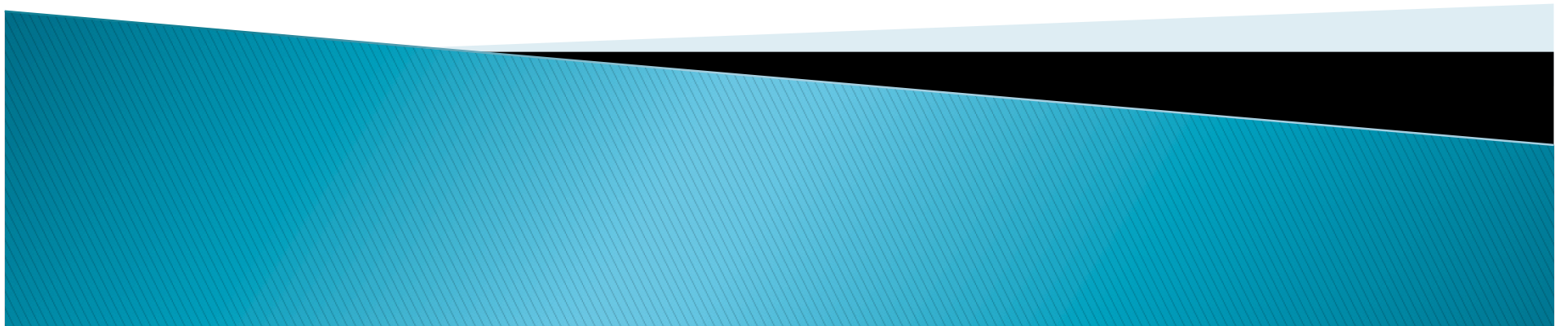


Visualising words

Uncovering hidden information in large UK primary care databases

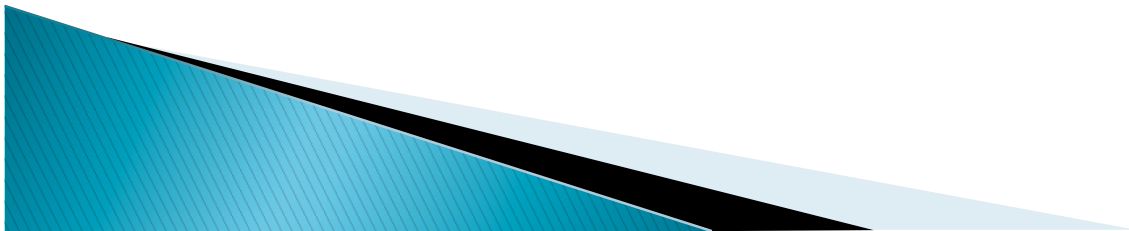
.

Rosemary Tate on behalf of the "Patient Records Enhancement Project", Brighton and Sussex Medical School and Department of Informatics at University of Sussex,

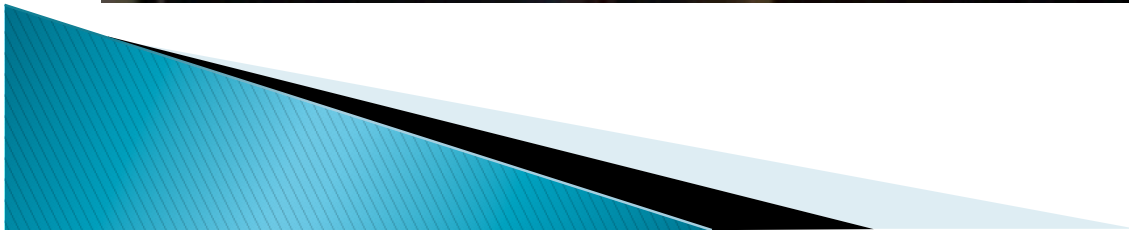


Structure of talk

- ▶ The role of General Practice in the UK
- ▶ The General Practice Research Database
- ▶ Why GPRD researchers need visualisation tools
- ▶ Example - ovarian cancer study
- ▶ Open questions



The UK National Health Service: Role of General Practitioner



General practitioner (GP) systems in the UK

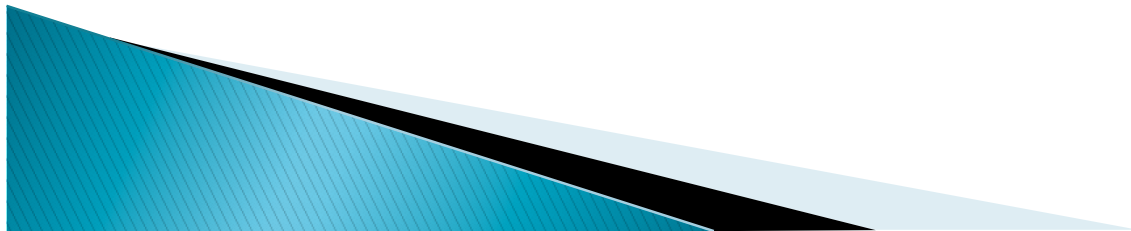
- ▶ Have been recording electronically since the early 1980s
- ▶ Systems provide a huge amount of flexibility for recording
- ▶ Many different (Read) codes for the same thing
- ▶ May also use free text



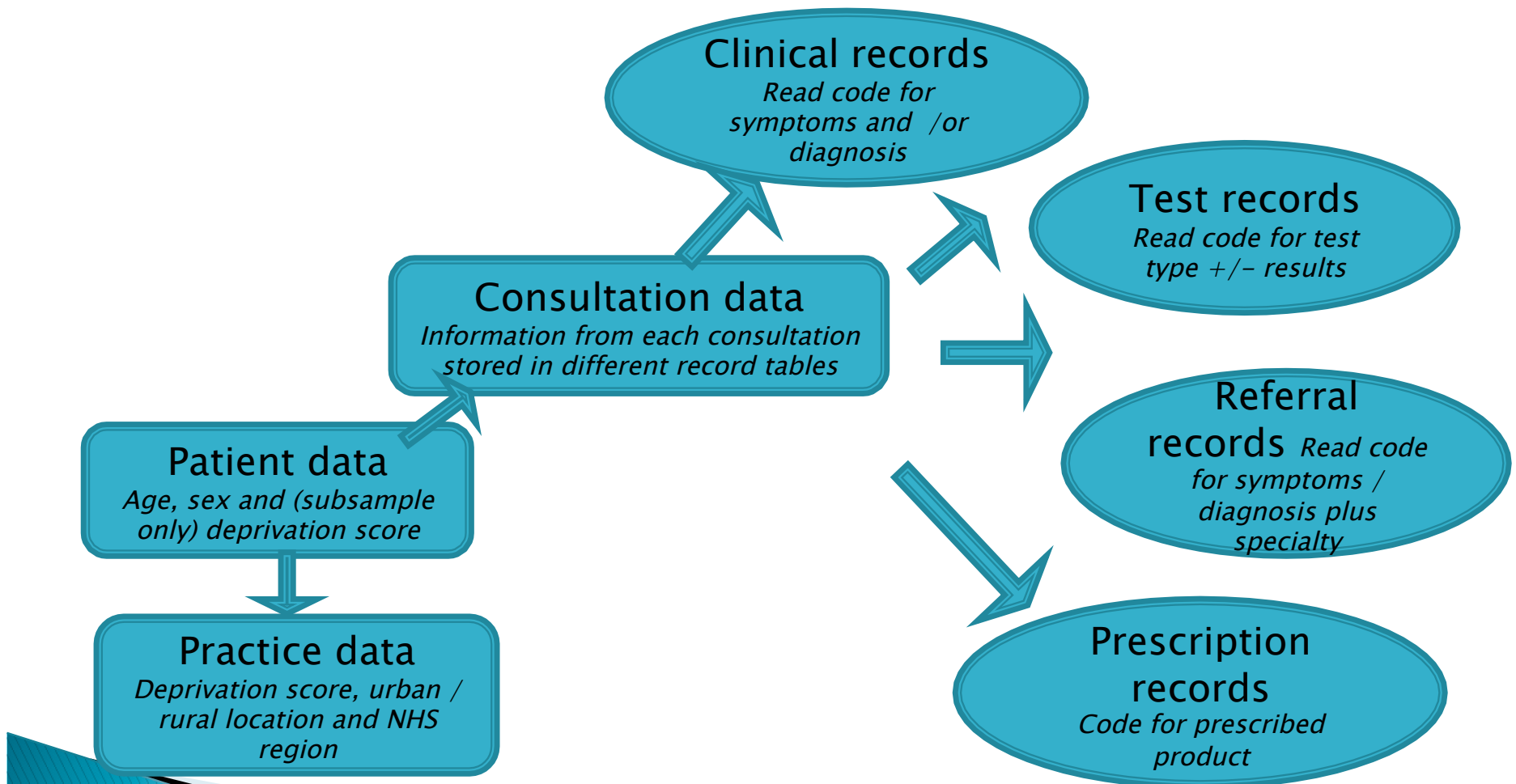
General Practice Research

Database (GPRD) <http://www.gprd.com>

- ▶ Collections from 620 practices (~7% of UK practices)
- ▶ Contains over 12.5 million patients
- ▶ 5 million active (alive) patients (~8% popn.)
- ▶ Contains over 62 million patient-years of follow up
- ▶ Geographically representative of UK

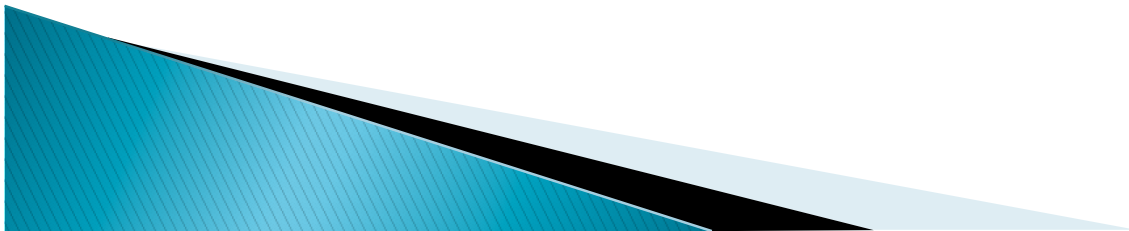


Structure of GPRD database



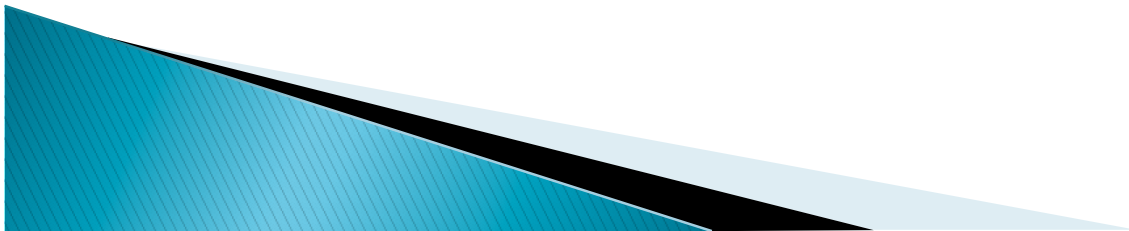
Why GPRD researchers need visualisation tools

- ▶ Track for early indicators of diagnoses
- ▶ Check data quality
- ▶ Develop models of patients' journey through cancer
- ▶ Define definitions of 'cases'
- ▶ Annotate and edit the data; e.g., add causal links



Example: Ovarian cancer study

- ▶ Investigating symptoms and delay in diagnosis
- ▶ 344 patients with ovarian cancer
- ▶ Using data from the General Practice research database



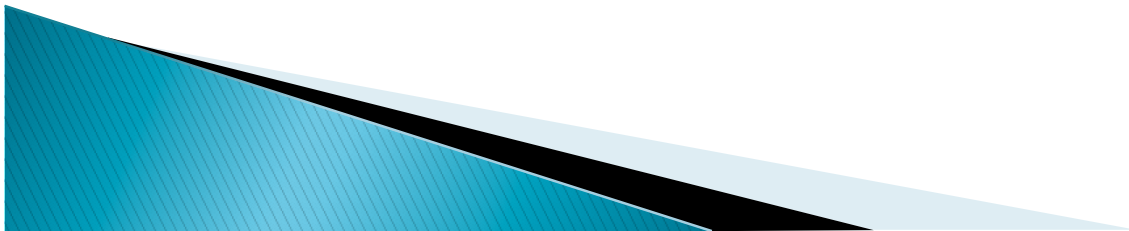
Ovarian cancer study

Compared incidence of 5 most common symptoms in the year prior to diagnosis using

- ▶ 1. codes
- ▶ 2. free text (notes and letters) – mapped symptoms in text to “pseudo” Read codes

Results: Incidence doubled for some symptoms when we used the free text

Challenge – how to extract and present this extra information?

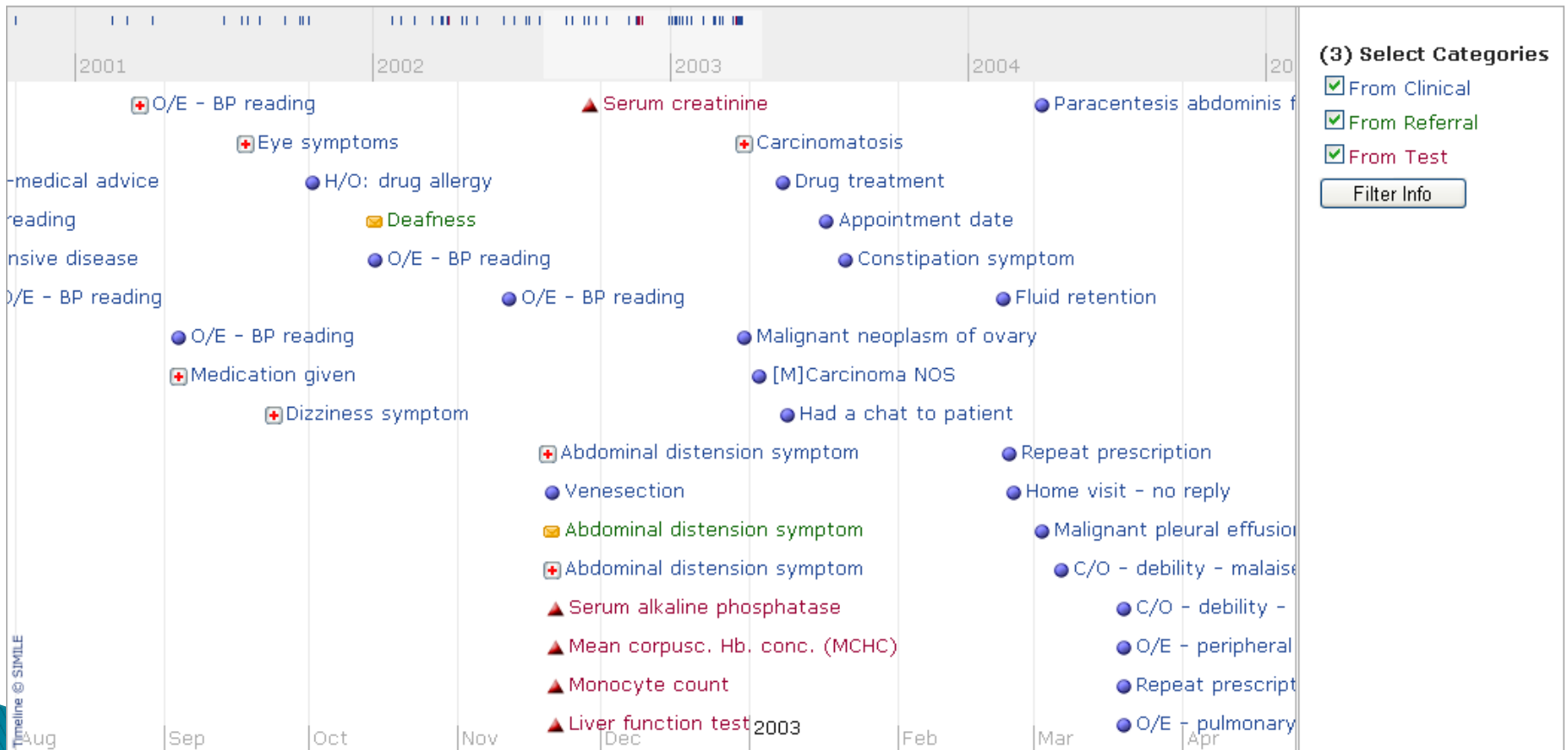


Typical patient record shown on a timeline

(1) Select Dataset

(2) Select Patient OR Enter PatientID

Filter: Highlight:



Abdominal pain

Patient ID: 6432912

Description: Abdominal pain

Category: From Clinical

Free Text: ~ bloating in 90's && relapsed last 1/52 - worse after meals , no wind. BO daily - now alternating , no rectal bleeding , gen health app + wt ok. No dyspepsia , no dysuria , only 2 K this year. FH - no Ca bowel / Ovary o / we abdo - soft no tender no LKKS. Imp IBS - tca if persists

Thu, 25 May 2006 00:00:00 GMT

us - acquired never smoked tobacco
us - acquired  Abdominal pain

Icon indicating the presence of Free Text for event

SOURCES & TOOLS

CONTACT

LO

Load Patient

00

2010

2020

2007

 Digestive system diseases NOS

Mapped text to “pseudo” Read codes

(1) Select Dataset:

(2) Select Patient: OR Enter PatientID:

Filter: Highlight:

(3) Select Categories & Subcategories

- ReadCodes
- PseudoCodes Keywords
- PseudoCodes Annotation

- Abdom. distension
- Abdom. mass
- Abdom. pain
- Appetite/Weight
- Back pain
- Breathing Probs
- Constipation/CBH**
- Indigestion
- Nausea/Vomiting
- Pelvic pain
- Tiredness
- Urinary/genital
- No Subcategory

Follow steps 1, 2, 3 to view patient information on the timeline.

Click and drag one of the three timeline bands (decade/year/week) horizontally to scroll events.

Icon indicates the presence of free text information for the event. Click event to view event information.

Icon indicates negation.

LifeLines2 to show differences between free text and coded information

The screenshot displays the LifeLines2 software interface, which is used for comparing free text and coded information over time. The main window is titled "LifeLines 2" and has a menu bar with "File", "Jump To...", and "About". Below the menu bar are tabs for "Record View", "Comparison View", and "Group View". The "Comparison View" is currently selected.

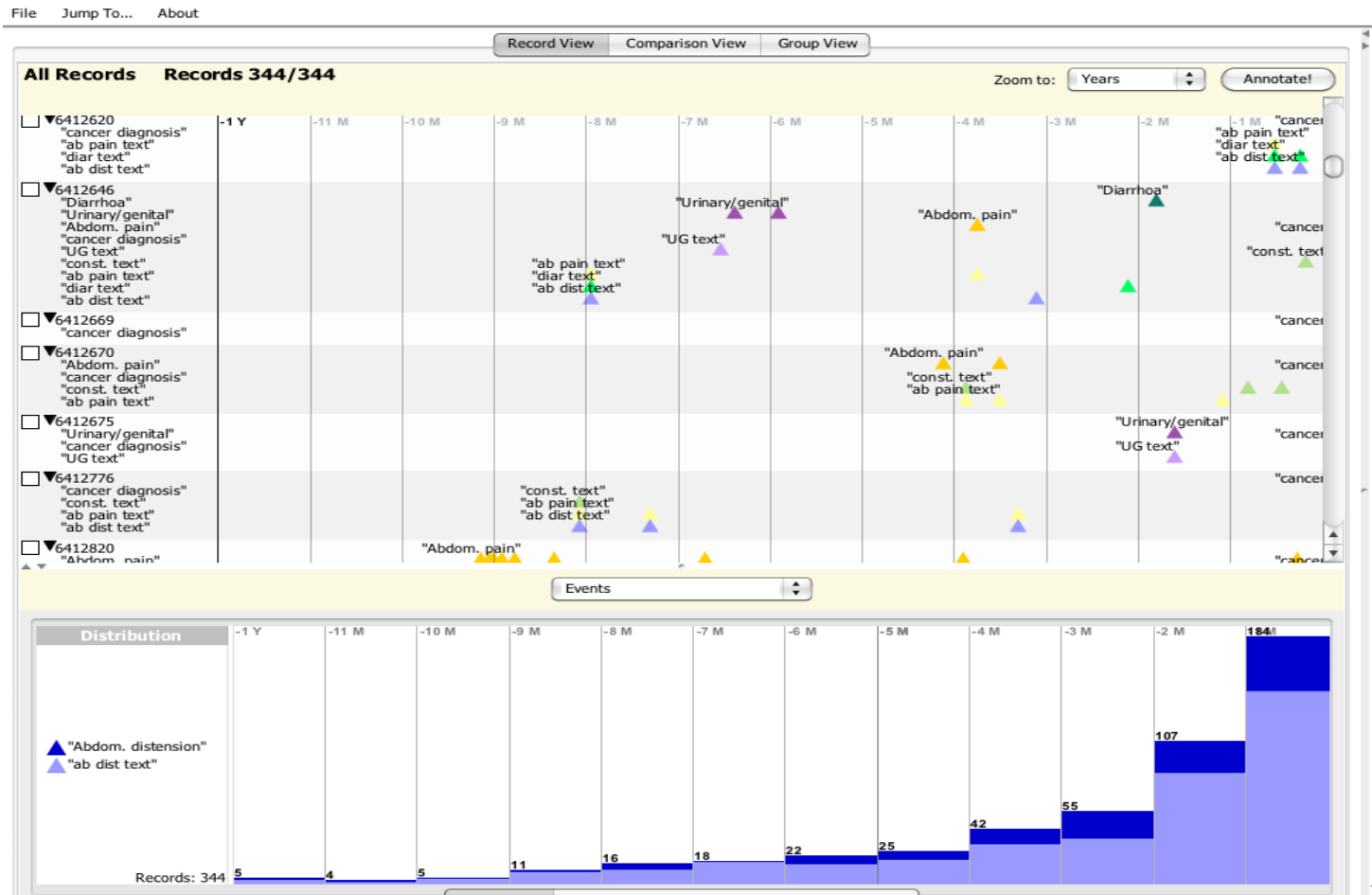
The main area shows a list of records on the left, each with a checkbox and a list of associated text. The records are sorted by their occurrence date. The timeline at the top is labeled "Zoom to: Years" and has a scale from -11 M to -1 M. The records are plotted on this timeline, with free text represented by triangles and coded information by squares. The records are:

- 6400008: "cancer diagnosis", "ab dist text"
- 6400405: "Abdom. pain", "cancer diagnosis", "ab pain text", "diar text", "ab dist text"
- 6400435: "Abdom. distension", "cancer diagnosis"
- 6400547: "cancer diagnosis"
- 6401148: "cancer diagnosis", "diar text", "ab dist text"
- 6401560: "cancer diagnosis", "ab pain text", "ab dist text"
- 6401896: "cancer diagnosis", "ab dist text"
- 6401956: "Abdom. pain", "cancer diagnosis"
- 6402062: "cancer diagnosis", "ab pain text", "ab dist text"
- 6402200: "Abdom. pain", "cancer diagnosis"
- 6402341: "cancer diagnosis", "ab pain text", "ab dist text"
- 6402453: "cancer diagnosis", "ab pain text", "ab dist text"
- 6402939: "Abdom. pain", "cancer diagnosis", "ab pain text"
- 6402948: "Abdom. pain", "cancer diagnosis", "ab pain text", "ab dist text"
- 6403220: "Diarrhoa", "cancer diagnosis"

On the right side, there is a control panel with tabs for "ARF", "Comparison", and "Groups". The "Comparison" tab is selected. It includes sections for "Align by..." (set to "cancer diagnosis"(344) with "1st occurrence" selected), "Rank by..." (set to "Record ID"), and "Filter by..." (set to "No Filter"). Below these are buttons for "Keep Selected", "Remove Selected", "Reset Filters", "Collapse all", "Save Records as New Group", "Add Records To...", "New Group...", "Show Group Membership", and "More Controls". At the bottom of the control panel, there are tabs for "Events", "Icons", "Labels", and "Alignment". The "Events" tab is selected, showing a list of events with checkboxes and icons: "Urinary/genital", "Abdom. distension", "Abdom. pain", "cancer diagnosis", "First suspicious/referral", "Constipation/CB+*", "UG text", "const. text", "ab pain text", "diar text", and "ab dist text". An "Edit Events" button is at the bottom of this list.

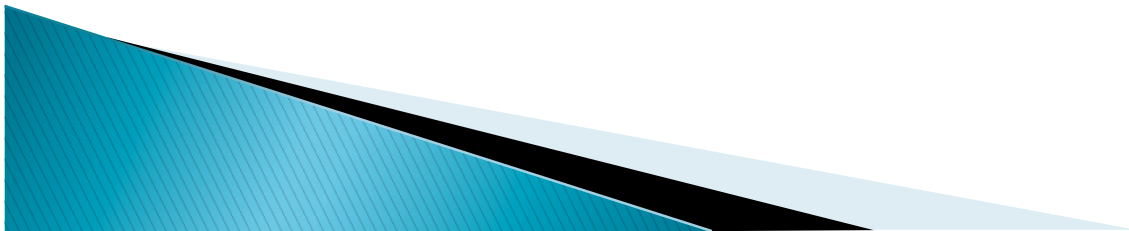
At the bottom right of the window, the copyright information is displayed: "Copyright 2007 - 2010 HCIL UMD".

LifeLines2 to show differences between free text and coded information



Open questions

- ▶ How to present information at different levels of detail – show both the big picture and the individual records
- ▶ How to present information hidden in the free text?
- ▶ Provide for user interaction
- ▶ Accommodate for a range of user tasks



Acknowledgements

Aishath Ali¹

John Carroll²

Fliss Henwood³

Donia Scott²

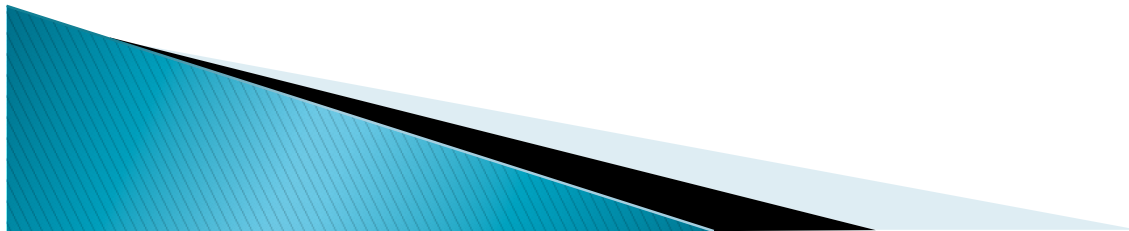
Lesley Axelrod²

Jackie Cassell¹

Rob Koeling²

Tim Williams⁴

- ▶ ¹Division of Primary care, Brighton and Sussex Medical School
- ▶ ² Dept of Informatics, University of Sussex
- ▶ ³ Applied Social Sciences, University of Brighton
- ▶ ⁴ General Practice Research Database, MHRA, London



The Team

