



QUANTIFYING RISK, ENABLING OPPORTUNITY

Health Market Inquiry: Data and Methodological Considerations

Adam Lowe
Willis Towers Watson
4 May 2017

Introduction

- WTW Role
 - Warehousing of data and some analysis work
 - Specific focus on medical scheme data
 - Analysis work focused on medical expenditure by or on behalf of medical scheme beneficiaries
- HMI Context
 - Presenting work WTW has done for the HMI
 - 3 Specific Conditions:
 - Presentation limited to data and methodologies
 - No discussion or presentation of results
 - No identification of individual stakeholders

Report Structures

- WTW has prepared six reports to the HMI:
 - Data and Descriptive Statistics
 - Overall Expenditure/Claims Trends
 - PMB and PMBD Analyses
 - Facility (Hospital) Analyses
 - Practitioner Analyses
 - Funder Analyses
- HMI takes WTW reports and uses them as an input into their reports

Structure and Contents

Raw Data Used

Data Completeness

Other Data Considerations

Analysis Datasets and Key Variables

Descriptive Methodologies

Statistical Methodologies

Other Methodological Considerations

Raw Data Used

- Medical Scheme Claims and Membership Data for the period from 2010 to 2014
- Schemes were asked to supply four main files:
 - Policy (Member) file;
 - Beneficiary (Life) file;
 - Claims file; and
 - Provider file.
- A number of tables were also requested separately to provide additional information and allow for appropriate protection of personal information

Data Completeness

- Schemes excluded if:
 - Data not supplied;
 - Incomplete data supplied; or
 - Quality checks failed.
- Quality checks performed both within datasets (e.g. do claims map accurately to members) and against the CMS annual reports
- Final dataset covers 94% of 2014 industry, with lower coverage in earlier years (due to incomplete submissions and deregistered schemes)

Other Data Considerations

- Schemes enter and (occasionally) exit the dataset during the period:
 - Membership trends should be analysed using the CMS data which is complete
 - Claims trends should be exposure adjusted
 - Trend analyses which are not risk adjusted should be interpreted in light of this
- Scheme data submissions are variable in format and quality:
 - Tariff, NAPPI and ICD10 codes not in standard formats
 - Modifiers captured inconsistently, if at all
 - PMB flags reflect multiple treatments of the PMBs
- Analyses must be based on what is universally available, meaning a 'lowest common denominator' approach is required

Analysis Datasets – Beneficiary Dataset

- One line per active beneficiary per year
- Contains:
 - Demographic information (age, gender, scheme, benefit option etc.)
 - Clinical profile indicator
 - Member movement indicator
 - Utilisation indicators:
 - Medical practitioner usage
 - Multiple medical practitioners indicator
 - Pathology and radiology usage
 - Summarised claims information
 - OH claims by benefit category
 - Admissions
 - IH claims (total)
- Used to analyse overall utilisation and claims patterns

Key Variables – Clinical Profile

- Intended to provide a proxy for disease burden
- ICD10 codes from consultations and NAPPI codes grouped into 18 disease groups
- A beneficiary must have two consults or one consult and two scripts to be placed in a group
- Priority order (most severe to least severe) applied to generate a single group for each beneficiary in each year

Key Variables – PMB Claims

- A key issue within the HMI is the impact of the PMB regulations on expenditure
- Two different PMB definitions are included in the analysis data:
 - PMB flags as provided by the administrators (PMBs); and
 - PMB diagnoses as listed by the Council for Medical Schemes (PMBDs).
- Most analysis in subsequent reports is based on PMBDs for multiple reasons

Key Variables – Plan Mix Groups

- Schemes have submitted that ‘buy-downs’ cause a material understatement of annual claims increases
- Variable created to test these submissions
- Options grouped by their benefit design, both OH and IH
- Groupings made using information in the public domain, so some options are ‘Unknown’

Analysis Datasets – Admissions Dataset

- One line per admission
- Contains:
 - Admission and discharge dates
 - Demographic and clinical profile variables (as per beneficiary file)
 - Treating practitioners (doctor, anaesthetist, pathologist, radiologist)
 - Diagnosis and procedure groups
 - Admission type (surgical, medical, maternity)
 - Days spent in ward types (general, high care, ICU etc.)
 - Claims information
 - By discipline
 - Hospital fees split into ward, theatre, NAPPI, ARM and Other
- Used to analyse intra-admission trends (LoS, LoC, CPA)

Key Variables – ‘Case Mix’

- The model uses individual factors as opposed to the more common DRG approach
- Age, gender and disease profile are supplemented by an ‘Admission Type’ indicator
- This indicator is built using a combination of treating specialist discipline, diagnosis and procedure
- Groups are created within each specialist discipline such that the majority of cases fall into the specific group – rest are grouped together within each discipline
- Admissions are grouped into one of 177 groups by this procedure

Descriptive Methodologies

- Claims trends presented as per beneficiary per year figures
- Claim increases split into three parts:
 - Utilisation (% claiming lives)
 - Intensity (# of services per claiming life)
 - Price (CPI)
- Admission trends presented per 1 000 lives, usually split between day and overnight stay

Statistical Methodologies – Overall Process

- GLM modelling used, with varying parameters, response distributions and link functions depending on the response variable
- Models parameterised on 2014 data and parameters applied to all five years
- Stepwise addition of variables to allow for the impact of each individual variable to be assessed
- Annual impact calculated as change in the average model prediction per beneficiary/admission from year to year

Statistical Methodologies – Cost Models

- Base dataset is the beneficiary dataset i.e. units of analysis are beneficiary years
- Two step modelling process:
 - First probability of claiming modelled using a binomial GLM with logistic link function
 - Then cost modelled (conditional on claiming) using a GLM with gamma distribution and log link function
- Final prediction calculated by multiplying the predicted probability of claiming by the predicted claim conditional on claiming
- ‘Risk Index’ (essentially average predicted cost) then created to allow for the necessary calculations

Statistical Methodologies – Other Models

- Admission rate (per 1 000 lives) model run using a Poisson GLM with log link function, and the same independent variables as the cost per beneficiary models
- Models relating to admissions use the admission dataset and a different set of independent variables:
 - Cost per Admission (CPA) model uses a gamma distribution with log link function
 - Length of Stay model uses a Poisson distribution with log link function
 - Level of Care model uses a normal distribution with an identity link function (essentially multiple linear regression)
- ‘Risk Index’ (essentially average predicted cost) then created as for the cost per beneficiary models

Questions?

