ Proposed strategy

Limitations

Conclusions

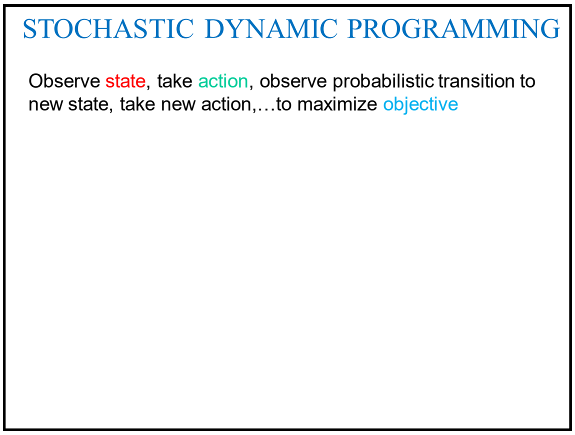
46



PROPOSED vs. BENCHMARK STRATEGY

It solves cases much more quickly: at E[workload] = 2000, it solves 71% of cases vs. 27% for Benchmark

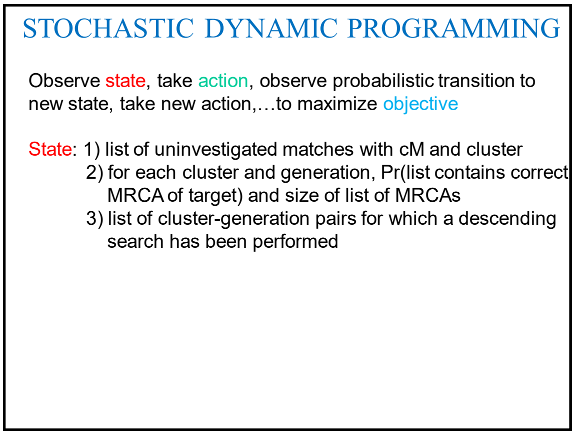
47



STOCHASTIC DYNAMIC PROGRAMMING

Observe state, take action, observe probabilistic transition to new state, take new action,…to maximize objective

48

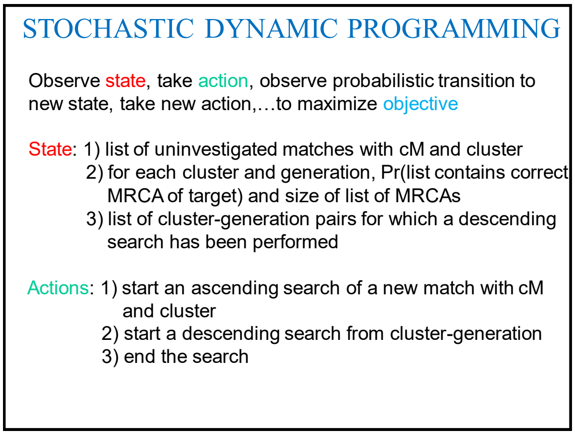


STOCHASTIC DYNAMIC PROGRAMMING

Observe state, take action, observe probabilistic transition to new state, take new action,…to maximize objective

State: 1) list of uninvestigated matches with cM and cluster
2) for each cluster and generation, Pr(list contains correct MRCA of target) and size of list of MRCAs
3) list of cluster-generation pairs for which a descending search has been performed

49



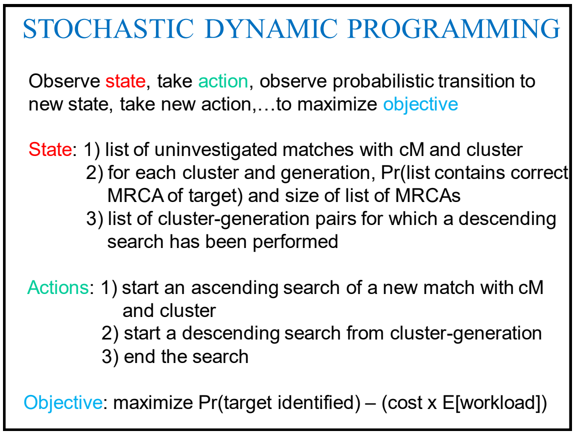
STOCHASTIC DYNAMIC PROGRAMMING

Observe state, take action, observe probabilistic transition to new state, take new action,…to maximize objective

State: 1) list of uninvestigated matches with cM and cluster
2) for each cluster and generation, Pr(list contains correct MRCA of target) and size of list of MRCAs
3) list of cluster-generation pairs for which a descending search has been performed

Actions: 1) start an ascending search of a new match with cM and cluster
2) start a descending search from cluster-generation
3) end the search

50



STOCHASTIC DYNAMIC PROGRAMMING

Observe state, take action, observe probabilistic transition to new state, take new action,…to maximize objective

State: 1) list of uninvestigated matches with cM and cluster
2) for each cluster and generation, Pr(list contains correct MRCA of target) and size of list of MRCAs
3) list of cluster-generation pairs for which a descending search has been performed

Actions: 1) start an ascending search of a new match with cM and cluster
2) start a descending search from cluster-generation
3) end the search

Objective: maximize Pr(target identified) – (cost x E[workload])

51

PROPOSED STRATEGY

SDP problem was too hard to solve (huge state space)

Use a greedy (ie, myopic) approach

52

PROPOSED STRATEGY

SDP problem was too hard to solve (huge state space)

Use a greedy (ie, myopic) approach

Given current state:

1) for a descending search, compute ΔP (increase in probability of identifying target) and ΔW (increase in E[workload]) for each undescended cluster-generation

53

PROPOSED STRATEGY

SDP problem was too hard to solve (huge state space)

Use a greedy (ie, myopic) approach

Given current state:

1) for a descending search, compute ΔP (increase in probability of identifying target) and ΔW (increase in E[workload]) for each undescended cluster-generation

2) for an ascending search, compute ΔP and ΔW for each uninvestigated match, assuming that we descend right after ascending

54

PROPOSED STRATEGY

SDP problem was too hard to solve (huge state space)

Use a greedy (ie, myopic) approach

Given current state:

1) for a descending search, compute ΔP (increase in probability of identifying target) and ΔW (increase in E[workload]) for each undescended cluster-generation

2) for an ascending search, compute ΔP and ΔW for each uninvestigated match, assuming that we descend right after ascending

3) compute the marginal increase in the objective (ie, ΔP – cost x ΔW) from 1) and 2) and choose the action with the maximum increase (and stop after n investigations)

PROPOSED STRATEGY

SDP problem was too hard to solve (huge state space)

Use a greedy (ie, myopic) approach

Given current state:

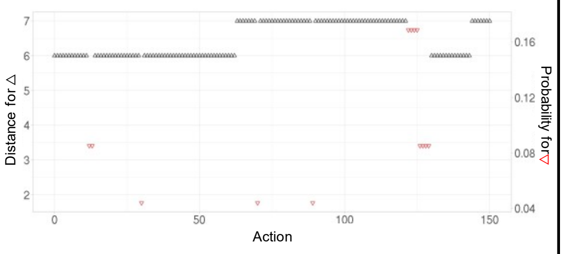
1) for a descending search, compute ΔP (increase in probability of identifying target) and ΔW (increase in E[workload]) for each undescended cluster-generation

2) for an ascending search, compute ΔP and ΔW for each uninvestigated match, assuming that we descend right after ascending

3) compute the marginal increase in the objective (ie, ΔP – cost x ΔW) from 1) and 2) and choose the action with the maximum increase (and stop after n investigations)

Vary n to generate Pr(target identified) vs. E[workload] curve

PROPOSED STRATEGY FOR CASE #1



Closest matches in Case #1 are at distance 6

△ = ascending from distance (left vertical axis)

▽ = descending from cluster-generation with probability of true MRCA (right vertical axis)

OUTLINE

Introduction: research questions and data

Main results

Parameter estimation

Modeling the two-stage genealogy process

Proposed strategy

➡ Limitations

Conclusions

58

LIMITATIONS

Cases: small sample size and not chosen randomly

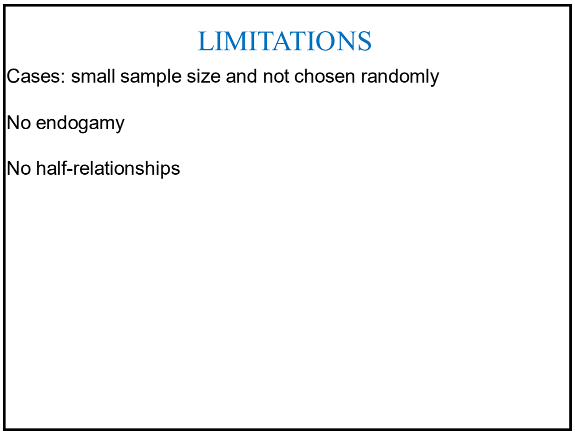
59

LIMITATIONS

Cases: small sample size and not chosen randomly

No endogamy

60



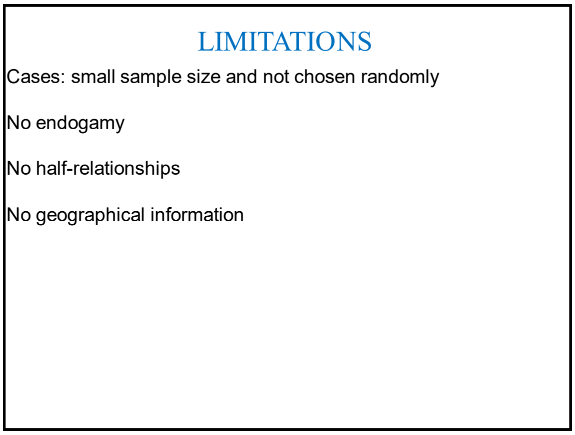
LIMITATIONS

Cases: small sample size and not chosen randomly

No endogamy

No half-relationships

61



LIMITATIONS

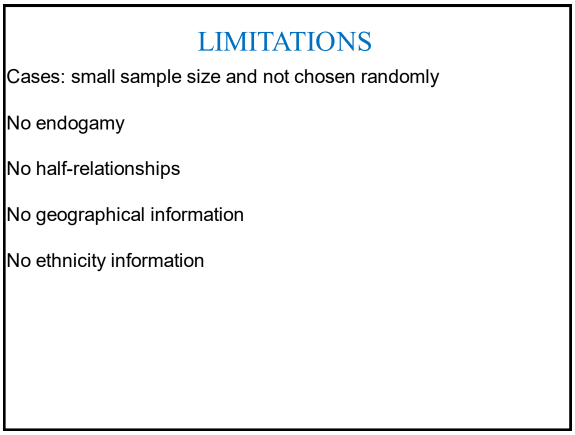
Cases: small sample size and not chosen randomly

No endogamy

No half-relationships

No geographical information

62



LIMITATIONS

Cases: small sample size and not chosen randomly

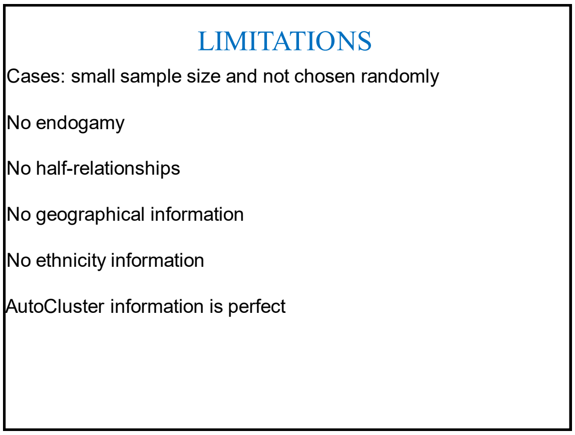
No endogamy

No half-relationships

No geographical information

No ethnicity information

63



LIMITATIONS

Cases: small sample size and not chosen randomly

No endogamy

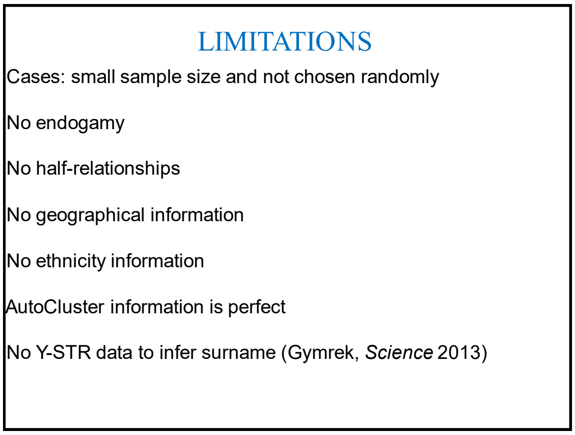
No half-relationships

No geographical information

No ethnicity information

AutoCluster information is perfect

64



LIMITATIONS

Cases: small sample size and not chosen randomly

No endogamy

No half-relationships

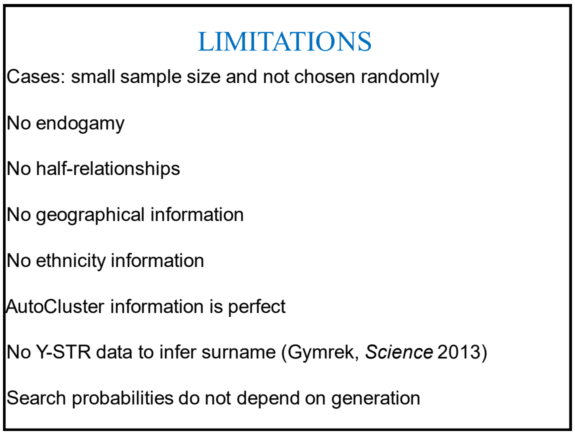
No geographical information

No ethnicity information

AutoCluster information is perfect

No Y-STR data to infer surname (Gymrek, *Science* 2013)

65



LIMITATIONS

Cases: small sample size and not chosen randomly

No endogamy

No half-relationships

No geographical information

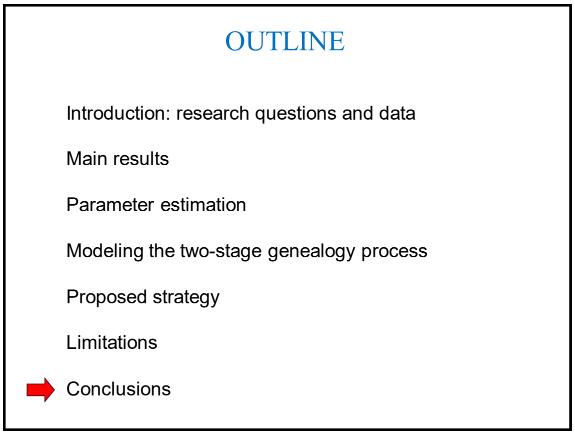
No ethnicity information

AutoCluster information is perfect

No Y-STR data to infer surname (Gymrek, *Science* 2013)

Search probabilities do not depend on generation

66



OUTLINE

Introduction: research questions and data

Main results

Parameter estimation

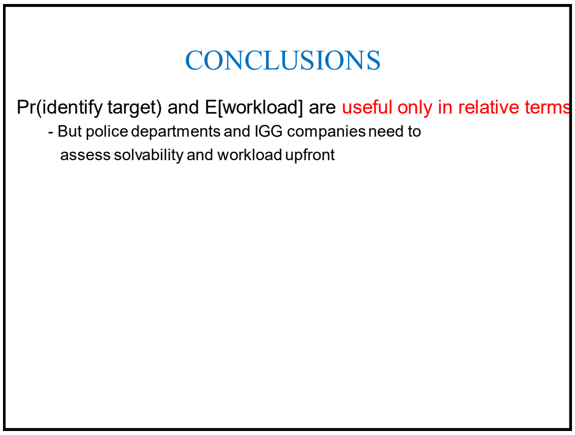
Modeling the two-stage genealogy process

Proposed strategy

Limitations

➡ Conclusions

67

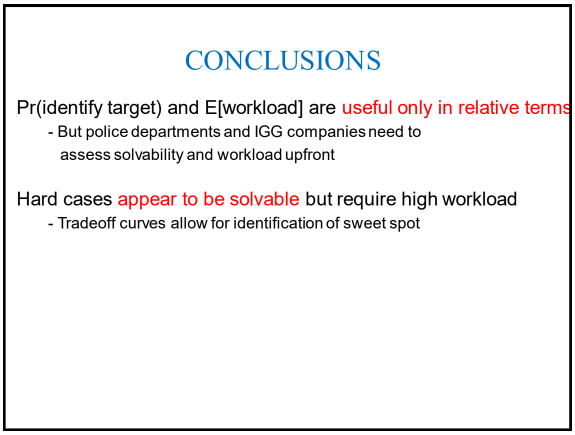


CONCLUSIONS

Pr(identify target) and E[workload] are useful only in relative terms

- But police departments and IGG companies need to assess solvability and workload upfront

68



CONCLUSIONS

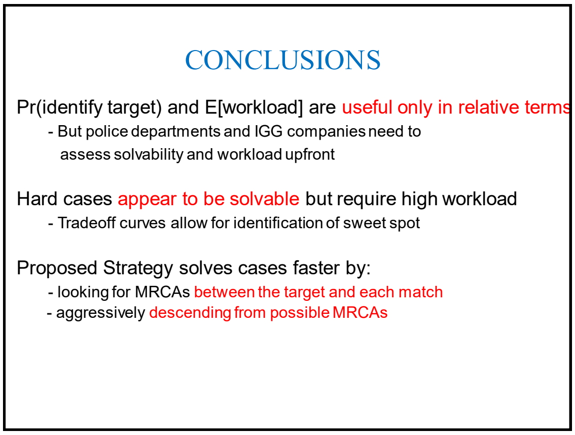
Pr(identify target) and E[workload] are useful only in relative terms

- But police departments and IGG companies need to assess solvability and workload upfront

Hard cases appear to be solvable but require high workload

- Tradeoff curves allow for identification of sweet spot

69



CONCLUSIONS

Pr(identify target) and E[workload] are useful only in relative terms

- But police departments and IGG companies need to assess solvability and workload upfront

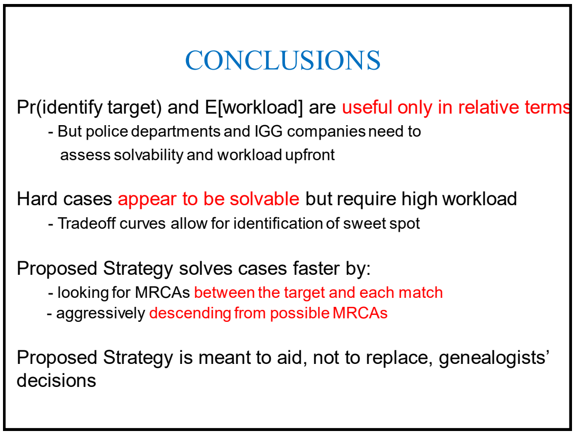
Hard cases appear to be solvable but require high workload

- Tradeoff curves allow for identification of sweet spot

Proposed Strategy solves cases faster by:

- looking for MRCAs between the target and each match
- aggressively descending from possible MRCAs

70



CONCLUSIONS

Pr(identify target) and E[workload] are useful only in relative terms

- But police departments and IGG companies need to assess solvability and workload upfront

Hard cases appear to be solvable but require high workload

- Tradeoff curves allow for identification of sweet spot

Proposed Strategy solves cases faster by:

- looking for MRCAs between the target and each match
- aggressively descending from possible MRCAs

Proposed Strategy is meant to aid, not to replace, genealogists' decisions

71