Assessment of current disease burden and unmet needs in MDD

SCAN-2030 Work Package 1 [WP1] Analysis Plan

1. OBJECTIVE

To assess the ten-year current disease burden, unmet care needs and economic burden related to major depressive disorder ("depression" thereafter) in Hong Kong.

2. RATIONALE

Information on disease burden, unmet needs and economic burden will form the foundation for innovative medicine decision-making for both supply and demand. We will demonstrate how real-world data could be used to understand current care needs as potential tools to guide health policy and marketing decisions.

3. DATA SOURCE

We will utilise Clinical Data Analysis and Reporting System (CDARS), a territory-wide electronic medical record (EMR) database managed by the Hospital Authority in Hong Kong. Real-time records in patient demographics, dates of registered death, dates of hospitalization and service attendance, all-cause diagnoses, prescriptions, procedures and laboratory tests across inpatient, outpatient and emergency settings are centralized for audit and research purposes, and de-identified to protect patient confidentiality.

4. STUDY POPULATION

Patients with clinical diagnosis of depression between 1 January 2014 and 31 December 2022 will be identified from the EMR database using International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnostic codes (296.2, 296.3, 300.4 and 311). The patients will be divided into 11 cohorts based on year of diagnosis. The cohort follow-up period will be from the date of cohort identification until death or study end date (31 December 2022).

5. METHODOLOGY and STUDY OUTCOMES

5.1 Prevalence of active patients using Hospital Authority services

"Prevalent patients" mean the current cases living with the disease, no matter if they are newly diagnosed or pre-existing cases in the current time point. "Prevalence", or technically "period

prevalence", refers to the number of current cases divided by the population or sample during a specified period. In this part, we will identify the **annual unique number of patients whoever received a clinical diagnosis of depression** in each year from 2014 to 2022, no matter if they were newly diagnosed in the current year or the years before. We will also calculate the **annual crude and age-standardardized prevalence** by dividing the number of cases by the total mid-year local population in the corresponding year*. For both types of outcome (annual case number and prevalence per 10,000 persons), we will describe and illustrate the trend of overall, age-specific and sex-specific figures from 2014 to 2022. Finally, we will calculate the **ten-year period prevalence** by dividing the total number of unique prevalent patients from 2014 to 2022 (nominator) by the total mid-year population of 2017.
*Source: https://www.censtatd.gov.hk/en/web_table.html?id=110-01002#

Tips for analysis: Combine the "all diagnosis" data, which recorded the all-cause diagnoses from 1993 to 2022 of each cohort. Subset the data to include only records with ICD-9-CM codes of 296.2, 296.3, 300.4, 311. To find the number of patients whoever had a depression diagnosis in, for example 2014, simply subset the records with reference dates fallen into the year of 2014. Remove the duplicated reference keys then count the number of reference keys. Please note that this would be an **underestimated count of true prevalence** because it only accounts for the patients who actively sought consultation in the public healthcare.

5.2 Incidence & newly diagnosed patients

"Incident patients" are the new cases of the disease. "Incidence", or technically incidence proportion, refers to the number of new cases during a specific period divided by population at the start of the interval. In this part, we will identify the **annual number of patients newly diagnosed with depression** in each year from 2014 to 2022. We will also calculate the **annual crude and age-standardized incidence** by dividing the number of new cases by the total local population at the start of the corresponding year. The trend of both outcomes will be described and illustrated at overall, age-specific and sex-specific settings.

Tips for analysis: Within the list of prevalent patients obtained from the section 5.1, check whether they had previous depression diagnosis from 1993, which is the year in which the database was available, to the year before diagnosis. Patients with no history before could be concluded as new cases in that calendar year. ICD-9 code 296.3 (Major Depressive Disorder, Recurrent Episode) applies to patients with multiple episodes of major depression. This means

they cannot be newly diagnosed with this code, as it assumes at least one prior depressive episode. Therefore, patients newly diagnosed with ICD-9 code 296.3 should be excluded from the incident cohort.

5.3 Survival probability analysis based on all-Cause mortality

We will follow up on the incident cohorts, i.e., patients who were newly diagnosed with prostate cancer in each defined year, starting from the earliest confirmatory diagnosis date (index date) until death or the end of the study period (31 December 2022). The analysis will focus on estimating survival probabilities among these patients. To evaluate long-term survival, we will generate Kaplan-Meier (KM) survival curves for the combined cohort (2014-2022) and conduct subgroup analyses stratified by age groups and sex. These survival curves will illustrate the probability of survival over time, helping to identify variations in outcomes across different patient demographics.

5.4 One-year costs of care under Hospital Authority

We will report the annual cost of all-cause care from 2014 to 2022. Based on the 11 prevalent cohorts identified in section 5.1, we will follow up on the patterns of healthcare resource utilization from the index date to death or the annual window cut-off date for each cohort. Taking the 2014 prevalent cohort as an example, the follow-up period of new cases in 2014 will be from the first date of diagnosis to death or 31 December 2014, and the follow-up period of pre-existing cases in 2014 will be from 1 January 2014 to death or 31 December 2014. There will be in total 11 one-year costs which trend can be illustrated.

During the follow-up period of each cohort, we will identify the total number of attendance episodes in the outpatient settings and the total lengths of stay (LOS) in the accident & emergency (A&E) and inpatient settings in a service-type-specific manner. The total episodes or LOS in the 15 service types will be multiplied by the service-specific unit costs (https://www.ha.org.hk/visitor/ha_visitor_index.asp?Content_ID=10045&Lang=ENG) charged as non-eligible persons by the Hospital Authority. There will be 11 aggregated costs which will then be used to plot the graph.

5.4.1 Inpatient by-ward bed-days

For each annual prevalent cohort, inpatient service utilization will be quantified by analyzing all hospitalization records associated with patients with depression, with particular attention to episode duration and ward type. Hospitalization data will be meticulously cleaned and structured to ensure accuracy in length-of-stay (LOS) calculations and cost assignment.

Ward Type Classification:
Inpatient bed-days are categorized into the following mutually exclusive ward types by mapping specialized care-type codes:

- **General wards:** Sum of LOS for "Acute General - Acute" and "Convalescent/Rehabilitation/Infirmary."
- **Psychiatric wards:** Sum of LOS for "Psychiatry/Mentally Handicapped."
- **High Dependency Units (HDU):** Sum of LOS for "Acute General - High Dependency."
- **Intensive Care Units (ICU):** Sum of LOS for "Acute General - Intensive Care."

All LOS variables are converted to numeric, and missing values are set to zero to avoid errors in aggregation. Only records with valid discharge dates are included in the final analysis.

Cohort Attribution and Follow-up:
To ensure that only relevant hospitalizations are included for each cohort, admissions are restricted to those occurring between the patient's cohort entry (first diagnosis or start of the calendar year, whichever is later) and the earliest of death or the end of the calendar year. Any stays with negative LOS (i.e., if death occurred before the episode) are excluded.

Tips for analysis:

Data Cleaning and Processing:

- Each admission and discharge date is standardized to ensure consistency across various date formats.
- Admissions with discharge dates earlier than first diagnosis are excluded.
- For hospital episodes that span multiple calendar years, the record is split so that the LOS is apportioned to each year according to the actual number of days spent in each calendar year.

- To avoid double-counting, episodes that are contained within other, longer episodes (in terms of admission and discharge dates) are removed.
- For partially overlapping records (where a new admission occurs before the previous discharge), LOS is proportionally adjusted to ensure that overlapping days are not counted twice.

5.4.2 Outpatient service-specific episodes

Outpatient service episodes for each prevalent cohort will be identified and analyzed to quantify the disease burden and associated costs of outpatient care for depression under the Hospital Authority. Outpatient service classification will utilize detailed service-type codes and specialty designations available in the electronic medical record.

For each analysis year, all outpatient attendance records will be extracted for patients in the prevalent cohort, ensuring that only those episodes occurring within the individual's defined follow-up window (from cohort entry to death or annual cut-off) are included. Records are filtered to remove patients not in the cohort and to restrict to the appropriate observation period for each prevalent case.

Each outpatient record will then mapped to a specific service category in HA charging list (https://www.ha.org.hk/visitor/ha_visitor_index.asp?Parent_ID=10044&Content_ID=10045 &Ver=HTML) using a standardized outpatient service cost library. The outpatients service cost library (protocol file costs_mapping_final.docx) was developed according to service group, service type code, specialty, and sub-specialty fields of the outpatients' services. Independent cross-checks were conducted during the generation process to ensure the objectivity and accuracy of the document.

For each unique outpatient episode with a valid service-type mapping, the corresponding unit cost (as charged to non-eligible persons by Hospital Authority) is assigned. The total number of outpatient episodes and the aggregate cost are then summarized by service category for each cohort year.

Finally, the total cost of outpatient services is calculated both by category and in aggregate, excluding Accident & Emergency (A&E) and inpatient-related categories to ensure that the

outpatient-specific burden is accurately measured. Results are summarized and exported to facilitate year-to-year comparison and to enable illustration of trends in outpatient burden and resource utilization for depression care across the study period.

5.5 Unmet Needs for innovative medicines

Using the prescription data, we will visualize the number of patients in each incident cohort (2014-2022) who developed treatment-resistant depression during the follow-up until 2022. Treatment-resistant depression is defined as at least two trials/switches of antidepressant regimens with adequate duration and dosage, with the presence of the third regimen to confirm the failure of the previous two lines of treatment. To be more specific, one regimen is defined as the same choice of antidepressant or combination regimen of at least 28 days with gaps no longer than 14 days. The date of turning into a treatment-resistant patient is defined as the date of commencing the third regimen. Patients who do not meet the criteria for TRD throughout the entire follow-up period (until death or end of study period) will be classified as non-TRD. We will plot the cumulative proportion of patients who develop TRD by year in all cohorts.