Portinari: A Tool for Visualizing Risk in the Cervical Cancer Screening Database
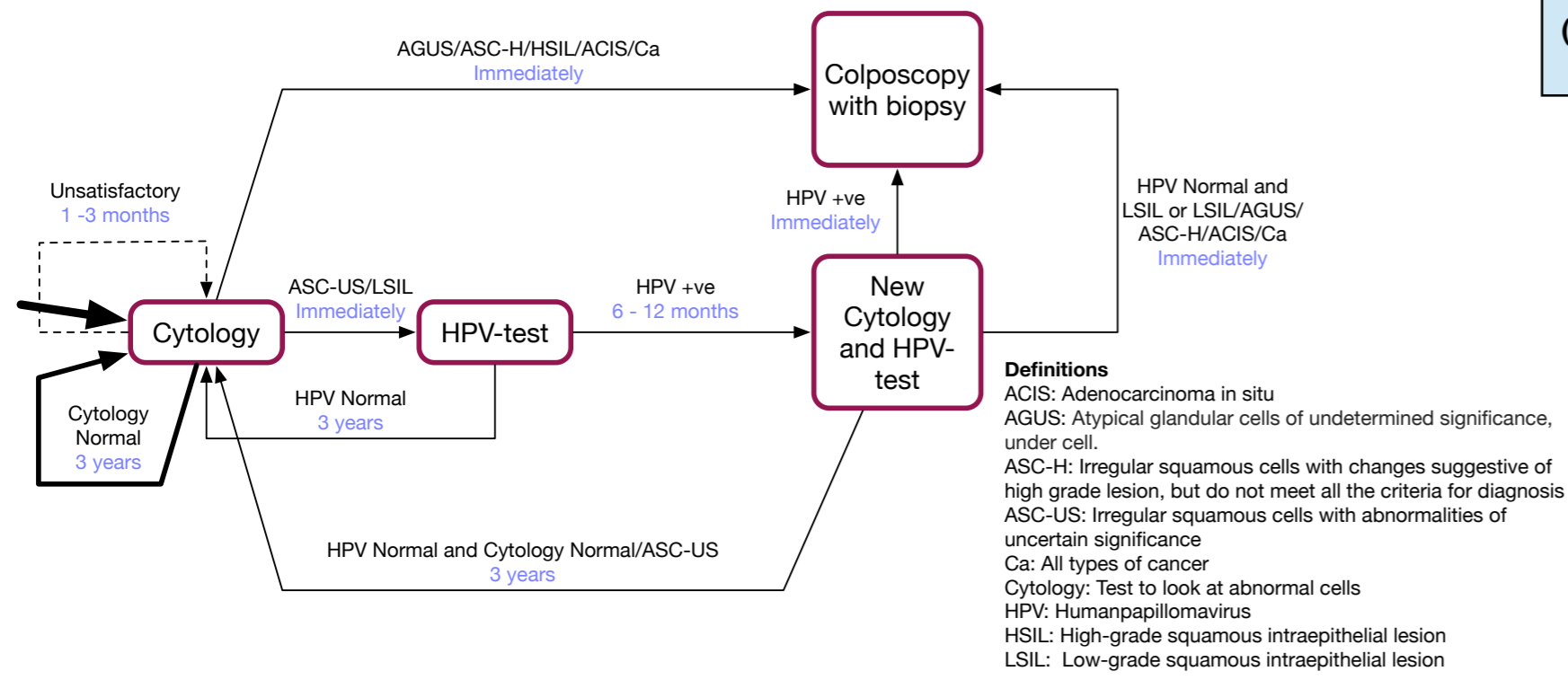
Sagar Sen
HPV Research Group, Cancer Registry of Norway
Certus V&V Center, Simula Research Lab, Oslo, Norway
sagar@simula.no

**Manoel Horta Ribeiro   Raquel de Melo Minardi
Wagner Meira Jr.**
Department of Computer Science, UFMG, Belo Horizonte, Brazil
manoelribeiro@dcc.ufmg.br

Mari Nygård
HPV Research Group, Cancer Registry of Norway
mari.nygard@kreftregisteret.no

Fig 1. Triage algorithm for Cervical Cancer Screening in Norway.

Introduction

- Cervical cancer incidence and mortality rates have significantly decreased in countries that adopted screening programs [1].
- A challenge faced by these programs is to simultaneously minimize unnecessary screenings [2] and the population's risk of developing cervical cancer [3].
- The usual approach is the development of *screening guidelines*. For example Norway's Cervical Cancer Triage Algorithm is depicted in **Fig 1.**
- However, screening guidelines for cervical cancer (and others) don't take into account many predictive factors [4], and in practice, few people follow the recommendations.
- We present **Portinari**, a tool intended to assess screening guidelines and communicate risk through intuitive diagrams. We validate it analyzing cervical cancer screening data collected from *906* thousand women in Norway, as shown bellow:

ID	birthdate	date	exam type	diagnosis	censor date
Internal numeric identifier	Month and year of birth	Date of diagnosis	Type of visit/exam: cytology, histology, HPV test	Result of the exam	date of emigration/ death/cervical cancer diagnosis

Graph Databases

- Screening exams and their respective diagnosis are usually stored as a record in a *relational database*.
- *Queries* of interest are often time related, and thus hard to write in relational databases.
- We advocate using a **graph database** [5] to represent this temporal ordering.
- Querying becomes a **subgraph matching problem**, which is harder computationally, but much easier to express. Furthermore, our system allows the user to express the subgraph to be matched in an easy *drag-and-drop* fashion.
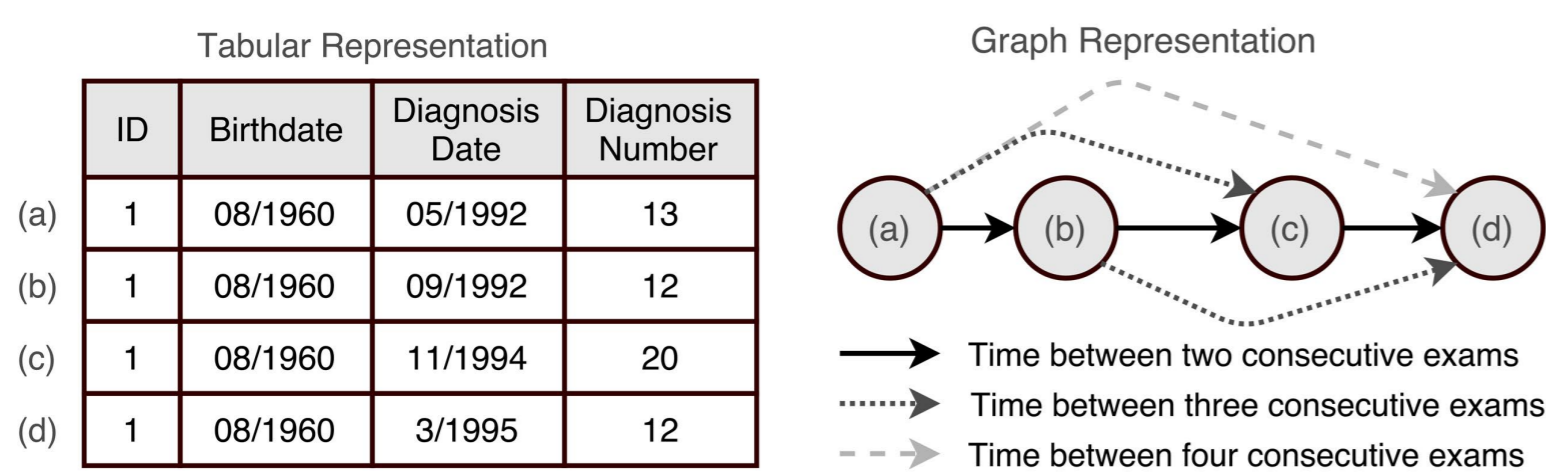
Fig 2. Transformation from tabular schema to a graph schema. Extra edges help the querying process.

Querying

- A query graph is built using the drag-and-drop interface in shown in **Fig 3 a**, and the forms shown in **Fig 3 c-d**.
- The user creates the topology of the graph using nodes and edges in **(a)**. Each nodes represent an exam to be matched, while the edges give the trajectory order.
- The user specifies the attribute to be seen in the next exams, how many exams be explored ahead of the matching and a time range for the return in **(d)**.
- **Portinari** matches all patients that had the specified *query graph* in their very first exam.

Matching

- We visualize the trajectories of the patients who matched the *query graph* using a flow diagram, as shown in **Fig 3 e-f**.
- The individuals who matched the query are grouped in the **Origin** node, and then, reading left to right, one may see the development of the individuals.
- **Did not return** represents people that didn't return within to the specified return times, or at all.
- **Other** contains trajectories too specific to be shown in the sankey diagram without making it unreadable.

[1] M. Arbyn, A. Anttila, J. Jordan, G. Ronco, U. Schenck, N. Segnan, H. Wiener, A. Herbert, and L. Von Karsa, "European guidelines for quality assurance in cervical cancer screening." *Annals of Oncology*, vol. 21, no. 3, pp. 448–458, 2010.
[2] M. A. Nobbenhuis, J. M. Walboomers, T. J. Helmerhorst, L. Rozendaal, A. J. Remmink, E. K. Risse, H. C. van der Linden, F. J. Voorhorst, P. Kenemans, and C. J. Meijer, "Relation of human papilloma virus status to cervical lesions and consequences for cervical-cancer screening: a prospective study," *The Lancet*, vol. 354, no. 9172, pp. 20–25, 1999.
[3] S. J. Goldie, L. Kuhn, L. Denny, A. Pollack, and T. C. Wright, "Policy analysis of cervical cancer screening strategies in low-resource settings: clinical benefits and cost-effectiveness," *Jama*, vol. 285, no. 24, pp. 3107–3115, 2001.

(a) A drag and drop canvas to create the query graph. **(b)** Show constraints on specific node or edge. **(c)** Create constraints on properties. **(d)** Allows to specify a property, the number of events to show in the future and the time range. **(e-f)** Sankey diagram with patients trajectories.
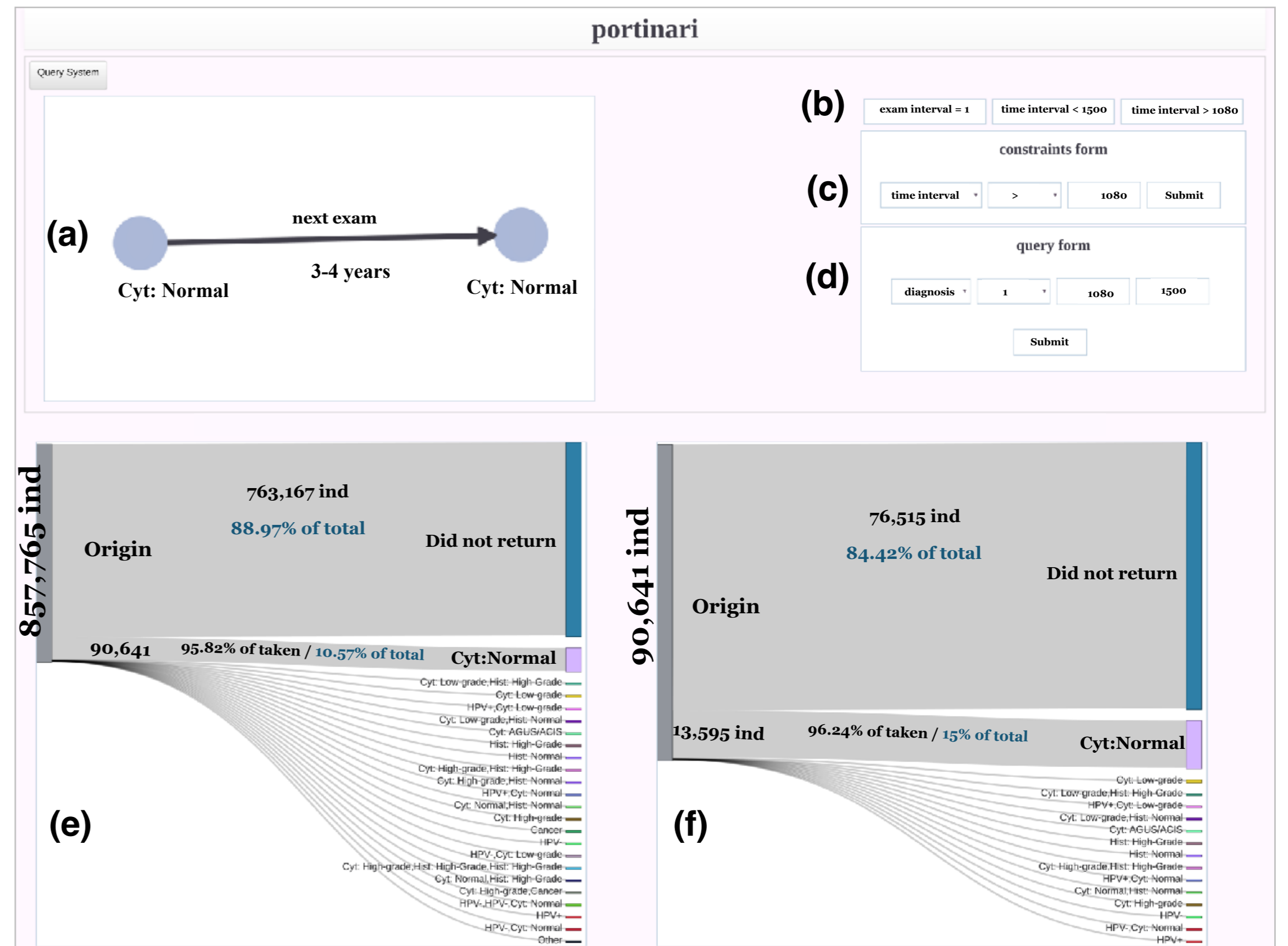

Fig 3. System's overview. The **taken** exams accounts for the patients who returned, while the **total** accounts for all matched patients.
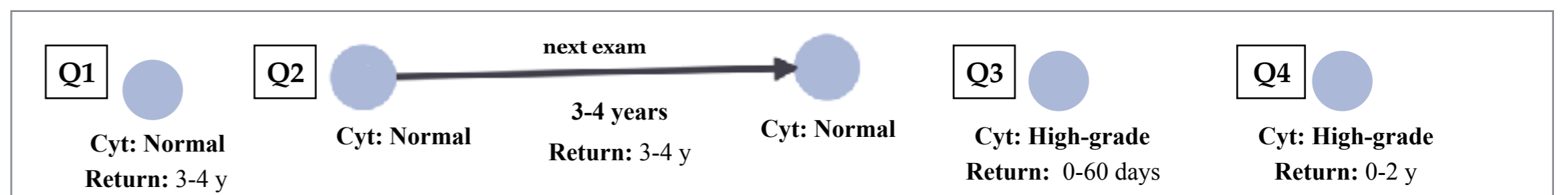

Fig 4. Queries made in the exploration section.

Explorations

- We hope that exploring patient's trajectories may give insight into how to personalize the Cervical Cancer Screening Program in Norway.
- We use **Portinari** to explore the differences between people that follow the guidelines and people who do not. This is essential to understand how effective are the guidelines.
- We now discuss the queries shown in **Fig 4**:
- **Fig 3 e, f** shows results for *Q1-2* respectively.
  - There is a increase on the percentage of individuals diagnosed with **Cyt:Normal** on *Q2* if compared to *Q1* (from 95.82% to 96.24%).
  - This trend continues. If individuals follow the screening algorithm for three, four or five iterations, always getting diagnosed normal, their chances of having being diagnosed normal upon return goes to 96.81%, 97.3% and 99%, respectively . The flow diagram of those are omitted for reasons of space.
- **Fig 5 a, b** shows results for *Q3-4* respectively.
  - People don't follow the guidelines. **Fig 1**, instruct patients to immediately return for a histology after a ASC-H/ HSIL diagnosis (**Cyt:High-Grade** in our program). However, as shown in **Fig 5 a**, most of them do not.
  - Patients who return within 60 days as recommended have worst outcomes than those who return later. They are diagnosed:
    - less w/ **Cyt:Normal**: 7.8% v. 16.7%
    - more w/ **Hist:High-Grade**: 65.6% v. 44.4%
    - more w/ **Cancer**: 1.47% v. 0.9%
  - This may be due to individuals perceiving symptoms, and returning because of that.

Fig 5. Results for queries 3 and 4.

Conclusion

- We presented **Portinari**, a tool that allows its users to visually query and understand the trajectories of patients in the screening program.
- We explained how to formulate query and to interpret the results given by the tool.
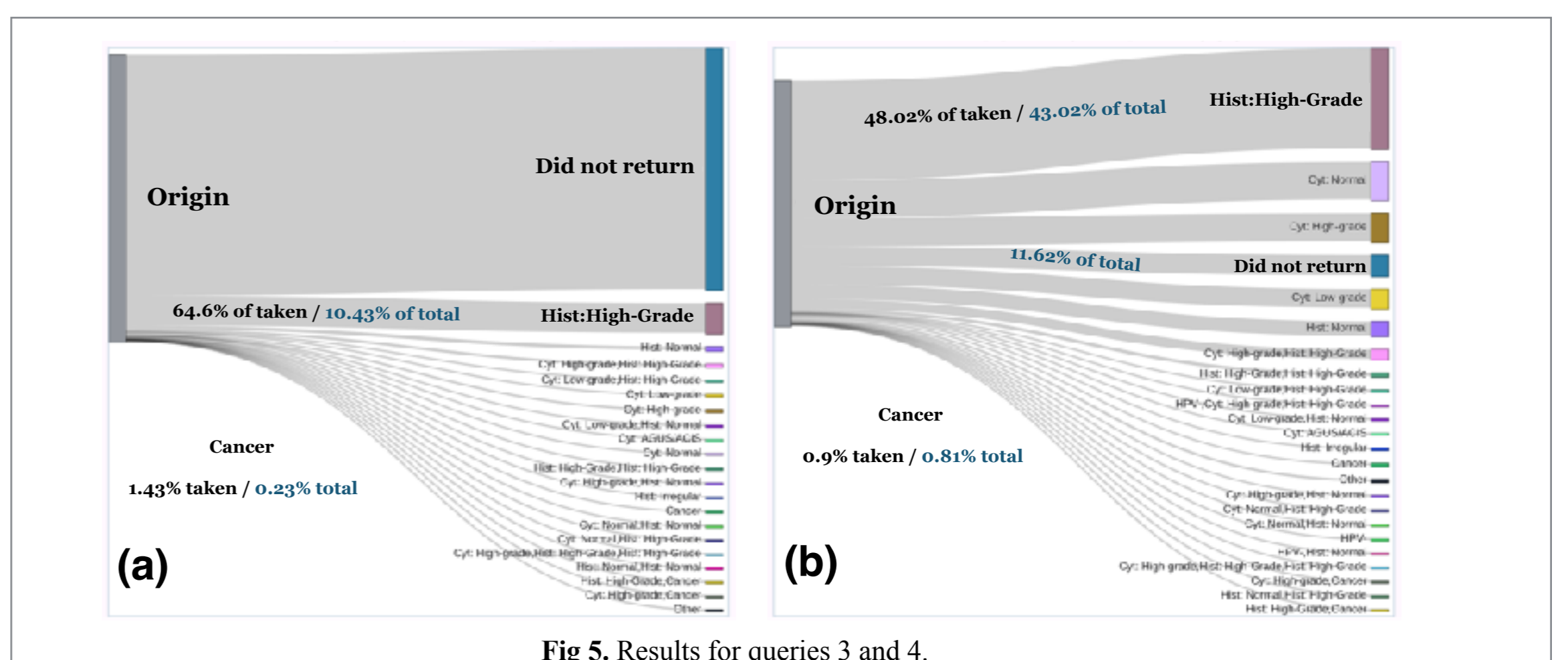- We presented several queries that showed how Portinari can aid to understand personalized scenarios, enabling researchers to evaluate guidelines and gain insights.