|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-B-020-R2 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | Plenary | New York City, 15-16 November 2018 |
| **DOCUMENT** |
| **Source:** | Mental Health Data Science – Department of Psychiatry Columbia University and the New York State Psychiatric Institute (NYSPI) |
| **Title:** | Developing flexible prediction models for new cases of substance use disorders with a focus on prescription opioid misuse from complex national survey mental health and substance abuse survey data |
| **Purpose:** | Discussion |
| **Contact:** | Melanie M WallMental Health Data ScienceDepartment of Psychiatry Columbia UniversityNew York State Psychiatric InstituteUnited States  | Tel: +1 646-774-5458Cell: +1 612-239-1677E-mail: mmwall@columbia.edu  |

|  |  |
| --- | --- |
| **Abstract:** | This document contains the summary of a possible project on applying AI models to develop more predictive tools for substance use disorders from nationally representative survey data, and proposes the development of assessment approaches for predicting substance use disorders, based on nationally representative survey data.  |

U.S. national longitudinal survey data like the NESARC Wave 1 and 2 have been used to address many epidemiological questions related to risk factors for substance use disorders including cannabis use disorder and prescription opioid misuse. But, there has been only little progress on developing highly predictive models for who will develop substance use disorders. **The goal of this project would be to apply AI models to develop more predictive tools for substance use disorders from nationally representative survey data.** The NESARC 1 was conducted in 2001 with 43,093 participants by experienced lay interviewers and collected information on over 2,500 attributes (questions) for each participant across domains of childhood psychological distress, environmental liabilities, and comorbid psychiatric and medical disorders. All procedures, including informed consent, received full ethical review and approval from the US Census Bureau and the US Office of Management and Budget. The NESARC Wave 2 interview was conducted approximately 3 years later on the same participants. Excluding ineligible respondents (e.g., deceased), the Wave 2 response rate was 86.7% reflecting 34,653 completed interviews. The number of cases of targeted substance use disorders is: cannabis use disorder n=491, and prescription opioid misuse n= 1282. We propose that the FH AI4H develops assessment approaches for predicting substance use disorders, based on nationally representative survey data.

Annex:
Answer to the Questions in DOC B-006

1. **Relevance** - How relevant is the health problem to be addressed?

Drug overdose death killed 72,000 Americans in 2017 higher than the deaths due to HIV, car crashes, or gun deaths. Over 2/3rds of those drug overdoses were due to opioids (either in combination with other drugs or alone)

1. **Impact** - What level of impact will a benchmark in the context of the proposed project have?

If a useful (accurate) prediction algorithm for opioid use disorder, or non-fatal opioid use overdose, or fatal opioid overdoes could be developed then it could be used as a screener for “critical time intervention” to PREVENT future overdose fatalities.

1. **Existing work** - Does the project start from scratch, or are there preliminary experiences?

Many (thousands) of epidemiological studies conducted related to risk factors for substance use disorders, prescription opioid misuse. These studies have examined and identified many domains across the lifecourse that are risk factors for substance use disorders from childhood, adolescence, young adult, and adulthood. For example, genetic predispositions, childhood maltreatment, adverse childhood events, poverty, personality traits, comorbid internalizing psychiatric disorders, social networks, chronic pain.

1. **Feasibility** - Is the project feasible, based on the current state of the art?

We expect that by pooling information across numerous risk factors that useful prediction models can be developed

1. **Data Availability** - Is there sufficient data available? How much of it can be openly available? How much of it as part of the non-disclosed data set?

The NESARC 1 was conducted in 2001 with 43,093 participants by experienced lay interviewers and collected information on over 2,500 attributes (questions) for each participant across domains of childhood psychological distress, environmental liabilities, and comorbid psychiatric and medical disorders. All procedures, including informed consent, received full ethical review and approval from the US Census Bureau and the US Office of Management and Budget. The NESARC Wave 2 interview was conducted approximately 3 years later on the same participants. Excluding ineligible respondents (e.g., deceased), the Wave 2 response rate was 86.7% reflecting 34,653 completed interviews. The number of cases of targeted substance use disorders is: cannabis use disorder n=491, and prescription opioid misuse n= 1282. OTHER possible sources: Electronic health records, Medicaid.

1. **Data Quality** - Is the available data of high quality?

Validated and reliable measures of various domains across the lifespan were asked and self-reported by participants.

1. **Annotation / Label Quality** - Are the annotations / labels of the data of high quality?

The NESARC uses gold standard structured interviews to identify DSM IV substance use disorder diagnosis

1. **Data Provenance** - Has the data been obtained in a professional and ethically correct way?

Data were collected by U.S. Census (NESARC 1) and also Westat (NESARC 2) following highest standards of human protections

1. **Benchmarking** - Do the applicants have a clear proposal about what exactly should be evaluated / measured?

The benchmark for how to weigh the tradeoff between sensitivity and specificity can be determined depending upon where the critical time intervention would take place.

1. **Organizers** - Can the Focus Group work with the applicants, and do they have the time / resources to work with the Focus Group on the problem?

Team members at Mental Health Data Science at NYSPI has percent time available to allocate to development projects

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_