Context aware evidence tools

Masters project proposals 2024
ETHZ

Project 1 - Ontology Development and Context Gap Analysis

Research Hypothesis

Changes in context (both internal/behavioral and external/environmental) can explain variation in clinical outcomes. If contextual changes are accounted for, clinical trials can reduce variability and improve power.

Description

This project aims to systematically explore how "context" is conceptualized and represented within conceptual models across various Therapeutic Areas (TAs) in clinical research. The goal is to review existing literature, analyze qualitative research, and identify common themes and differences in how context, environment, and adaptation are integrated into these models. The project will result in the development of an initial ontology, a gap analysis of existing datasets and sensors used to capture contextual information, and exploratory analysis of how context can influence outcomes, taking a supervised approach.

Context (i.e. the surroundings and setting in which clinical evidence is measured) is critical in influencing how clinical outcomes are interpreted. However, there is variability in how context is defined and integrated into conceptual models across different TAs and conditions. By systematically reviewing existing literature and datasets, this project aims to standardize the understanding and representation of context, thereby improving the robustness and applicability of conceptual models, and creating a foundation for the measurement of contextual information which can be used in the interpretation and modeling of clinical evidence. Proof-of-concept will be demonstrated through exploratory analysis of the StudentLife and GLOBEM datasets.

Project outline

Phase 1: Ontology Development and Gap Analysis (8 weeks)

1. Ontology Development (based on WHO ICF model)
    - Task 1: Develop an initial ontology for context, integrating the WHO ICF (International Classification of Functioning, Disability, and Health) model .
    - Optional Extension: Investigate disease-specific contextual differences using available qualitative data from various therapeutic areas (TAs).

    - Output: Initial ontology for describing contextual data, especially in relation to symptoms, internal context, and external context.
2. Systematic Literature Review and Gap Analysis
    - Task 2: Perform a structured literature review focusing on current methods and technologies for capturing contextual information in clinical trials across TAs.
    - Task 3: Identify gaps in available datasets and tools compared to the developed ontology.
    - Output: A review paper describing the ontology and comparing it with current measurement capabilities, highlighting gaps. Discussion on the minimal dataset for context across TAs, with a focus on gaps and existing capabilities.

Phase 2: Exploratory Analysis of Contextual Impact on Outcomes (16 weeks)

1. Feature Engineering and Covariate Creation
    - Task 1: Select a few ontology items where data is rich (e.g. one behavioral/internal and one environmental/external item).
    - Task 2: Perform feature engineering to create covariates that summarize contextual changes over time using data from StudentLife or GLOBEM datasets.
2. Impact of Contextual Changes on Outcomes
    - Task 3: Explore the relationship between contextual changes and variations in clinical outcomes, focusing on visual and statistical analysis.
    - Task 4: Compare outcome variability between groups with differing levels of contextual changes.
    - Output: A paper prepared for a formative journal describing exploratory findings and potential benefits of accounting for contextual changes.

Project 2 - Modeling Contextual Data and Its Impact on Clinical Trial Variability

Research Hypothesis

Contextual changes explain part of the variation in clinical outcomes. By modeling contextual data, we can estimate how much trial sizes could be reduced through variance minimization.

Description

This project aims to systematically explore how "context" can be conceptualized and used to model outcomes in clinical research. This project focuses on analyzing the StudentLife and/or GLOBEM datasets to explore how capturing contextual data can reduce variation in key outcome measures, such as the PHQ-9 (a measure of depression). The project aims to model how contextual covariates can explain variation over time, particularly during periods of increasing stress, such as a college semester. The findings will contribute to a viewpoint paper and a detailed analysis of the impact of contextual data on reducing trial sizes in research.

Context (i.e. the surroundings and setting in which clinical evidence is measured) is critical in influencing how outcomes are interpreted. In longitudinal clinical studies, variability in

primary outcomes often complicates data interpretation and increases the required sample size. By incorporating covariates summarizing contextual data, we believe it's possible to account for factors that influence variation in outcomes, potentially reducing the required sample size and improving the accuracy of models. The StudentLife and GLOBEM datasets, which includes a range of contextual data, offers a unique opportunity to investigate this potential.

Project outline

Phase 1: Dataset Analysis and Trial Size Modeling (8 weeks)

1. Data Ingestion and Cleaning
    - Task 1: Ingest and clean the StudentLife and GLOBEM datasets . These datasets include anchor measures (e.g., PHQ-9 in StudentLife) along with contextual data (activity, Bluetooth encounters, GPS, self-report surveys, etc.).
        i. StudentLife Dataset: Analyze the StudentLife dataset (https://zenodo.org/records/3529253), which includes self-report questionnaires (such as PHQ-9), activity data, audio, Bluetooth encounters, conversations, light exposure, GPS coordinates, phone usage (screen on/off, charge status), and Wi-Fi IDs from 48 participants over 66 days.
        ii. GLOBEM Dataset: Analyze the GLOBEM dataset (https://the-globem.github.io/datasets/overview) which includes longer surveys and short daily questionnaires, as well as a wide range of contextual data for several hundred participants over several years (average 78 days of data per participant per year).
    - Output: Clean, well-prepared datasets for further analysis.
2. Modeling Impact of Contextual Data on Trial Sizes
    - Task 2: Develop models to estimate the impact of reducing outcome variability by 0.5%, 1%, 5%, and 10% through the inclusion of contextual covariates (e.g., the PHQ-9).
    - Output: Viewpoint paper discussing the potential of contextual data to minimize trial size requirements and improve power.

Phase 2: Machine Learning for Contextual Change Metrics (16 weeks)

1. Metric Development and Feature Engineering
    - Task 1: Use machine learning techniques to create metrics that describe contextual variation over time (e.g., distance metrics comparing how “normal” an individual's context is based on prior data).
    - Task 2: Alternatively, explore individuals with more or less stable outcomes and compare their contextual stability.
2. Outcome Variation and Contextual Differences
    - Task 3: Analyze differences in outcome variation between individuals with high and low contextual stability.

- Output: A second paper prepared for publication, focusing on how contextual data can be leveraged to improve outcome modeling in trials.