

Moving beyond RCTs: Tapping the potential of Real-World Data for advancing clinical research and improving post-treatment in breast cancer

Ioannis Sarafis

Multimedia Understanding Group (MUG)
Department of Electrical and Computer Engineering,
Aristotle University of Thessaloniki, Greece

sarafis@mug.ee.auth.gr

16 September 2021



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 965231.

The REBECCA project

- REBECCA: REsearch on BrEast Cancer induced chronic conditions supported by Causal Analysis of multi-source data
- **Overarching goal:** to improve the quality of life of breast cancer patients or former patients
 - **Clinical Research:** Analyze data to better understand how treatment affects quality of life
 - **Patient Management:** Improve interventions or care choices at the individual level



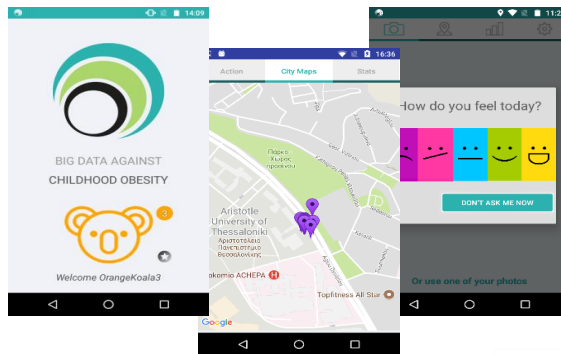
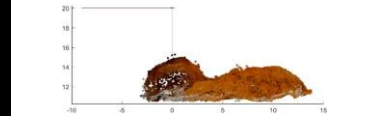
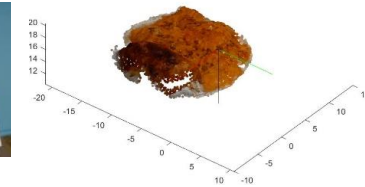
The REBECCA consortium



Moving beyond RCTs: Tapping the potential of Real-World Data

16 September 2021

Multimedia Understanding Group



Complex Chronic Conditions in breast cancer

- Cancer has several associated comorbidities and disabilities
- A patient can suffer from more than one comorbid condition
- In REBECCA, we focus on:
 - Breast cancer treatment-induced **peripheral neuropathy**
 - Breast cancer-related **fatigue**
 - Adjuvant treatment-induced **osteopenia/osteoporosis**



The baseline: Randomized Control Trials

- RCTs are the current Gold Standard in clinical research
- Powerful tools to decide whether an intervention is effective
 - **Causality** = Intervention causes the effect /improvement

Methodology:

- Formulate a hypothesis about a causal relation
- Assign participants to interventions or control group
 - Ideally, the statistics across the groups should be the same
- Collect prospective data



Limitations of RCTs in studying CCCs

- **Combinatorial explosion** due to large number of factors
 - Patients may suffer from n different conditions (not necessarily independent)
 - Conditions may depend on m variables
 - Then, assuming binary variables, there are 2^{n+m} configurations that we would ideally need to test!
 - (even worse if variables have multiple levels)



Limitations of RCTs in studying CCCs (2)

- **Gap** between RCTs and clinical practice
 - RCTs include a small population sample and thus a **subset of relevant variables**
- **Feasibility** of RCTs
 - Researchers do not have control over the independent variables.
 - For example, there is no control over the progression of the comorbid conditions



Real World Data (RWD)

- RWD are **observational** data
- RWD were generated and collected for purposes **other than research**
- Examples:
 - Patient diagnoses
 - Prescriptions
 - Examination records
 - Visit records
 - Patient discharge notices
 - etc.



RWD compared to RCTs

Real World Data

- Observational data
- Uncontrolled conditions
- May exist already in large volumes
- Large number of subjects
- Can include population samples with various characteristics

Randomized Control Trials

- Experiments
- Controlled interventions
- Must be collected for each RCT
- Typically, smaller samples
- Limited to the characteristics of the included participants



RWD in REBECCA



- Electronic health records



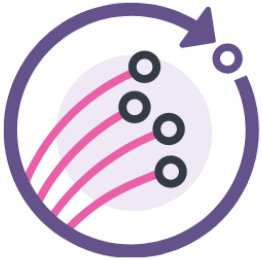
- Wearable devices (smartwatch, fitness bands)
 - Daily activity (e.g., steps)
 - Automatic eating detection
 - Sleep quality and duration
 - Locations visited / Movement / Life-space
 - Pictures of food



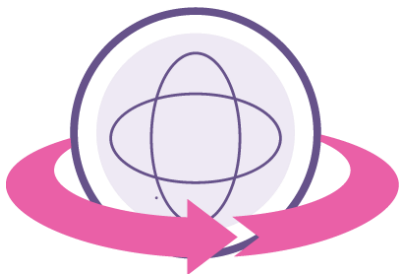
- Mobile apps
 - Short and frequent questionnaires answered by the patient
 - Questionnaires answered companion of the patient (e.g., spouse, friend)



RWD in REBECCA



- Web browser plugin
 - Website types
 - Text of post in social media
 - Search queries



- Environment of the patient
 - Characteristics of the local urban environment
 - GIS and map data (e.g., OpenStreetMap)
 - Satellite data (e.g., Copernicus)
 - Data from Statistical Authorities
 - Home/work environment



RWD to indicators

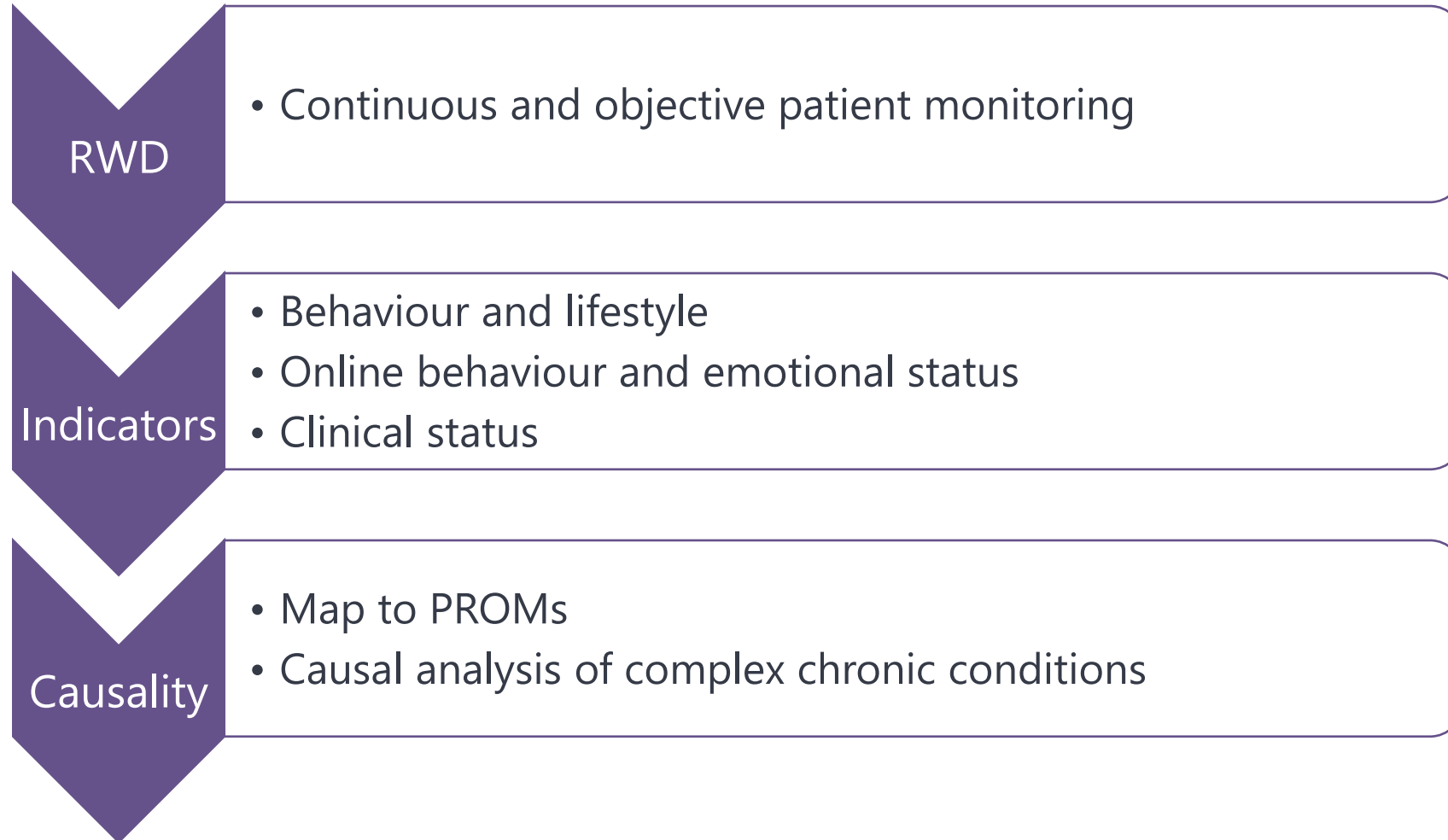
- Collected RWD are processed to extract **indicators**
 - Higher level of information

Examples:

- Prescriptions → Symptoms
- Smartwatch accelerometer → Meal detection → Eating schedule variability
- Smartwatch accelerometer → Physical activity → Exercise sessions
- Smartphone GPS → Detection of visits → Estimate time spend outside
- Search queries → Sentiment analysis → Emotional status

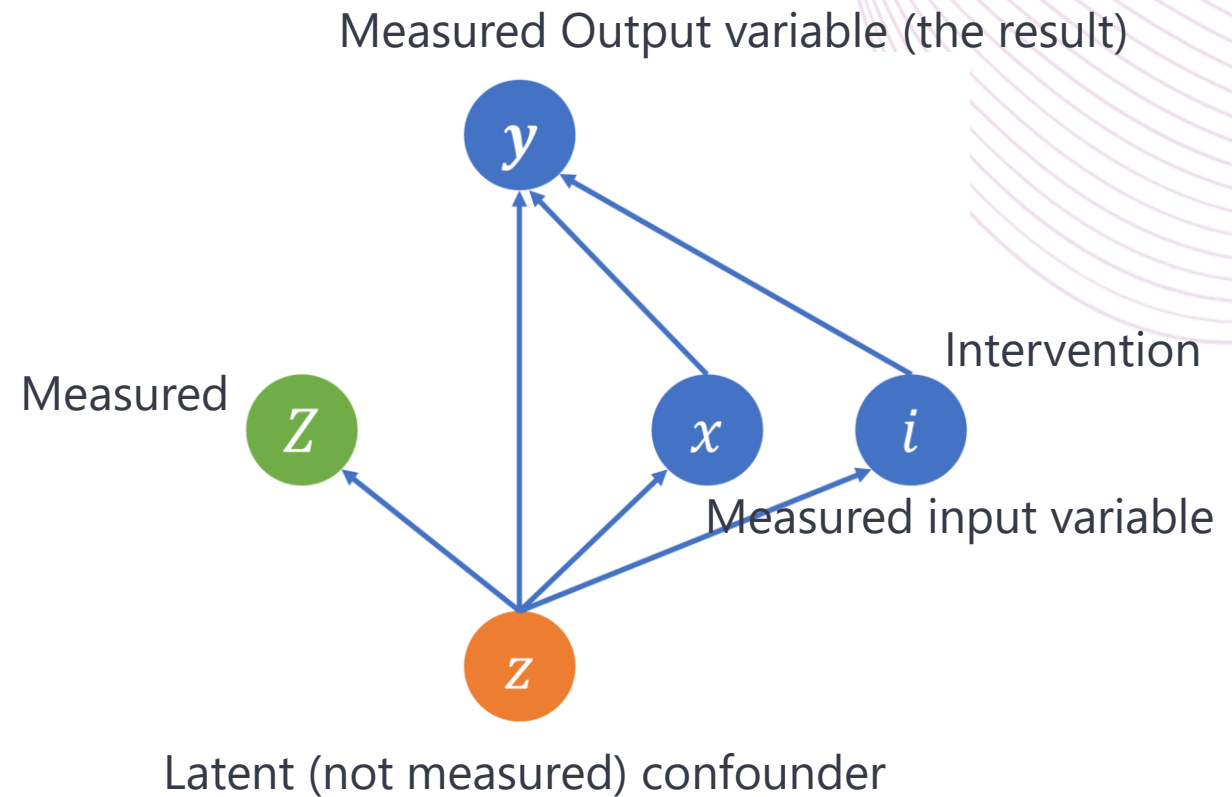


From indicators to PROMs and causal associations



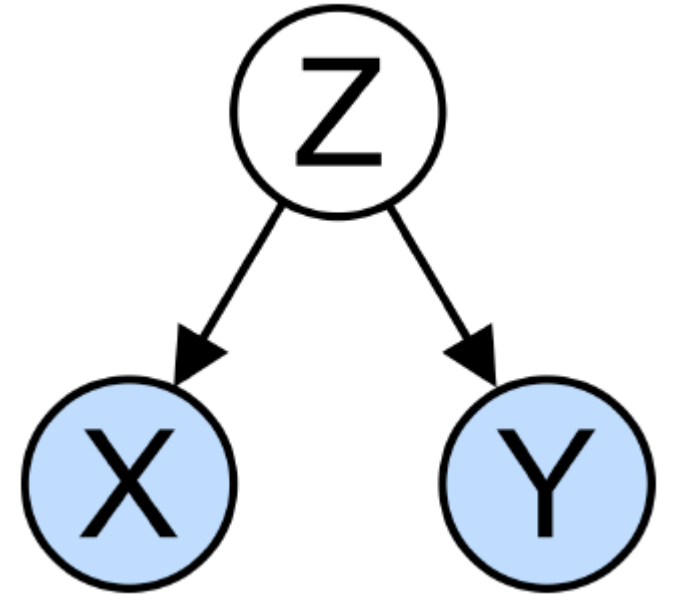
Causal modeling

- Directed Acyclic Graphs
- Constructed by experts or generated from data
- Arrows indicate causalities
- Include dependencies on confounders



The curse of confounders

- In RCTs we would control for Z and we would find no relation between X and Y
- That's not possible using observational data!
- Too many confounders and/or not observed
- Methods to break the curse:
 - Structural Causal Models (SCM)
 - Deep latent confounder modelling



How can we use causal modelling?



- Measure the safety and effectiveness of breast cancer treatments, especially their impact to the quality of life after BC treatment
 - **Measure** causal effects



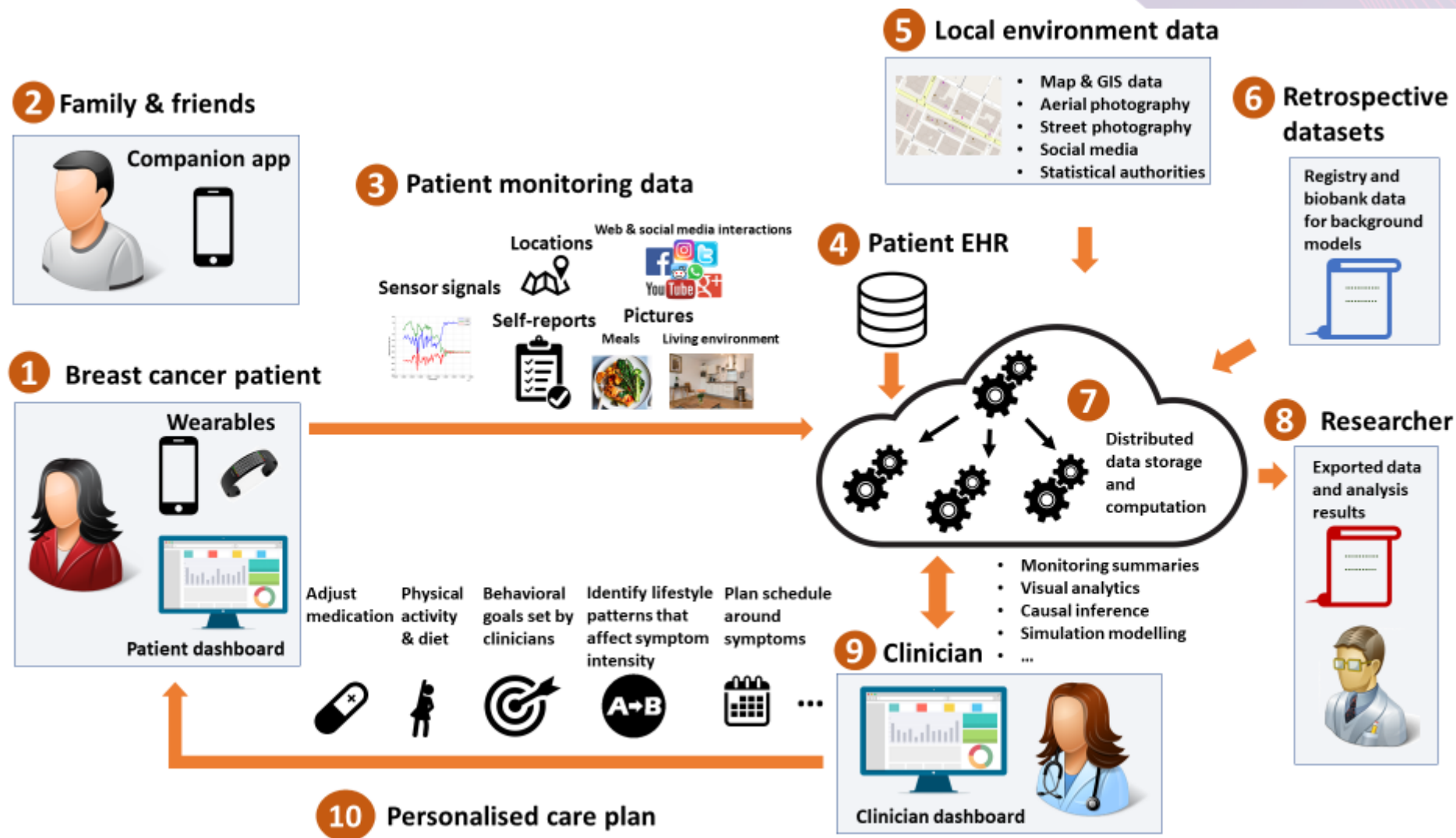
- What can we do to improve clinical outcomes and PROMs?
 - Predict the outcome of **interventions**



- How can we develop **personalized** recommendations and treatments based on the data collected by each patient?
 - Counterfactual reasoning



The REBECCA system



Other Caveats & Challenges

- Data heterogeneity from using of different measurement sources
- Noise in RWD leads to errors in estimation of indicators
- RWD have incomplete or missing values

New for additional countermeasures, for example:

- Estimate error characteristics
- Adopt “error in variables” models
- Estimate and remove bias



Thank you!

- For more information on REBECCA and the applications of Real-World Data

<https://rebeccaproject.eu>

