Portinari: A Data Exploration Tool to Personalize Cervical Cancer Screening

Sagar Sen*,

Manoel Horta Ribeiro†,

Raquel C. de Melo Minardi†,

Wagner Meira Jr.†,

Mari Nygård‡

* Certus V&V Center and HPV Research Group,

† Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil,

‡ HPV Research Group, Cancer Registry of Norway, Oslo, Norway

Software's Impact in Societal Processes

Through which means societies organize themselves?

Software's Impact in Societal Processes

Through which means societies organize themselves?

politics

Software's Impact in Societal Processes

Through which means societies organize themselves?

politics

Software's Impact in Societal Processes

Through which means societies organize themselves?

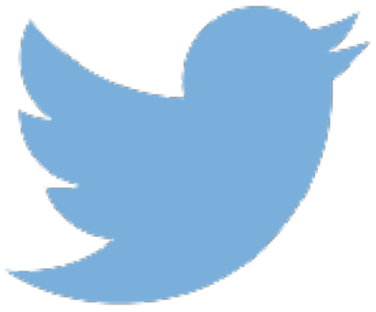
politics

social bonds

Software's Impact in Societal Processes

Through which means societies organize themselves?

politics

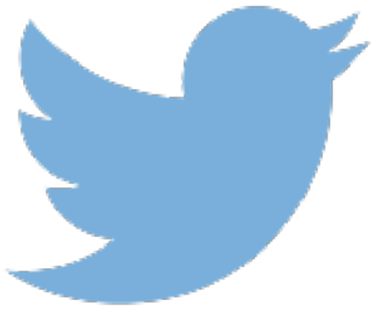


social bonds

Software's Impact in Societal Processes

Through which means societies organize themselves?

politics



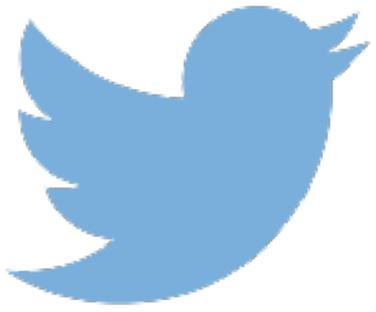
social bonds

labour

Software's Impact in Societal Processes

Through which means societies organize themselves?

politics



social bonds

labour

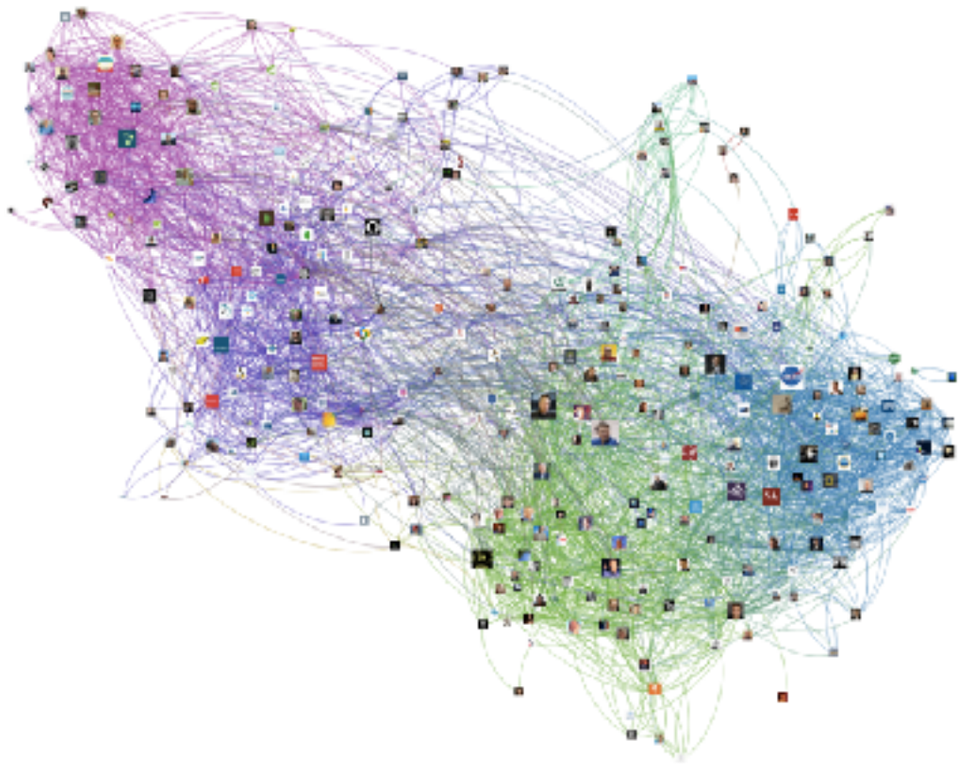
Self-organizing vs. Centralized

Self-organizing vs. Centralized

Self-organizing vs. Centralized

vs

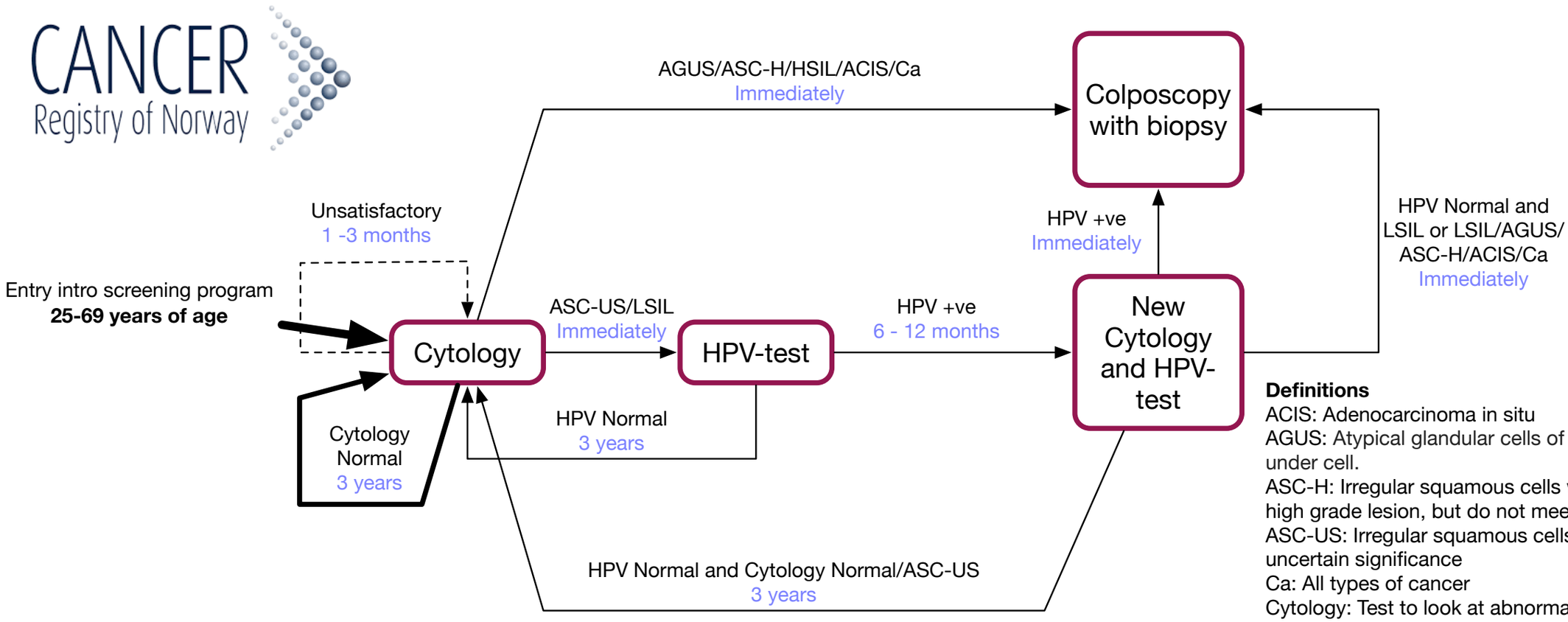
Self-organizing vs. Centralized



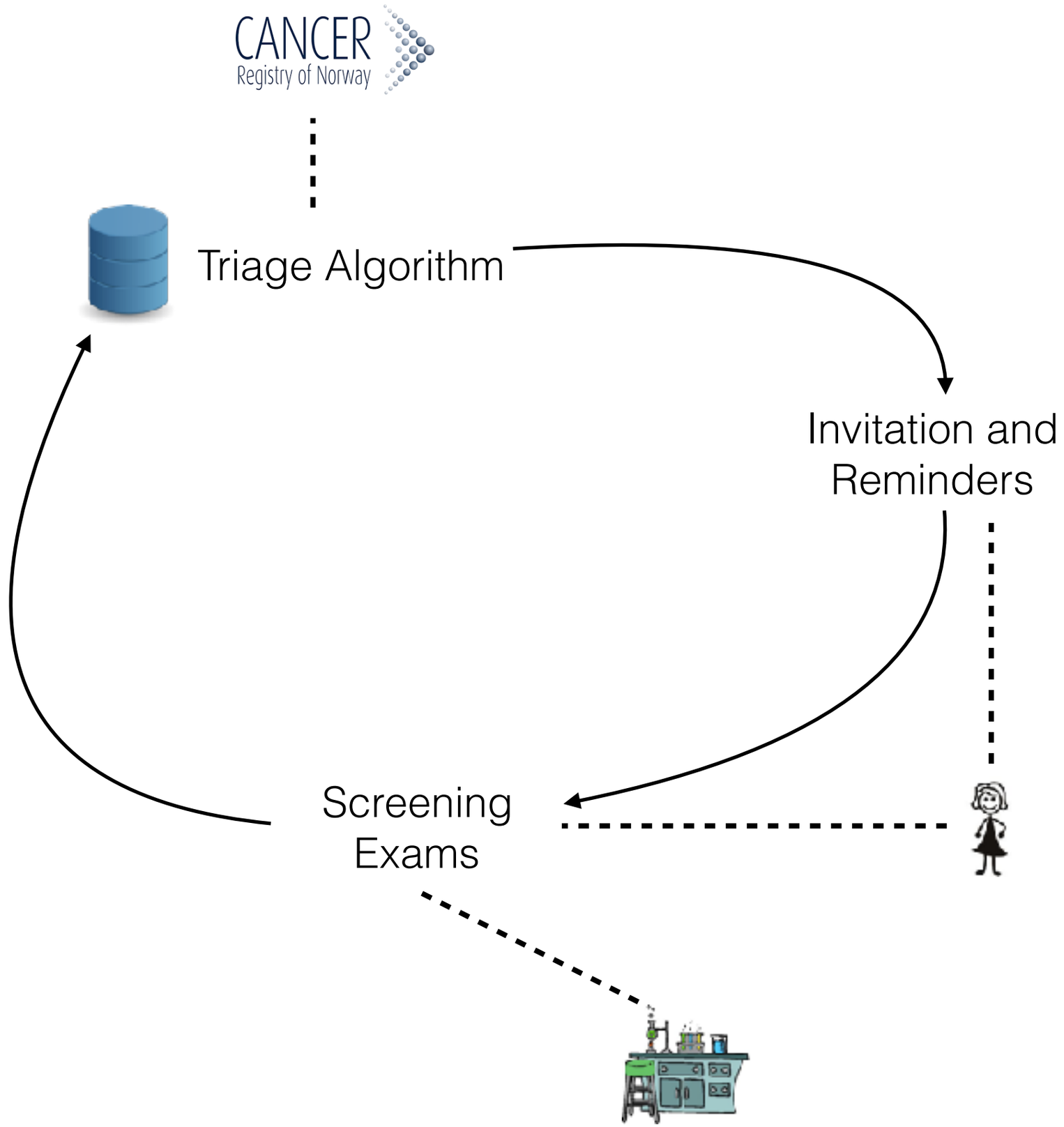
vs

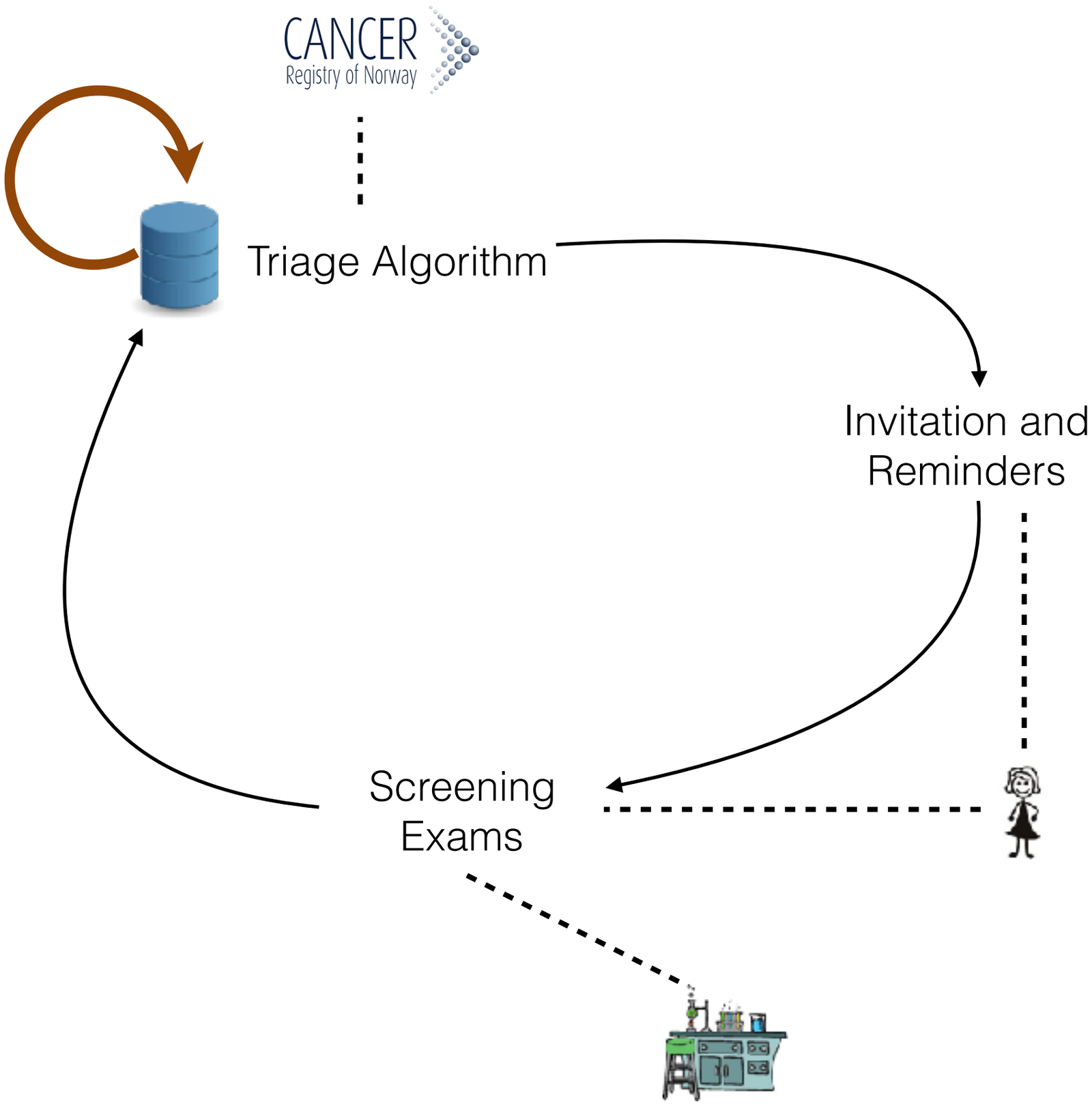
CANCER
Registry of Norway

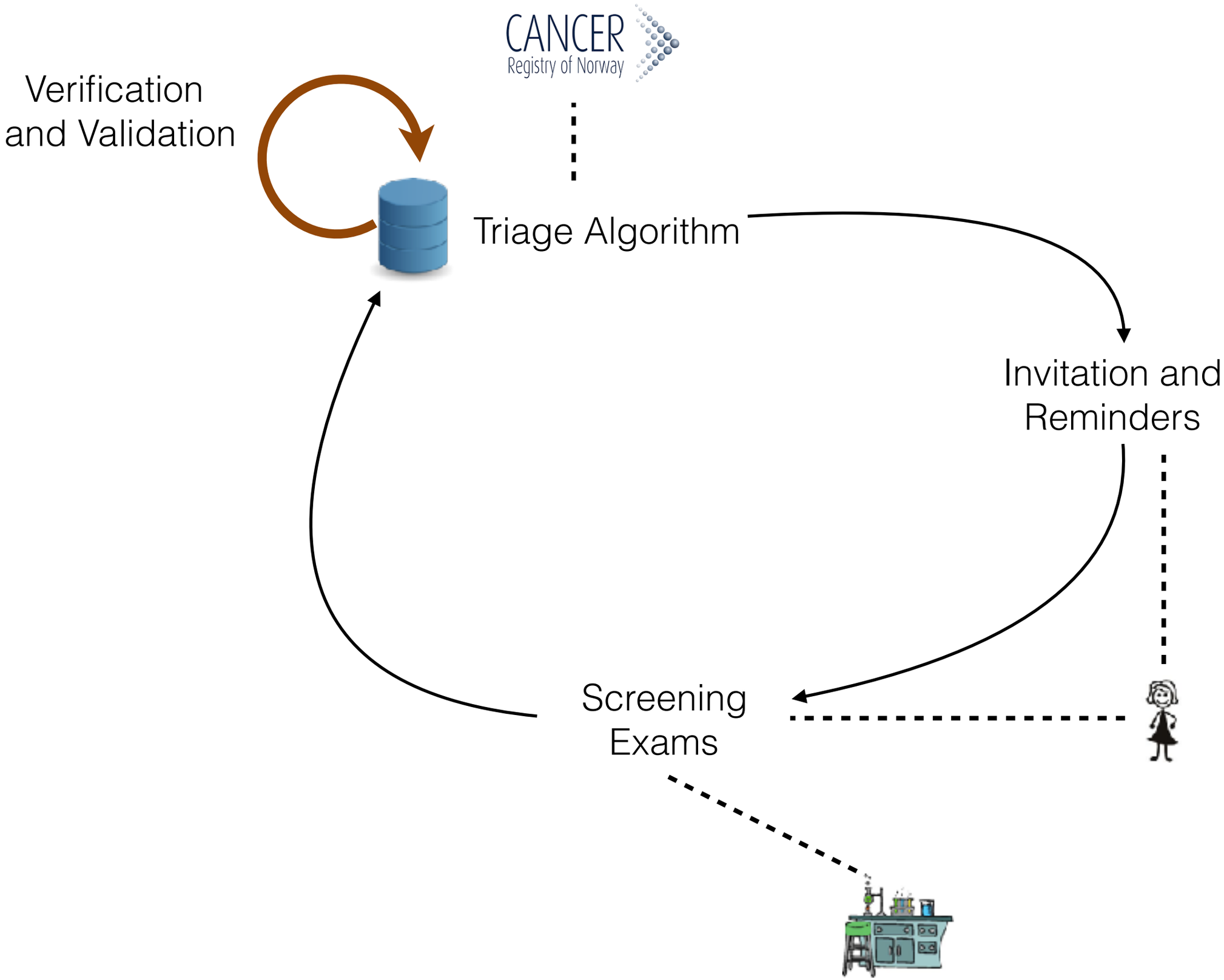
Norway's Screening Program

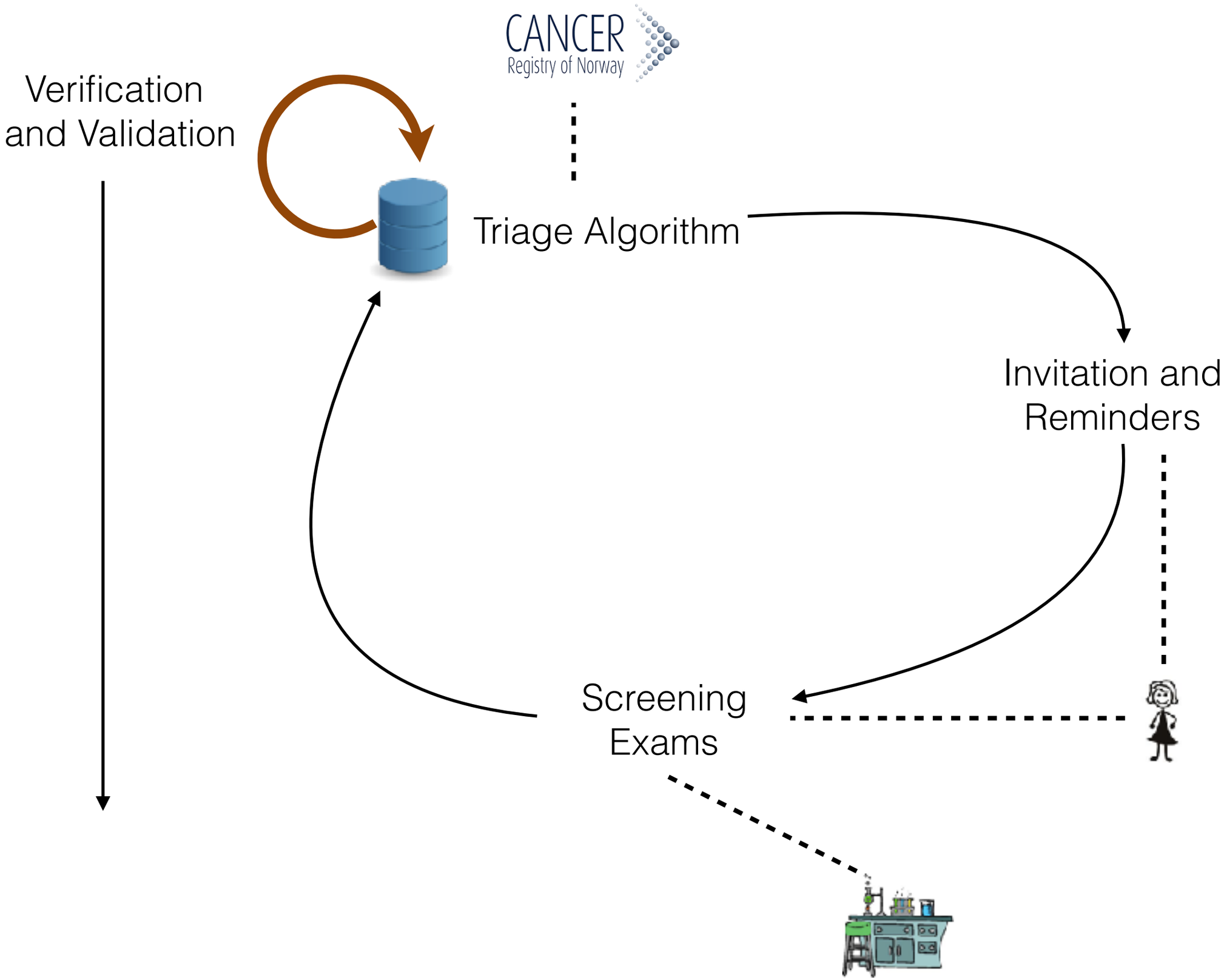


Definitions
ACIS: Adenocarcinoma in situ
AGUS: Atypical glandular cells of undetermined significance, under cell.
ASC-H: Irregular squamous cells with changes suggestive of high grade lesion, but do not meet all the criteria for diagnosis
ASC-US: Irregular squamous cells with abnormalities of uncertain significance
Ca: All types of cancer
Cytology: Test to look at abnormal cells
HPV: Humanpapillomavirus
HSIL: High-grade squamous intraepithelial lesion
LSIL: Low-grade squamous intraepithelial lesion


CANCER
Registry of Norway
Triage Algorithm
Invitation and
Reminders
Screening
Exams


CANCER
Registry of Norway
Triage Algorithm
Invitation and
Reminders
Screening
Exams


CANCER
Registry of Norway
Verification
and Validation
Triage Algorithm
Invitation and
Reminders
Screening
Exams


CANCER
Registry of Norway
Verification
and Validation
Triage Algorithm
Invitation and
Reminders
Screening
Exams

CANCER Registry of Norway

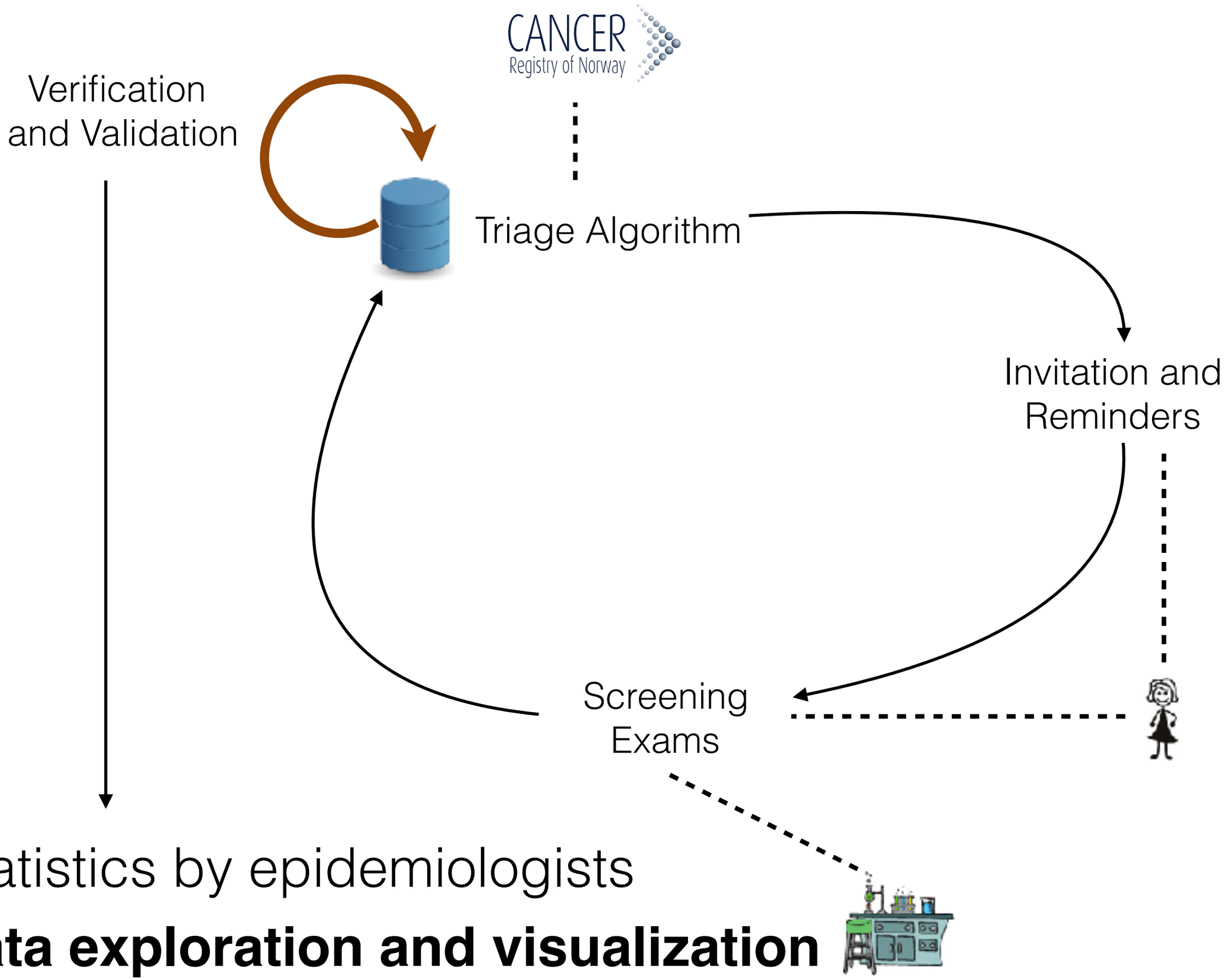
Verification and Validation

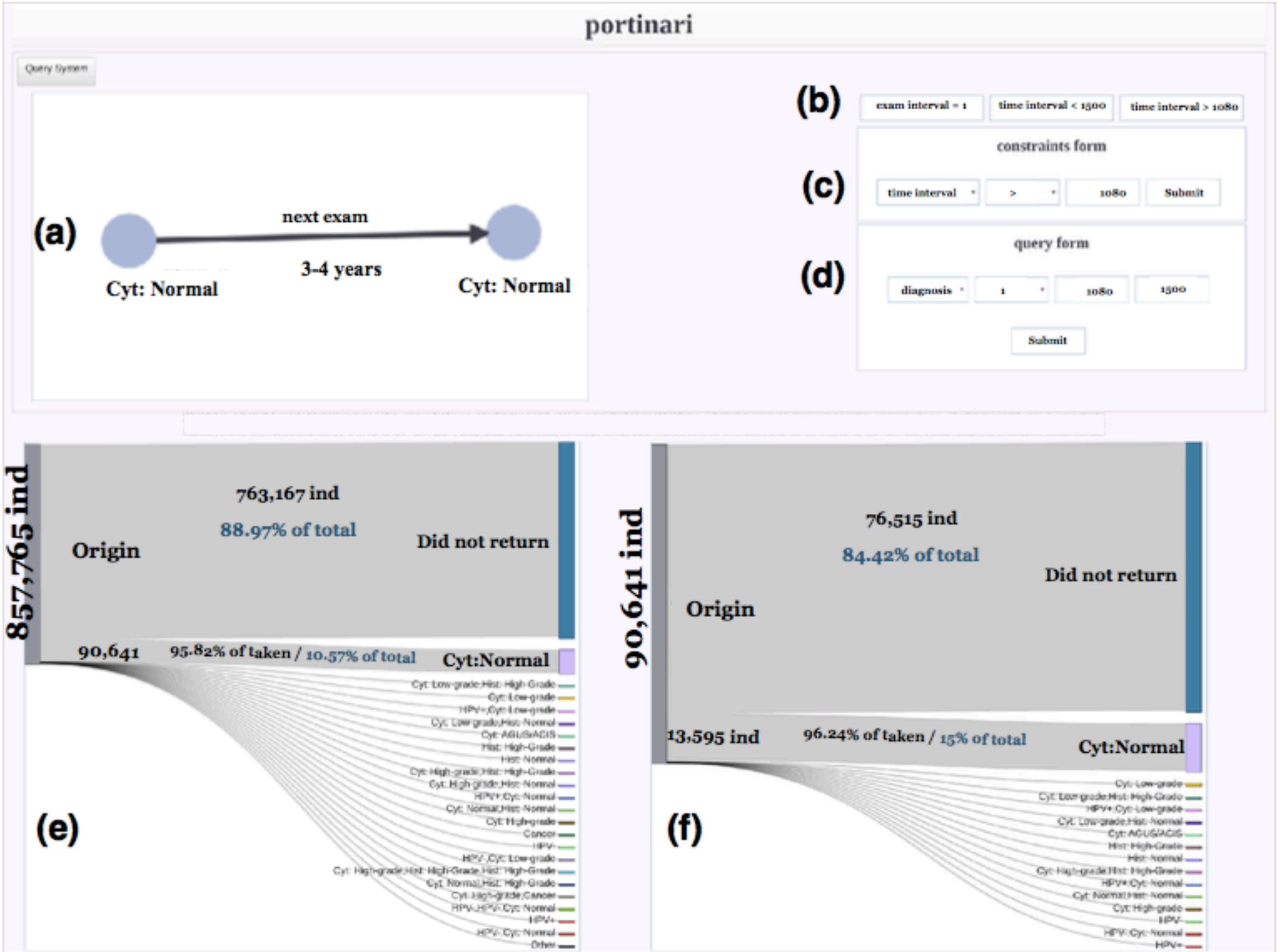
Triage Algorithm

Invitation and Reminders

Screening Exams

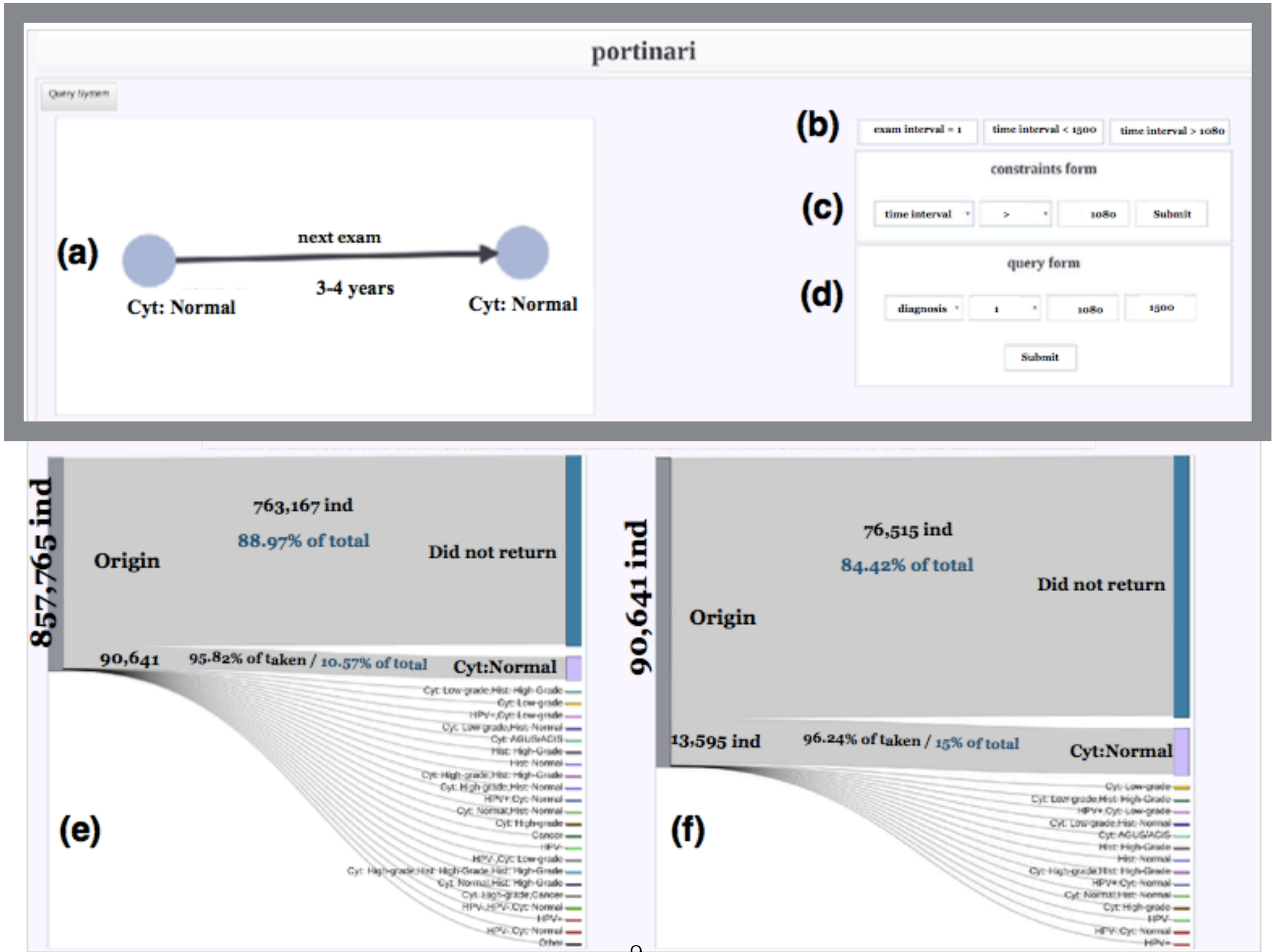
Statistics by epidemiologists


CANCER
Registry of Norway
Verification
and Validation
Triage Algorithm
Invitation and
Reminders
Screening
Exams
Statistics by epidemiologists
Data exploration and visualization


portinari
Query System
(a)
next exam
3-4 years
Cyt: Normal
Cyt: Normal
(b)
exam interval = 1
time interval < 1500
time interval > 1080
constraints form
(c)
time interval
>
1080
Submit
query form
(d)
diagnosis
1
1080
1500
Submit
857,765 ind
763,167 ind
88.97% of total
Origin
Did not return
90,641
95.82% of taken / 10.57% of total
Cyt:Normal
(e)
90,641 ind
76,515 ind
84.42% of total
Origin
Did not return
13,595 ind
96.24% of taken / 15% of total
Cyt:Normal
(f)

What data?

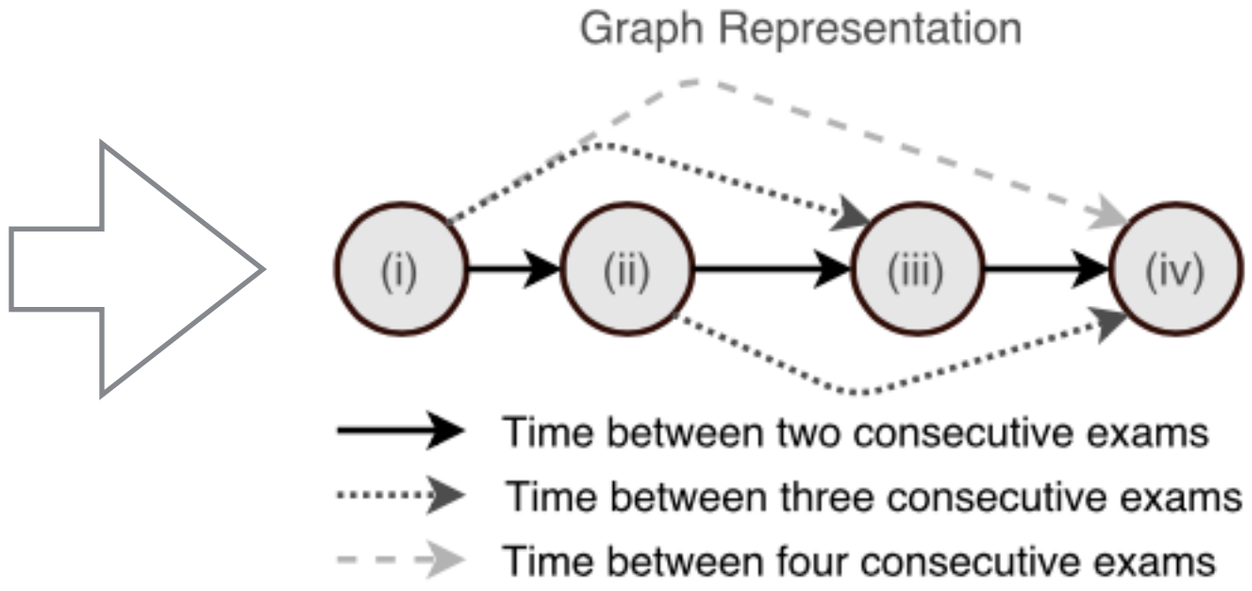
ID	Internal numeric identifier.
birthdate	Month and year of birth of the woman.
diagnosisdate	Date of diagnosis.
type	Type of visit/exam.
diagnosis	Result given by the exam/visit.
stage	Stage for cancer diagnosis.
lab_nr	Code for the laboratory, which is a 5-digit number.
region	Norwegian health care region.
censordate	Date for emigration/death/cervical cancer diagnosis.


portinari
Query System
(a)
next exam
3-4 years
Cyt: Normal
Cyt: Normal
(b)
exam interval = 1
time interval < 1500
time interval > 1080
constraints form
(c)
time interval
>
1080
Submit
query form
(d)
diagnosis
1
1080
1500
Submit
857,765 ind
763,167 ind
88.97% of total
Origin
Did not return
90,641
95.82% of taken / 10.57% of total
Cyt:Normal
(e)
90,641 ind
76,515 ind
84.42% of total
Origin
Did not return
13,595 ind
96.24% of taken / 15% of total
Cyt:Normal
(f)

What representation?

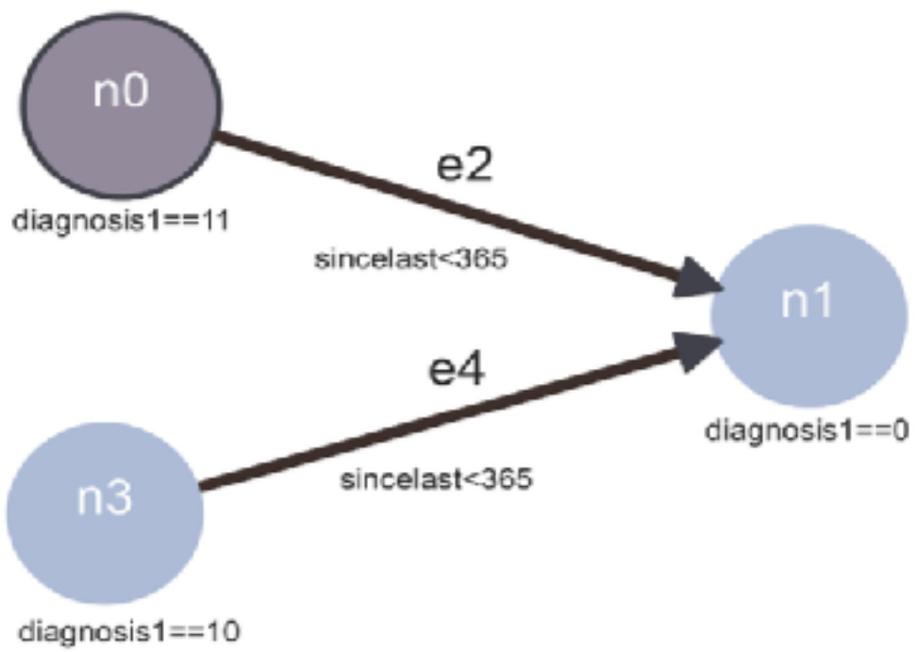
Tabular Representation

	ID	Birthdate	Diagnosis Date	Diagnosis Number
(i)	1	08/1960	05/1992	13
(ii)	1	08/1960	09/1992	12
(iii)	1	08/1960	11/1994	20
(iv)	1	08/1960	3/1995	12

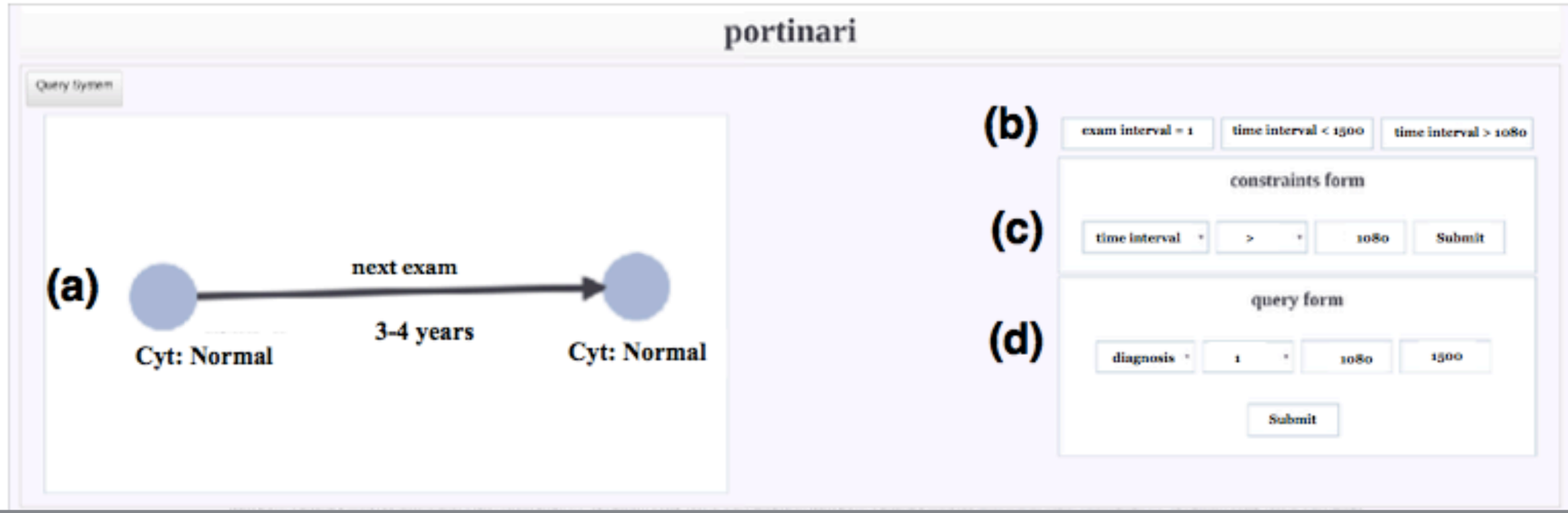



```
CREATE
(p1:Event // Create first node
  {PatientID:1, DiagnosisDate:"15.05.1992", BirthDate:"12.08.1960", ExamType:"cyt",
   Diagnosis:13, MorphologyCode:76700, CDate:NULL, LaboratoryNbr:19, Region:2}),
(p2:Event // Create second node
  {PatientID:1, DiagnosisDate:"15.09.1992", BirthDate:"12.08.1960", ExamType:"cyt",
  Diagnosis:12, MorphologyCode:69000, CDate:NULL, LaboratoryNbr:19, Region:2}),
(p3:Event // Create third node
  {PatientID:1, DiagnosisDate:"15.11.1994", BirthDate:"12.08.1960", ExamType:"cyt",
  Diagnosis:20, MorphologyCode:76700, CDate:NULL, LaboratoryNbr:19, Region:2}),
(p4:Event // Create third node
  {PatientID:1, DiagnosisDate:"15.03.1995", BirthDate:"12.08.1960", ExamType:"cyt",
  Diagnosis:12, MorphologyCode:76700, CDate:NULL, LaboratoryNbr:19, Region:2}),
// 1-hop edges
(p1)-[:Next1 {SinceLast:123}]->(p2),
(p2)-[:Next1 {SinceLast:791}]->(p3),
(p3)-[:Next1 {SinceLast:116}]->(p4),
// 2-hop edges
(p1)-[:Next2 {SinceLast:914}]->(p3), (p2)-[:Next2 {SinceLast:807}]->(p4),
// 3-hop edges
(p1)-[:Next3 {SinceLast:1030}]->(p4)
```

Querying

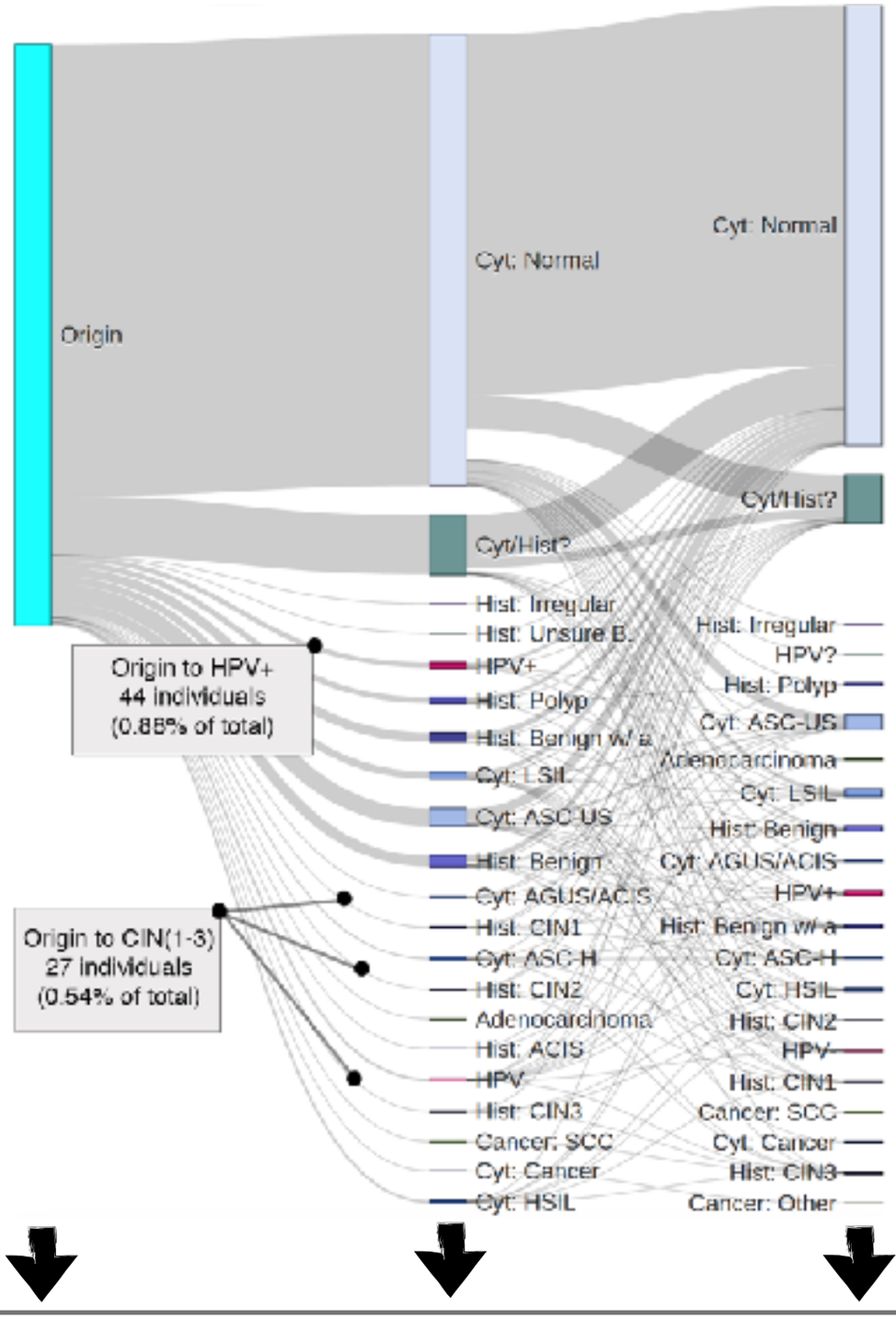



```
MATCH ( n0:Event { Diagnosis:11 })
MATCH ( n1:Event { Diagnosis:0  })
MATCH ( n3:Event { Diagnosis:10 })
MATCH (n0)-[e2]-(n1)
  WHERE e2.SinceLast < 365
MATCH (n1)-[e4]-(n3)
  WHERE e3.SinceLast < 365
MATCH (n3)-[ef1:Next1]->(nf1)
  WHERE ef1.SinceLast < 60
RETURN COUNT(distinct n3), nf1.Diagnosis
```


portinari
Query System
(a)
next exam
3-4 years
Cyt: Normal
Cyt: Normal
(b)
exam interval = 1
time interval < 1500
time interval > 1080
(c)
constraints form
time interval
>
1080
Submit
(d)
query form
diagnosis
1
1080
1500
Submit
(e)
857,765 ind
763,167 ind
88.97% of total
Origin
Did not return
90,641
95.82% of taken / 10.57% of total
Cyt:Normal
Cyt: Low-grade;Hist: High-Grade
Cyt: Low-grade
HPV+;Cyt: Low-grade
Cyt: Low-grade;Hist: Normal
Cyt: AGUS/AGIS
Hist: High-Grade
Hist: Normal
Cyt: High-grade;Hist: High-Grade
Cyt: High-grade;Hist: Normal
HPV+;Cyt: Normal
Cyt: Normal;Hist: Normal
Cyt: High-grade
Cancer
HPV-
HPV-;Cyt: Low-grade
Cyt: High-grade;Hist: High-Grade;Hist: High-Grade
Cyt: Normal;Hist: High-Grade
Cyt: High-grade;Cancer
HPV-;HPV-;Cyt: Normal
HPV+
HPV-;Cyt: Normal
Other
(f)
90,641 ind
76,515 ind
84.42% of total
Origin
Did not return
13,595 ind
96.24% of taken / 15% of total
Cyt:Normal
Cyt: Low-grade
Cyt: Low-grade;Hist: High-Grade
HPV+;Cyt: Low-grade
Cyt: Low-grade;Hist: Normal
Cyt: AGUS/AGIS
Hist: High-Grade
Hist: Normal
Cyt: High-grade;Hist: High-Grade
HPV+;Cyt: Normal
Cyt: Normal;Hist: Normal
Cyt: High-grade
HPV-
HPV-;Cyt: Normal
HPV+

Portinari Flow Diagrams of Future Paths

The flow diagram starts where the query patterns ends, showing



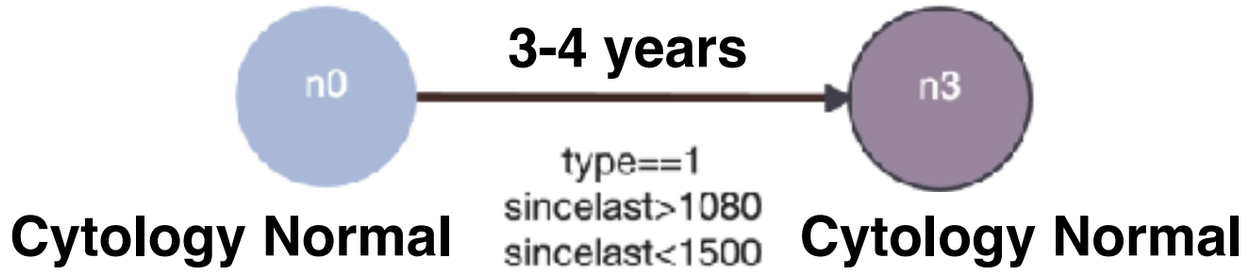
Portinari allows two charts to be displayed side by side!

Hovering over the edges or nodes of the graph shows stats!

Each layer represents the distribution of diagnosis n exams into the future.

Explorations

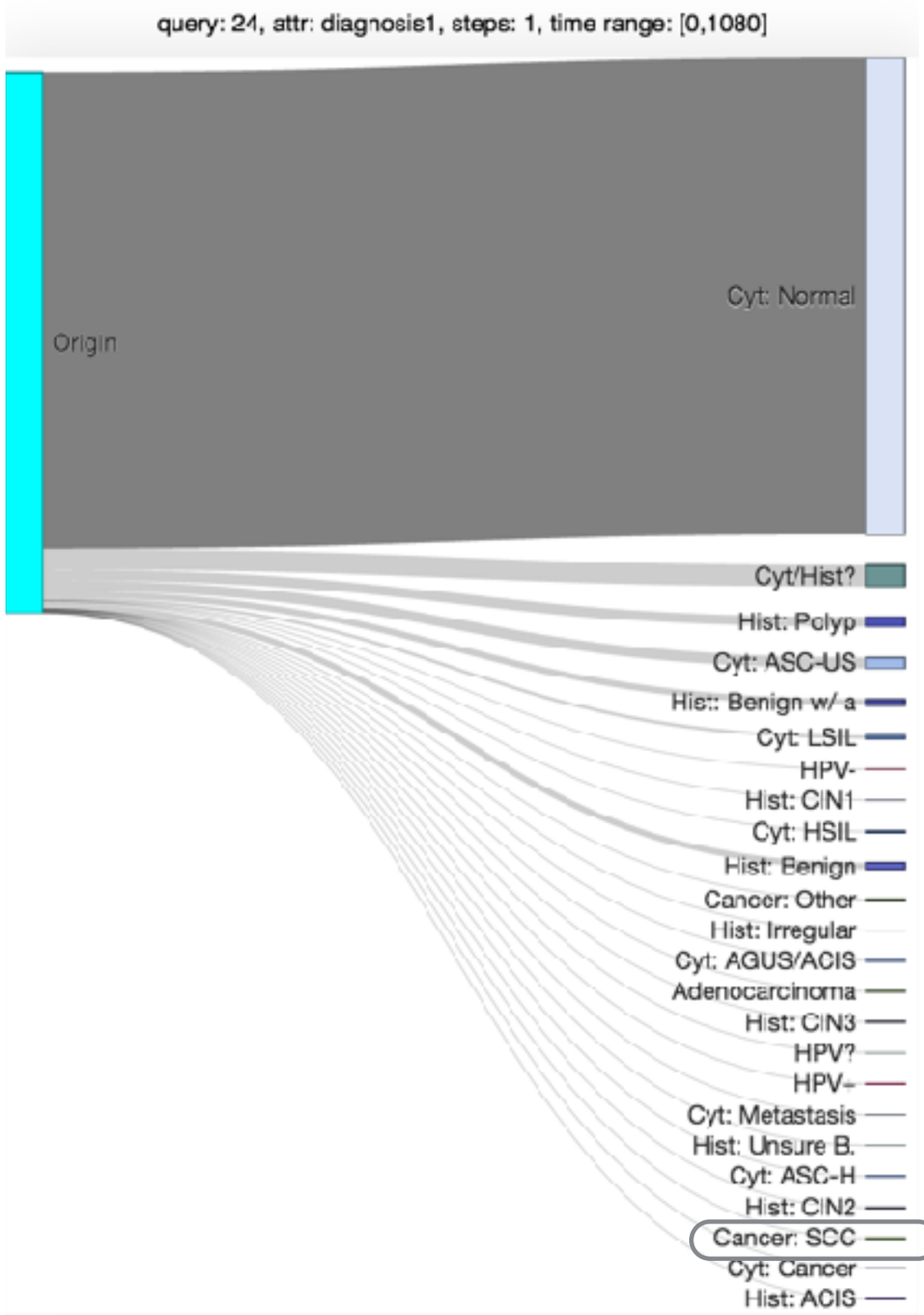
Exploring effects of screening guidelines



What happens after 3 years to this cohort?

Next Diagnosis in 3 years

261133 individuals

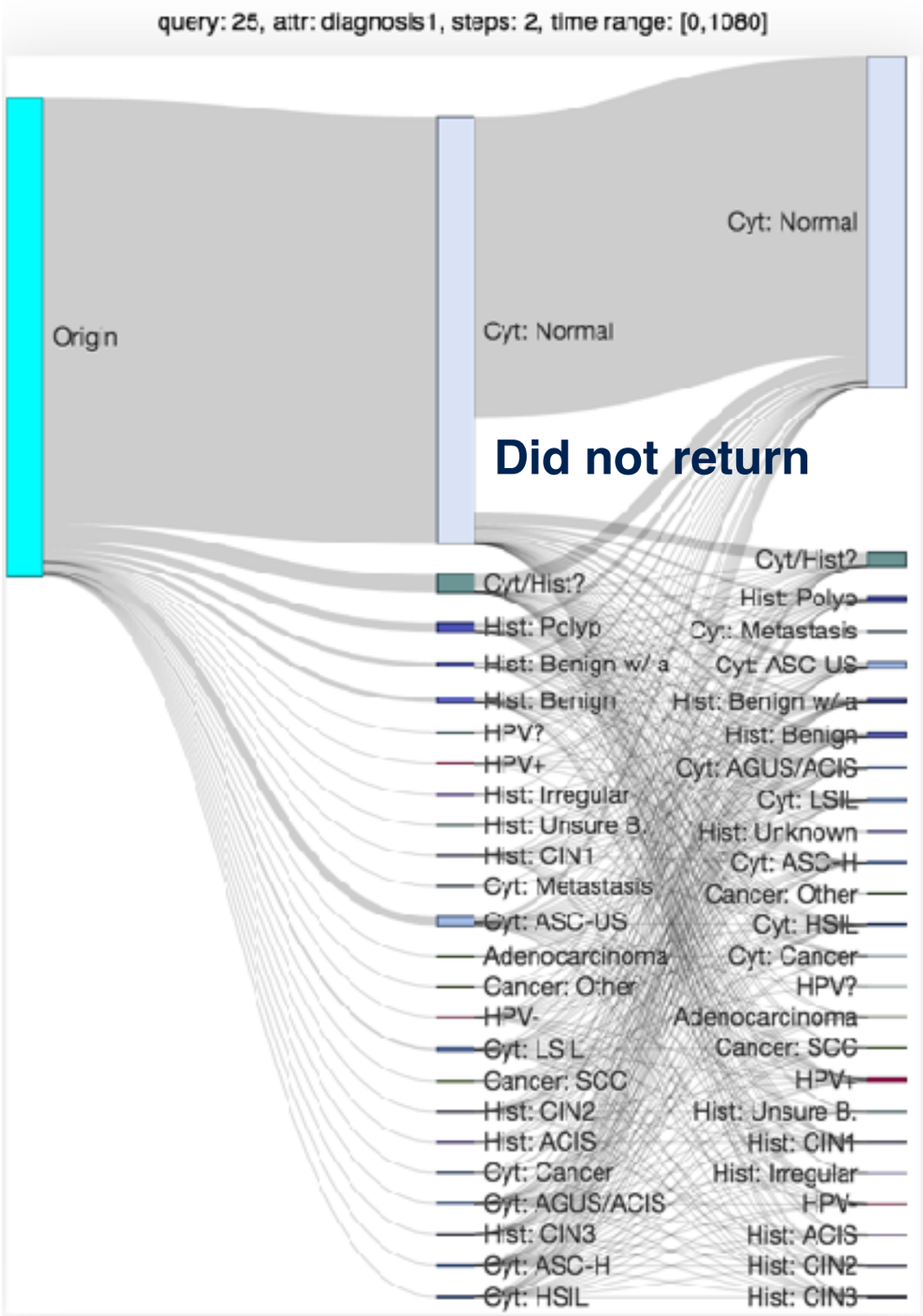


230,331 individuals

39 individuals

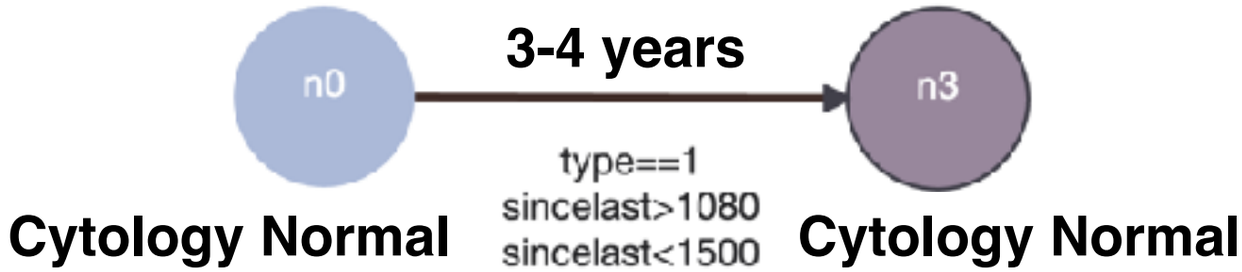
Next Two Diagnoses

280572 individuals



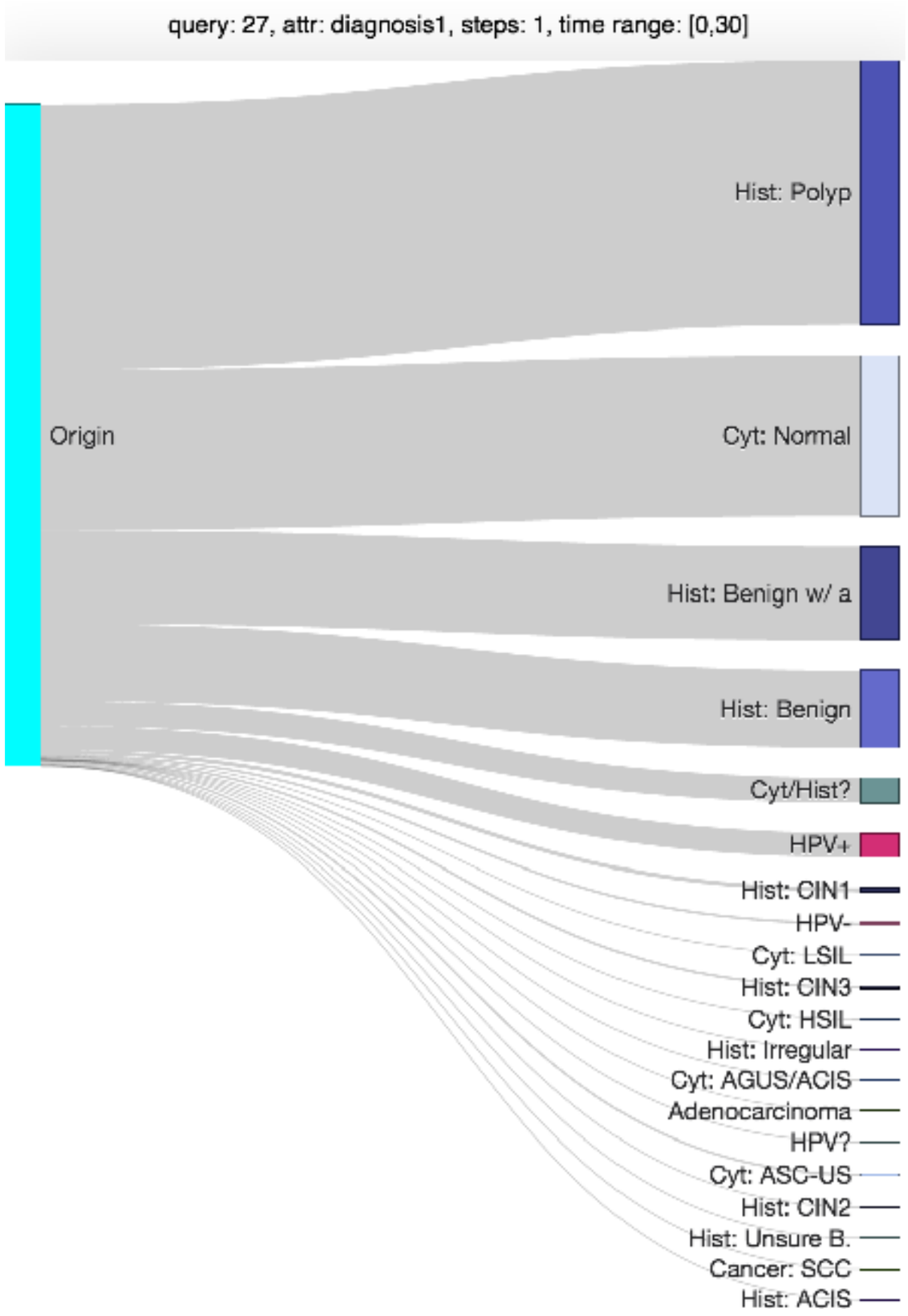
193543 individuals

Exploring existing screening guidelines



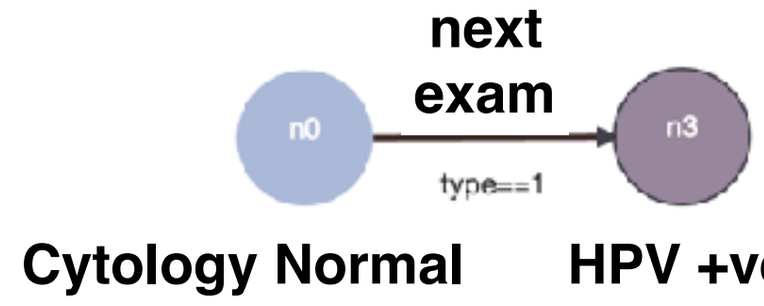
What if women come back in the next 30 days?

4337 individuals



Cytology not good enough for some?

Comparing two scenarios in the triage algorithm

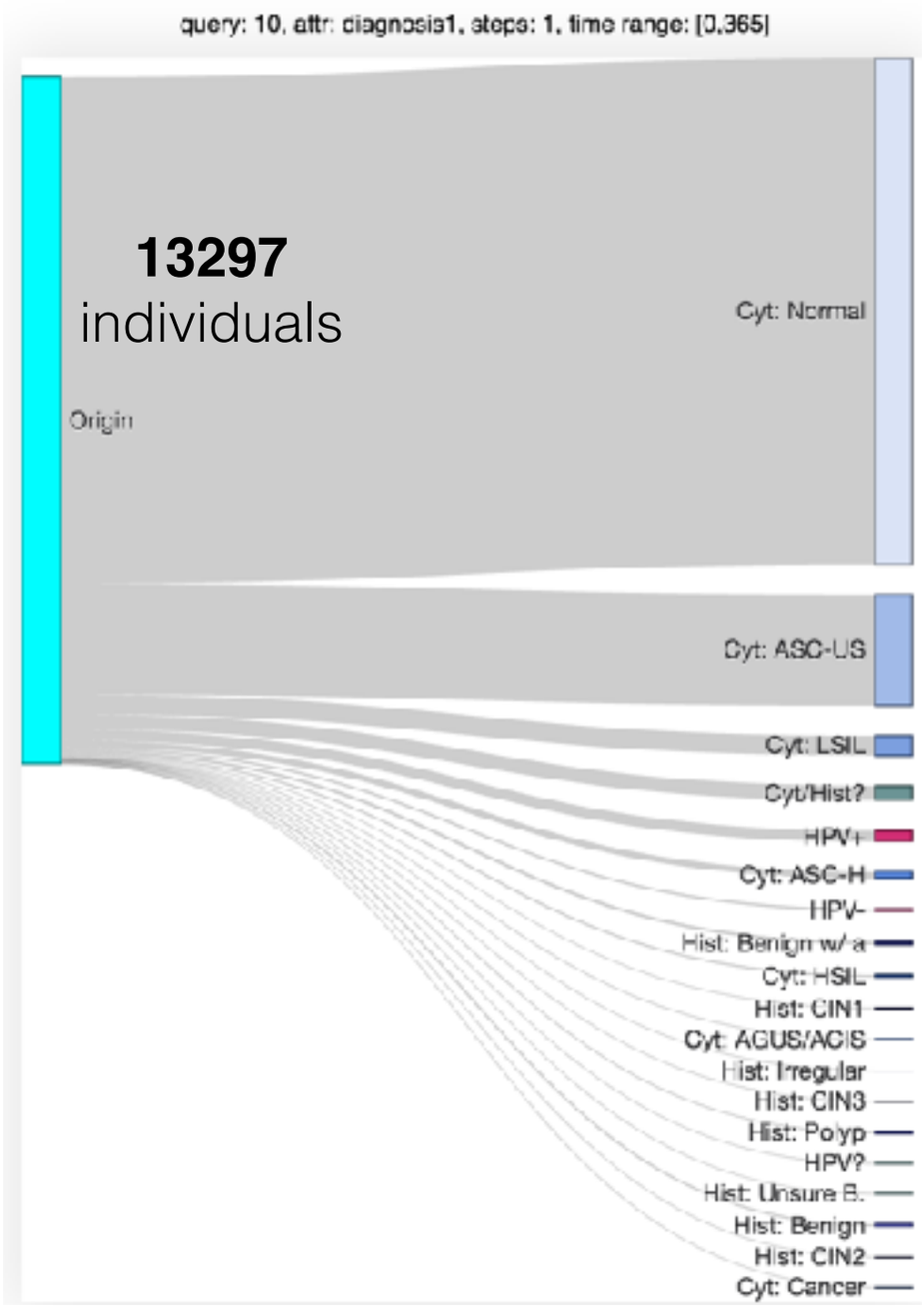


Come back within 1 year (Guideline)

vs.

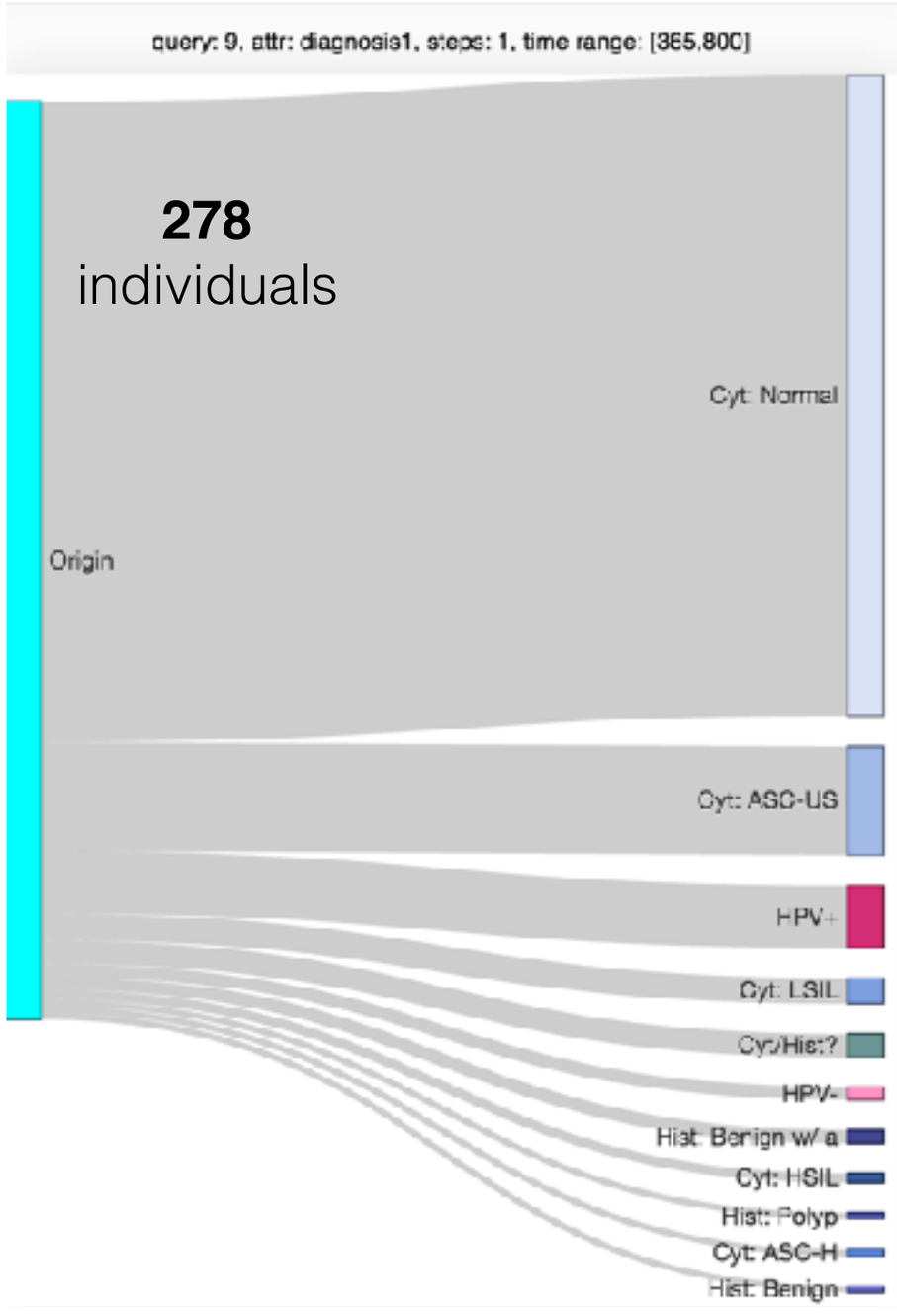
Come back in 1 to 2 years

Returned within 1 year



(227)
1.61%
HPV +ve

Returned in 1 -2 years

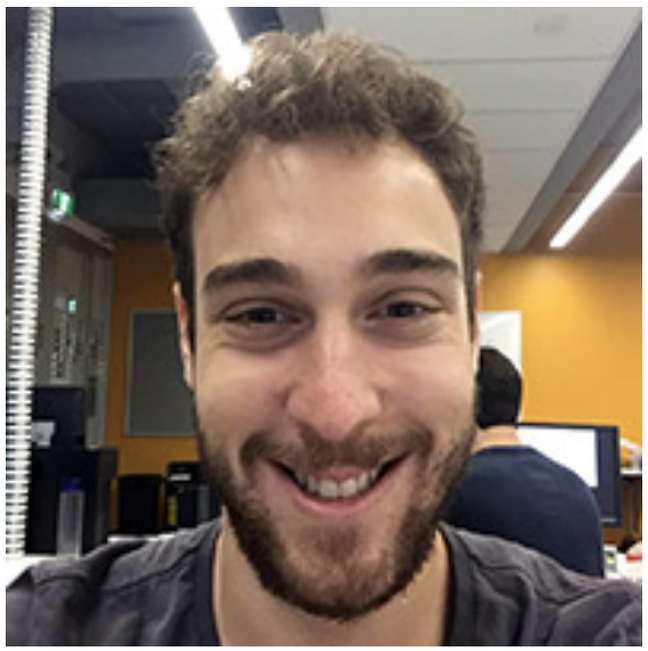


(19)
6.83%
HPV +ve

Conclusion

- Alternative way of verifying and validating paths taken by people in a socio-technical system.
- Use of easy-to-query graph database.
- Generalizable to event-driven data.
- Generation hypothesis for further statistical investigation.

Questions?



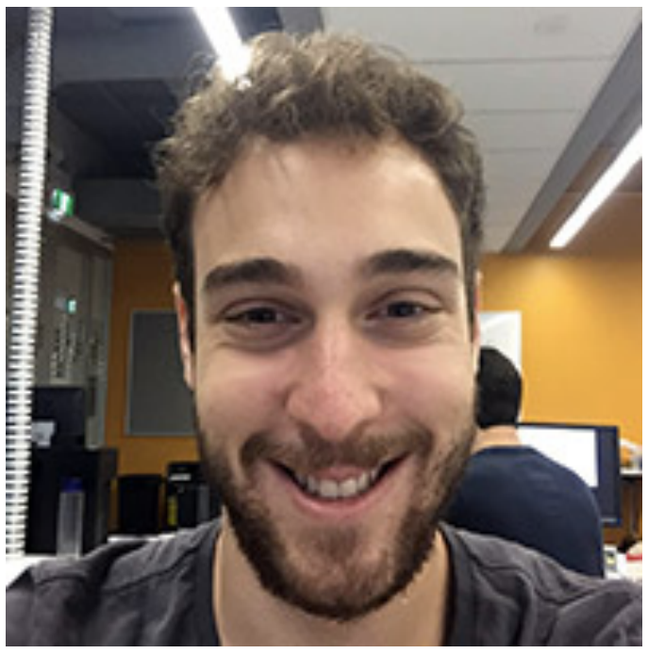
Manoel Horta Ribeiro

https://github.com/manelhr/portinari/

manoelribeiro

homepages.dcc.ufmg.br/~manoelribeiro/

Questions?



Manoel Horta Ribeiro

https://github.com/manelhr/portinari/

manoelribeiro

homepages.dcc.ufmg.br/~manoelribeiro/