Who's got what? A closer look at disease prevalence

By SL Bennett

Presented at the Actuarial Society of South Africa's 2024 Convention
Cape Town International Convention Centre, 20–21 November 2024

ABSTRACT
Disease prevalence is the proportion of people in a population who have a particular disease. Disease prevalence is a fundamental building block for many measures in an outcomes-based framework. Disease prevalence provides visibility on the burden of disease in a population and how it moves over time. Finally, disease prevalence can be used to risk adjust different populations relative to each other. This paper highlights the challenges of estimating the prevalence of chronic conditions in the context of medical schemes in South Africa.

KEYWORDS
Disease prevalence, case definition, disease burden, sensitivity, specificity

CONTACT DETAILS
Mrs Sarah Bennett, Medscheme, Cape Town; Email: sarahb@medscheme.co.za

1. INTRODUCTION

1.1 Disease prevalence is the proportion of people in a population who have a particular disease or characteristic and is a fundamental building block for many measures in an outcomes-based reporting and management framework.

1.2 In order to estimate the proportion of people in a population with a chronic condition (prevalence), a decision must be made on who to count as having the condition i.e., there has to be a case definition. A case definition is a set of standard criteria for classifying whether a person has a particular disease or characteristic.[1]

1.3 At its essence, disease prevalence is a simple concept, in that we are counting individuals with a particular disease or condition for a defined time period or at a given point in time.

1.4 Calculating disease prevalence can become complex as different diseases have different characteristics. Most of the challenges arise from determining case definitions that minimise false positive and false negative cases. In addition, the definition of prevalence must be appropriate for the purpose. For example, if we want a high degree of certainty that the person has a particular condition, we would impose additional criteria. For example, the Scheme Risk Management (SRM) or Risk Equalisation Fund (REF) criteria under the Health Utilisation Annual Statutory Returns (HUASR) requirements from the Council for Medical Schemes (CMS) are stricter in that they apply diagnostic criteria as well as requiring proof of treatment.

1.5 At Medscheme,[2] chronic medicine registrations and claims data are used to estimate prevalence of chronic conditions. Two key challenges of using chronic registrations and claims data for estimating prevalence are:

— defining what to count as a case (defining case definitions); and
— defining the period. We have sight of claims only from when beneficiaries join a scheme so we can't easily define a period for prevalence reporting (we can only look back as far as the beneficiary has been on a scheme, which may be months for some and years for others). Further, beneficiaries may be registered for chronic medication but not claiming chronic medication.

1.6 The purpose of this paper is to provide clarity on how disease prevalence is defined, to consider the complexities of defining disease prevalence and case definitions for certain conditions, including diabetes, hypertension, HIV, cancer and depression, and to identify and unpack the various disease prevalence definitions used currently. Further, we consider defining disease prevalence with a view to calculating the disease burden and/or applying an appropriate risk adjustment. This is very relevant in the light of the National Health Insurance roll-out and the Contracting Units for Primary Health Care (CUPs).

1 https://www.britannica.com/science/case-definition
2 https://www.medscheme.com/about-medscheme/

Section 37(2)(e) of the NHI Act[3] states that the CUPs must assist the NHI Fund to access information on the disease profile in a particular sub-district that would inform the design of the health care service benefits for that sub-district.

2. DEFINING DISEASE PREVALENCE

2.1 Definition of disease prevalence

2.1.1 Disease prevalence is the proportion of people in a population who have a particular disease or characteristic at a specified point in time (point prevalence) or over a specified period of time (period prevalence).

— True prevalence refers to all individuals in a population who have the disease.

— Apparent prevalence is the term we use at Medscheme to refer to individuals who appear to have (or can be reasonably inferred to have) a disease from their registration, claims data or pathology results.

— Registered prevalence refers to individuals who have a chronic disease or medicine registration

— Chronic medicine registered prevalence refers to individuals who have a chronic medicine registration.

Whether the person is claiming medication or not is an additional dimension that can be overlaid onto the disease prevalence definitions as shown in Figure 1.

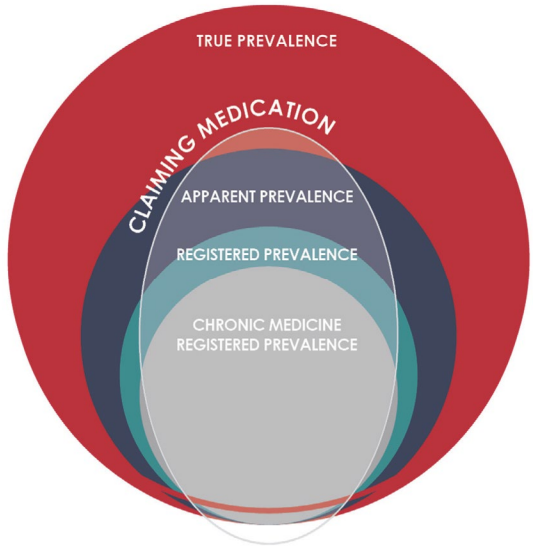


Figure 1. Venn diagram showing prevalence types vs those claiming medication

3 https://www.gov.za/sites/default/files/gcis_document/202405/50664nathealthinsuranceact202023.pdf

2.1.2 The approach to determining prevalence

According to Duncan et al. (2007) in a Practice Note on Disease Management prepared by the American Academy of Actuaries, a chronic population, whether consisting of a single disease or multiple chronic diseases, will exhibit changes in disease prevalence over time. This disease prevalence may arise from different sources, which the actuary may wish to consider carefully:

— Changes in underlying clinical disease prevalence;
— Changes in identification methodologies (and related claims issues);
— Changes in the composition of the population (e.g., the addition of a large new employer group);
— A change in the severity of the population (for example, increasing co-morbidities); and
— Statistical false-positives (members who met the chronic identification criteria in a prior period but who no longer do so in the current period).

Depending on the source of change in prevalence, different methods may be required to ensure equivalence between populations over time.

Different identification rules may be applied to address this problem. With any definition, a balance typically is struck between sensitivity (finding all the members who may have the condition) and specificity (excluding those members who are "false positives" and who may be identified through rule-out services).

According to St. Clair et al. (2017) in the paper titled "Using self-reports or claims to assess disease prevalence: It's complicated", comparisons of diagnosed diabetes and heart attack prevalence show similar trends by source, but claims differ from self-reports with regard to levels. Selection into insurance plans, disease definitions, and the reference period used by algorithms are identified as sources contributing to differences.

Case definitions should minimise false negative and false positive cases as far as possible. We often use what we term "registered" prevalence i.e. counting people as having the condition if they have a chronic medicine registration for the disease. This has advantages and disadvantages. For example, it means counting people who are on treatment for the disease but missing people who have the disease but are not on medication for the condition. People also do not tend to deregister from chronic medicine registrations, so they may continue to be counted as having the condition even if they were originally incorrectly registered for the disease, if they no longer have the disease, e.g. they outgrow a childhood asthma diagnosis, or if they are in remission, e.g. they had an episode of depression in the past but do not currently suffer from depression.

We can also look at what we refer to as "apparent prevalence" which involves determining a case definition that is broader than simply looking at chronic medicine registrations. It could include, for example, pathology tests/results and relevant ICD-10 codes linked to claims. Once again, there are advantages and disadvantages with this approach and the

case definition needs to attempt to minimise false positive and false negative cases. The case definition that is selected may depend on whether it is more important to identify anyone who may have the condition or whether it is more important to be confident that people who are counted as having the condition actually do have it.

2.1.3 Using a probabilistic algorithm to determine prevalence

One possible method of measuring prevalence is to consider the relevant features in the data that could indicate prevalence and use a regression model or machine learning techniques to assess likelihood of prevalence for a particular beneficiary.

According to Tonelli et al. (2015), validated algorithms were identified that use ICD-9 CM/ICD-10 data to ascertain the presence or absence of 40 morbidities. Algorithms with both positive predictive value and sensitivity ≥70% were graded as "high validity"; those with positive predictive value ≥70% and sensitivity <70% were graded as "moderate validity".

According to Souri et al. (2017), case definitions are automated computerised algorithms applied to secondary data that allow for identification of specific cohorts within EMR databases without the need for manual chart review by a researcher or clinician. Case definitions are validated against a gold standard for disease identification, most often manual review of patient charts.

According to Duncan (2005), there is no unique way of identifying who has a chronic disease. In order to be useful for measurement, however, an identification algorithm needs to be objective, stable over time, and cheap to apply. Many health plans use risk-adjuster methods to classify and rank members by risk class (Duncan, 2005). Chronically ill members are often identified from claims data, so data are a central issue. Unfortunately, there is no ideal source of data. Each source has its advantages and drawbacks, which must be weighed against each other. Five types of data commonly available to the health care analyst are medical charts, survey data, medical claims, pharmacy claims and laboratory values.

When using a probabilistic model for chronic identification, the problem of false negative and false positive identification will arise (Duncan, 2005).

According to Duncan (2005), we should also note that a set of criteria appropriate for identifying members for one purpose may not be the most appropriate for another. For example, one use of identification criteria may be to find members for a management programme, and another is to identify members for measurement. In the first instance, specificity is not as important as sensitivity (we need to identify as many members as possible with the condition to implement a successful programme). For measurement or other examples involving financial objectives, such as reimbursement of providers, we need to be reasonably certain that the identified population actually have the condition.

There is a place for probabilistic models for chronic identification, particularly at a population level. However, for interventional purposes where individuals need to be identified and managed, a deterministic approach may be more practical.

Table 1. The problem of false negatives and false positives

The problem of false negatives	False negatives are chronic members who are "missed" by an identification algorithm. These members are more of a problem for programme management than for programme measurement. To the extent that a beneficiary has a condition that is untreated, claims data will be unavailable and the member will be unidentified. A more difficult false negative problem occurs when the member's provider is not part of the data-submission system, for example, in the South African context, when HIV-positive beneficiaries claim their antiretroviral therapy (ART) from State.
The problem of false positives The problem of false positives (continued)	False positives are members who are falsely identified as having a chronic condition, when they do not have that condition. There are two types of false positives: – Clinical false positives, as the name implies, are those members who are identified with the condition and later found not to have it. – Statistical false positives, on the other hand, arise because the administrative claims used for identification will never be complete, unambiguous or correctly coded. When identification of chronic conditions takes place from administrative claims data, there is a chance of statistical false positives (which may be different than clinical false positive identification). We define statistical false positives as those members who meet a chronic definition in Year 1, but who do not requalify according to the same set of definitional criteria in Year 2. This issue is important for disease management outcomes evaluation because false positives, who do not have the condition according to the claims data, are less likely to have high costs. Therefore their continued inclusion in the chronic population, although they no longer meet chronic definition criteria, will likely reduce the average cost (and therefore the trend) in the chronic population, resulting in apparent reduction in cost due to the programme.

2.1.4 In biostatistics, according to Habibzadeh et al. (2022), so-called "apparent prevalence" (pr), is not necessarily an unbiased estimation of the true prevalence (π), the true proportion of diseased people in the population or the study sample. We can derive an unbiased estimation of π from the obtained pr and the test sensitivity (Se) and specificity (Sp). Sensitivity is a measure of how well a given test identifies the disease or trait in question (i.e., how well it avoids false negatives), while specificity is a measure of how well a given test identifies the absence of the condition being tested (i.e., how well it avoids false positives).[4]

According to Habibzadeh et al. (2022), the pr (the apparent prevalence) is defined as the portion of tested people with a positive test (T+). Therefore:

$$pr = P(T+) = TPR + FPR$$

where TPR and FPR are true-positive and false-positive rates, respectively.

Substituting the TPR and FPR, we have:

$$pr = TPR + FPR$$

4 https://www.britannica.com/science/sensitivity-medical-statistics

$$= \pi \, Se + (1 - \pi)(1 - Sp)$$

Solving the above equation for π (the true prevalence), yields:

$$\pi = (pr + Sp - 1) / (Se + Sp - 1)$$

Table 2. Results of the hypothetical test validity study (Habibzadeh et al., 2022)

		Disease		Total
		Present	Absent	
Test	Positive	70 (TP)	15 (FP)	85
	Negative	5 (FN)	135 (TN)	140
Total		75	150	225

TP – True positive. FP – False positive. FN – False negative. TN – True negative. N = TP + FP +FN + TN = 225. Se = TP/(TP+FN) = 0.93. Sp = TN/(TN+FP) = 0.90. TPR = TP/N = 0.31. FPR = FP/N = 0.07. FNR = FN/N = 0.02. Apparent prevalence = TPR + FPR = 0.31 + 0.07 = 0.38. True prevalence = TPR +FNR = 0.31 + 0.02 = 0.33. Using the equation above it can be calculated: True prevalence = (Apparent prevalence + Sp –1)/ (Se + Sp –1) = (0.38 + 0.90 –1)/(0.93 + 0.90–1) = 0.33.

According to Habibzadeh et al. (2022), depending on the Se and Sp of the diagnostic test used in a given prevalence study, the results obtained are generally biased estimates of the true prevalence of the condition of interest (e.g., a disease). The derived apparent prevalence values should therefore be corrected. Based on the variances of the seroprevalence, and the test Se and Sp, it is possible to calculate an unbiased estimation of the true prevalence.

2.1.5 Care should be taken when defining prevalence as it may impact on other measures, such as reported medicine adherence and hospital readmissions. For example, hospital readmissions can be measured for those deemed prevalent based on the various definitions:

— Chronic medicine registrations only
— Registered prevalence (chronic medicine or other chronic disease registrations), and
— Apparent prevalence.

It is not possible to measure readmissions based on true prevalence as true prevalence is unknown.

2.1.6 Clinically, disease prevalence is confirmed through diagnosis. According to Jutel (2009), diagnosis has been described as both a process and a classification scheme, or a "pre-existing set of categories agreed on by the medical profession to designate a specific condition". Clinical diagnosis is complex and beyond the scope of this paper. Medical schemes have to rely on the diagnosis made by the doctor and captured in the

ICD-10 code on the authorisation and/or claim, or encapsulated on the chronic medication authorisation application or the Ambulatory Prescribed Minimum Benefit (aPMB) Care Plan application for a particular disease; aPMBs are a basket of out-of-hospital benefits that must be covered as Prescribed Minimum Benefits (PMBs). Accurate diagnoses enable accurate case definitions, which in turn enable the accurate calculation of disease prevalence.

2.1.7 In the context of managed care,[5] consideration must be given to the chronic disease registration process at each Managed Care Organisation (MCO). The chronic disease registration process will directly impact the chronic registrations definition of prevalence as well as the registered prevalence definition.

At Medscheme, for example, chronic application requests (initially in the form of a script with the relevant ICD-10 codes) are received directly from doctors on behalf of their patients and will include information to confirm the diagnosis of the disease.

Certain requests will be automatically authorised by the systems, for example, first line treatment requests for common chronic conditions. Other requests will be automatically declined by the system; for example, medication requested that is considered to be exclusions, e.g. hypnotics and sedatives for chronic use.

Some requests will route for clinical intervention, for example, disease authorisation requests that need to meet reimbursement criteria set out in the clinical guidelines, e.g. hyperlipidaemia and medication that is not first line treatment.

There are administrative processes to deregister beneficiaries who have not claimed medication for a certain number of months.

Beneficiaries who regularly claim medication from acute medication benefits may be deemed to be claiming chronic medication and registered via the auto-chronic process. This only applies to certain diseases and specific medications with a high degree of certainty that the condition is of a chronic nature. Three months of medicine claims are required and there will also need to be a record of a doctor's visit confirming the chronic diagnosis. Only specific medication for conditions on the Chronic Disease List (CDL)– a list of chronic diseases that must be covered by all schemes as PMBs – is included in the auto-chronic process.

Care plans or baskets define standard outpatient health care services that are likely to be needed by patients with certain diagnosis and include, for example, doctor consultations, radiology and pathology tests. These aPMBs care plans can be generated/triggered by:

— Chronic medicine authorisations,
— Claim submissions for the particular disease,

5 Managed care, within the South African context, is a term used to refer to a diverse range of healthcare organisational strategies aimed at controlling cost, improving access and assuring higher levels of quality of care provided to those covered by medical schemes. https://www.medicalschemes.com/files/Guidelines%20and%20Manuals/Standard%20doc%20managed%20care%20final%2022%20Oct%2003.pdf

— Manual activation,
— Disease management questionnaires, and
— Hospital authorisations.

2.2 Definition of diabetes prevalence

2.2.1 According to Chawla et al., (2016) diabetes mellitus (diabetes) is a metabolic disorder characterised by hyperglycaemia that develops as a consequence of defects in insulin secretion, insulin action, or both.

According to the SEMDSA 2017 Guidelines for the Management of Type 2 diabetes mellitus, the long-term effects of diabetes include microvascular complications such as retinopathy, nephropathy and neuropathy; and macrovascular complications including cardiovascular diseases such as heart attack, stroke and peripheral arterial disease.[6]

The International Diabetes Federation (IDF) reports that the prevalence of diabetes among South African adults is 11.3%. Over 4.2 million adults in South Africa are estimated to be living with diabetes.[7]

2.2.2 Case definition for diabetes

There are challenges in determining the case definition for diabetes using the various sources of information available in medical scheme data.

— Chronic medicine registrations may be incorrect and beneficiaries may not be claiming medication.
— Care plan registrations may be incorrect and beneficiaries may not be claiming from the aPMB care plan or claims may not be allocated to the aPMB care plan.
— HbA1c results (and other pathology tests that could indicate diabetes) may not be available.
— Admissions and claims may be incorrectly coded.

Table 3. Case definition in use for some of our medical schemes

Definition of prevalence	Point prevalence (measured as at a particular date)
Chronic medicine registered prevalence	Registered for chronic medication for diabetes
Registered prevalence	Registered for chronic medication for diabetes AND/OR registered on diabetes care plan
Apparent prevalence	Registered for chronic medication for diabetes AND/OR registered on diabetes care plan AND/OR HbA1c > 6.5% AND/OR diabetes codes on three or more claims in three separate months AND/OR diabetic specific admission AND/OR any admission where diabetes codes are in the admission string, measured over a rolling 12 month period
True prevalence	Not measured or estimated

6 https://www.emcmobi.co.za/wp-content/uploads/2017/08/JEMDSA-2017.pdf
7 https://idf.org/our-network/regions-and-members/africa/members/south-africa/

Medicines are not generally included in the apparent prevalence definitions; however, it should be noted that these beneficiaries may be included as chronic medicine registered as a consequence of the auto-chronic process. Some medications can be used for more than one indication (for example metformin is licensed for use in type 2 diabetes but can also be used (off-label) to manage polycystic ovary syndrome). Antihypertensive drugs may also have more than one indication. The auto-chronic process would only automatically register someone for a condition based on their medication if there is a high degree of certainty that the patient has the condition.

2.2.3 We need to consider the fact that diabetes is reversible. There are a number of studies examining the impact of diet and lifestyle on chronic conditions. According to Noakes et al. (2019), high carbohydrate/sugar diets in susceptible people cause fatty livers, pancreatic damage, insulin resistance and ultimately diabetes. The changes can be picked up before the onset of diabetes by monitoring for components of the metabolic syndrome. It can be prevented or reversed by following low carbohydrate lifestyles. The terms "reversal" and "remission" have both been used to describe regression of hyperglycaemia and insulin resistance, but they are defined slightly differently (Noakes et al., 2019):

— **Reversal:** Sub diabetes level glycaemia (HbA1c <6.5%) that is achieved without diabetes medications or only with metformin.
— **Remission**: Similar to reversal but requires that no diabetes medication (including metformin) be used and that it is maintained for at least one year.

According to Riddle et al. (2021), remission should be defined as a return of HbA1c to <6.5%.

According to Hallberg et al. (2018), Virta Health focuses on diabetes reversal for the following reasons:

— It is not in every individual patient's best interest to stop metformin.
— Patients who have a history of diabetes remain at increased risk of developing diabetes again, even after HbA1c normalises, and metformin has been shown to decrease this risk.

Where diabetes is in remission, the beneficiary may deregister for medication (or be automatically deregistered if they do not claim for a certain period) and so will not be counted under Chronic Medicine Registration Prevalence. The beneficiary will most likely require regular monitoring and will remain registered on the care plan and so will be counted under Registered Prevalence.

2.2.4 The CMS report entitled "Quality of Care in Medical Schemes" (2024)[8] includes the enrolment into disease management programmes in 2022. Diabetes Type 2

8 https://www.medicalschemes.co.za/download/3738/research-note-2024/28789/quality-of-care-in-medical-schemes

has the second highest number of registered beneficiaries at 541 297 (or 6.0% based on 9 039 259 beneficiaries reported in the CMS Annexures for 2022[9]). Diabetes Type 2 registrations grew by 64.6% between 2015 and 2022.

2.2.5 According to Pheiffer et al. (2022), the prevalence of Type 2 diabetes in South Africa was estimated as 15.25% (11.07–19.95%) in individuals 25 years and older. This systematic review and meta-analysis pooled prevalence data from 11 population-based studies dating from 2001 to 2019.

2.3 Definition of hypertension prevalence

2.3.1 Hypertension, also known as high or raised blood pressure, is a condition in which the blood vessels have persistently raised pressure.[10]

According to Oparil et al. (2018) and Williams et al. (2018), hypertension is the most important modifiable risk factor for all-cause morbidity and mortality worldwide. It is associated with increased risk of cardiovascular disease (including coronary heart disease, heart failure, stroke, myocardial infarction, atrial fibrillation and peripheral artery disease), chronic kidney disease and cognitive impairment.

Globally, over one billion people have hypertension. The World Health Organization estimates that approximately 27% of adults in South Africa have hypertension,[11] although prevalence of up to 58% has been reported (Gaziano et al., 2017; Jongen et al., 2019). According to Folb et al. (2016), a study in the Western Cape found that 60% of a cohort of hypertensive primary care clinic attenders had uncontrolled blood pressure.

According to Seedat et al. (2014), hypertension is defined as office systolic blood pressure (SBP) values ≥ 140 mmHg and/or diastolic blood pressure (DBP) values ≥ 90 mmHg. Williams et al. (2018) recommend to base the diagnosis of hypertension on repeated measurements on more than one visit, except when hypertension is severe (e.g. grade 3).

2.3.2 The challenges with the case definition of hypertension (in addition to the challenges mentioned for diabetes is section 2.2.2) are the reliability of blood pressure readings and the definition of hypertension. Gupta et al. (2021) show that the prevalence of hypertension in South African adults would change significantly if the new ACC/AHA 2017 definition was adopted, compared to the existing JNC7 guideline. For medical schemes, registration for hypertension will typically follow the process described in Section 2.1.7 above and will be based on a script from the doctor confirming the diagnosis.

2.3.3 Based on medical scheme data, hypertension has the highest prevalence of all chronic conditions. Figure 2 shows the prevalence for all schemes in the industry relative to other diseases and, for illustration, according to the various definitions. Hypertension can be managed through diet and lifestyle and hence registered prevalence is higher than

9 https://www.medicalschemes.co.za/download/3697/industry-report-2022-23/28137/annexures-to-the-2022_23-industry-report.xlsx
10 https://www.who.int/health-topics/hypertension#tab=tab_1
11 https://apps.who.int/gho/data/node.main.A875STANDARD?lang=en

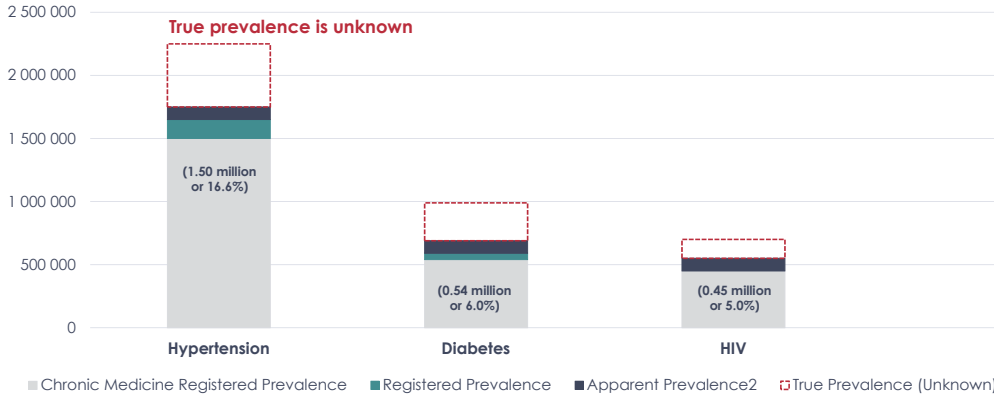


FIGURE 2. Disease prevalence across all medical schemes (extrapolated to include hypothetical registered and apparent prevalence)[12, 13]

chronic medicine registered prevalence. Registered prevalence includes beneficiaries who are on an aPMB care plan only. Apparent prevalence is based on ICD-10 codes in claims that indicate a diagnosis of hypertension.

2.3.4 The CMS report entitled "Quality of Care in Medical Schemes" (2024)[14] includes the enrolment into disease management programmes in 2022. Hypertension has the highest number of registered beneficiaries at 1 499 365 (or 16.6% based on 9 039 259 beneficiaries reported in the CMS Annexures for 2022[15]). Hypertension registrations grew by 52.3% between 2015 and 2022.

2.4 Definition of depression prevalence

2.4.1 According to Colman et al. (2011), depression is a common and often recurrent mood disorder that compromises daily functioning and is associated with a decrease in quality of life. People with depression experience a range of symptoms including persistent depressed mood, and diminished interest in or pleasure from activities.[16] At its worst, depression can lead to suicide. Depression results from a complex interaction of social, psychological and biological factors.[17]

12 https://www.medicalschemes.co.za/download/3738/research-note-2024/28789/quality-of-care-in-medical-schemes
13 https://www.medicalschemes.co.za/download/3697/industry-report-2022-23/28137/annexures-to-the-2022_23-industry-report.xlsx
14 https://www.medicalschemes.co.za/download/3738/research-note-2024/28789/quality-of-care-in-medical-schemes
15 https://www.medicalschemes.co.za/download/3697/industry-report-2022-23/28137/annexures-to-the-2022_23-industry-report.xlsx
16 https://www.who.int/teams/mental-health-and-substance-use/treatment-care/mental-health-gap-action-programme
17 https://www.who.int/news-room/fact-sheets/detail/depression

According to Global Burden of Disease (GBD) 2017 Disease and Injury Incidence and Prevalence Collaborators, globally, more than 264 million people suffer from depression. It is a leading cause of disability worldwide and a major contributor to the overall global burden of disease. Further, the burden of depression and other mental health conditions is on the rise globally.

Mental disorders are also an important cause of disease burden in South Africa. The South African Stress and Health (SASH) study indicated a lifetime prevalence of major depression of 9.7% and a 12-month prevalence of 4.9% (Tomlinson et al., 2009).

2.4.2 There are various challenges with measuring the prevalence of depression

— Firstly, many people with depression do not require medication so relying on chronic medicine registrations will grossly undercount depression prevalence.
— Secondly, depression is not a CDL condition and is therefore not covered on all schemes. Depression is generally under-reported as a result.
— Thirdly, the time period is suboptimal as we only have sight of claims from the time an individual joins a scheme.
— Finally, chronic diseases such as depression may be relapsing and remitting which makes it difficult to determine who is currently suffering from the condition (especially if the case definition relies on chronic medicine registrations that remain in place even when the person is no longer on medication for the condition).

2.4.3 Depression can be defined based on three overlapping case definitions, or any combination of these case definitions. Table 4 and Figure 3 illustrate the divergent depression prevalence results based on various definitions of depression.

Table 4. Depression prevalence results

Case definition	Proportion of total
Care plan for depression	74%
Chronic medicine registration for depression	39%
At least two separate months with relevant ICD-10 code(s) linked to claims	69%
All three case definitions	21%

2.5 Definition of HIV prevalence

2.5.1 According to Sharp et al. (2011), Acquired Immune Deficiency Syndrome (AIDS) was first recognised as a new disease in 1981. A retrovirus, now termed human immunodeficiency virus type 1 (HIV-1), was subsequently identified as the causative agent of what has since become one of the most devastating infectious diseases to have emerged in recent history. HIV-2, identified in 1986, is rarely found outside western Africa.

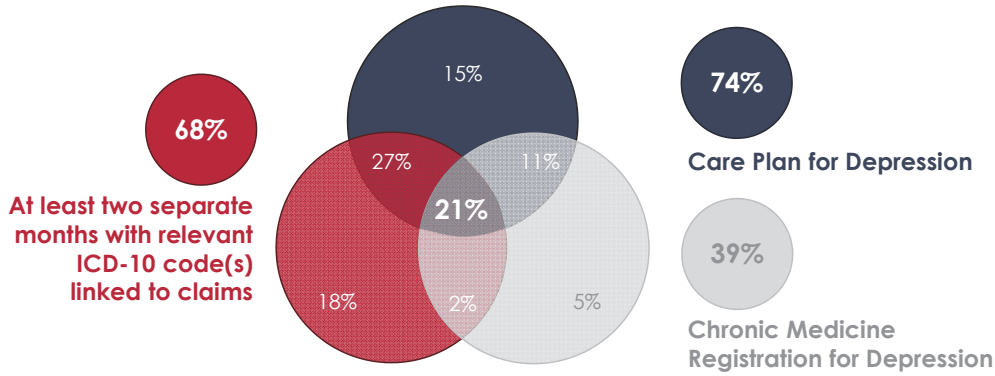


FIGURE 3. Venn diagram illustrating various definitions of depression prevalence for a group of medical schemes

HIV continues to be a major global public health issue.[18] There is, as yet, no cure or vaccine. However, with increasing access to effective HIV prevention, diagnosis, and treatment including with antiretroviral therapy (ART), HIV infection has become a manageable chronic health condition.[19]

Since the beginning of the epidemic, approximately 84.2 million people worldwide have been infected with HIV and about 40.1 million people have died of HIV. Globally, approximately 38.4 million people were living with HIV at the end of 2021.[20]

There were an estimated 8.45 million people living with HIV in South Africa in 2022, and the overall HIV prevalence rate was estimated to be approximately 13.9%.[21] For adults aged 15–49 years, 19.5% of the population was estimated to be HIV positive in 2021.[22]

2.5.2 The definition of HIV prevalence is critical to accurately measure the 95–95–95 targets as well as other measures such as ART adherence. HIV can be defined based on HIV Management Programme registrations, ART claims, pathology tests or other claims markers.

The challenge with the HIV case definition is those who have registered on the HIV management programme but are not claiming ART. There could be a number of possible explanations for this and the measures in Table 5 can be adjusted accordingly: these beneficiaries may be paying for ART out of pocket.

— These beneficiaries may be claiming ART from the State (or a wellness programme). This could be confirmed by the beneficiary when they are contacted.

18 https://www.who.int/data/gho/data/themes/hiv-aids
19 https://www.who.int/news-room/fact-sheets/detail/hiv-aids
20 https://www.who.int/data/gho/data/themes/hiv-aids
21 Statistics South Africa. Mid year population Estimates 2022.
22 Statistics South Africa. Mid-year population estimates 2021.

— These beneficiaries may be elite controllers. According to Gebara et al. (2019), HIV type 1 (HIV-1) elite controllers (ECs) represent a rare group of individuals with an ability to maintain an undetectable HIV-1 viral load over time in the absence of previous antiretroviral therapy.
— These beneficiaries may have been registered incorrectly in the first place.

Table 5. HIV case definitions

HIV measure	HIV case definition
First 95 target: People living with HIV who know their HIV status[23]	We assume that everyone who registered or is claiming for HIV knows their HIV status.
Second 95 target: People who know their HIV-positive status who are on antiretroviral therapy (ART)[24]	The denominator for this measure is people registered with the HIV management programme.
Third 95 target: People living with HIV who have suppressed viral loads[25]	The denominator is the number of people registered with the HIV management programme who are currently on ART and started ART at least six months ago.
Adherence to antiretroviral therapy (ART)	We report on: i) those who registered on the HIV management programme with no ART claim in the last six months; and ii) the number of months that ART was claimed by those registered on the HIV management programme who claimed ART in the last six months.

If viral load results are available for these beneficiaries this can be used to confirm prevalence as well as evaluate viral load suppression.

2.5.3 The CMS report entitled "Quality of Care in Medical Schemes" (2024)[26] includes the enrolment into disease management programmes in 2022. HIV has the third highest number of registered beneficiaries at 449 137 (or 5.0% based on 9 039 259 beneficiaries reported in the CMS Annexures for 2022[27]). HIV registrations grew by 67.2% between 2015 and 2022.

2.6 Definition of cancer prevalence

2.6.1 Cancer refers to a large group of diseases that are characterised by abnormal cells that grow beyond their usual boundaries and may spread to other organs (World Health

23 https://indicatorregistry.unaids.org/indicator/people-living-hiv-who-know-their-status
24 https://indicatorregistry.unaids.org/indicator/people-living-hiv-antiretroviral-therapy
25 https://indicatorregistry.unaids.org/indicator/people-living-hiv-who-have-suppressed-viral-loads
26 https://www.medicalschemes.co.za/download/3738/research-note-2024/28789/quality-of-care-in-medical-schemes
27 https://www.medicalschemes.co.za/download/3697/industry-report-2022-23/28137/annexures-to-the-2022_23-industry-report.xlsx

Organization (WHO)).[28] The transformation of normal cells into cancer cells results from the interaction between a person's genetic factors and three categories of external agents: 1) physical carcinogens (such as ultraviolet radiation); 2) chemical carcinogens (such as tobacco smoke and alcohol); and 3) biological carcinogens (such as infections from certain viruses, bacteria, or parasites). Cancer is a leading cause of death worldwide, with the most common cancers including breast, lung, colorectal and prostate cancer.

2.6.2 Measuring cancer prevalence can be challenging because cancer may not be a lifelong chronic illness. Episodes of cancer can be followed by periods of remission and the cancer may or may not recur. Depending on the purpose of the investigation, it may be important to determine who is currently being managed for cancer versus those with a past history of cancer that has been treated or that is in remission.

Further, since cancer authorisations tend to remain open for extended periods of time, a registered and claiming definition is more appropriate. The exception is newly diagnosed cases where claims may not have been incurred or received yet. Newly diagnosed cases are counted as prevalent whether or not there are claims.

2.6.3. The SEER[29] programme in the United States of America defines prevalence in various ways.

Table 6. Cancer prevalence definitions

Limited-duration prevalence	Represents the proportion of people alive on a certain day who had a diagnosis of the disease within the past x years (e.g. x = 5, 10 or 20 years). Registries of shorter duration, less than 40 or 50 years of data collection, can only estimate limited-duration prevalence. Limited-duration prevalence can be further classified into periods from year of diagnosis. Thus, the 20-years prevalence can be further classified into the prevalence of those diagnosed in the last 0 to < 5 years, 5 to < 10 years, 10 to < 15 years, and 15 to < 20 years. National Cancer Institute's SEER Program has information on cancer cases since 1975, thus a maximum of 46-year prevalence can be estimated from SEER cases diagnosed from 1975 through 2020.
Complete prevalence	Represents the proportion of people alive on a certain day who were diagnosed with the disease, regardless of how long ago the diagnosis was made. Complete prevalence can be estimated from self-reported population-based surveys (Byrne et al., 1992), although one must be concerned with under-reporting and misclassification of disease. Direct computation (the counting method) of complete cancer prevalence requires registry data that has been collected over a sufficiently long period of time to capture all prevalent cases of the disease. In the United States, the only registry with sufficient incidence and follow-up data to approximate complete cancer prevalence is the Connecticut Tumor Registry (Connelly et al., 1968; Gershman et al., 1976).

28 https://www.who.int/news-room/fact-sheets/detail/cancer
29 https://surveillance.cancer.gov/prevalence/measures.html

Care prevalence	Care prevalence is an estimate of prevalent cases that are still under care. Since population-based cancer data do not include follow-up information on cancer care, estimation of care prevalence is problematic. The SEER-Medicare linked data allow for longitudinal tracing of individuals with cancer using information from the Medicare claims. Mariotto et al. (2003) have estimated the prevalence of patients with colorectal cancer age 65 and older who are under care in the U.S.
Non-cure prevalence / Cure prevalence	Non-cure prevalence is an estimate of prevalent cases that have not been cured of disease. Statistical approaches (e.g. assuming survival models which are a mixture of cured and uncured patients) have been applied to model cure prevalence. (See Capocaccia & De Angelis, 1997; Coldman et al., 1992.)

There are several methods for determining which tumours to include, depending on the question of interest.

Table 7. Definition of malignancies

First malignant primary only	This method would include only the first malignant tumour. The SEER registries collect the number (but not the site or behaviour) of cancers that occur prior to the start of the registry, or prior to the person moving to a SEER catchment area. We make the assumption that these cancers are malignant, as is true of the majority of SEER cancers. Thus, if SEER indicates that the first SEER registered tumour is the person's second tumour (the other was a non-SEER cancer) this person's cancers are excluded from this method. This is the standard method used to calculate prevalence statistics at the NCI. It only counts a person once, and is useful if the user would like to sum various prevalence estimates across cancer sites without double counting individuals (i.e., the prevalence of specific cancer sites adds up to the prevalence of all cancers combined).
First malignant tumour per site in the last x years (observation period)	This method includes the first malignant tumour per cancer site diagnosed in SEER during the entire observation period. If 20-year prevalence on 1 January 1999 is being calculated, only the first malignant tumour per cancer site diagnosed from 1979 through 1998 would contribute to prevalence. For this method, the sum of the prevalence counts across specific sites may be greater than the prevalence for all sites combined. This is true because a person with multiple cancers may contribute to more than one estimate.
First malignant tumour per site and years since diagnosis	This method includes the first malignant tumour diagnosed in SEER for each considered site and each period since diagnosis. The sum of the prevalence counts across specific sites and across non-overlapping periods since diagnosis may be greater than the prevalence for all sites combined and for the total period, respectively. This is because a person with multiple cancers may contribute to more than one site-specific prevalence and to more than one period since diagnosis for the same site.

2.6.4 Given the challenges mentioned in ¶2.6.2 above, Medscheme has defined oncology prevalence as follows.

Table 8. Definition of cancer prevalence

The case definition for considering a beneficiary to be eligible to be included in the calculation of cancer prevalence is based on the "registered and claiming" description for disease prevalence.		
Scheme and plan cancer prevalence (across all cancers), case definition	In a given month, a beneficiary who is currently active on the scheme, has had a relevant oncology authorisation approved in the last 12 months (registered) and who has undergone active cancer treatment (claiming) within the last 12 months.	
Prevalence by cancer type, case definition	In a given month, a beneficiary who is currently active on the scheme, has had a relevant oncology authorisation with the relevant cancer ICD-10 code approved in the last 12 months (registered) and who has undergone active cancer treatment (claiming) within the last 12 months.	Each cancer type is defined by a grouping of related cancer ICD-10 codes.

An oncology claim is defined as follows:

— A claim with a primary cancer ICD-10 code has been paid for at least one of the specified tariff codes, subject to the included practice types.
— A claim with a primary cancer ICD-10 code has been paid for at least one of the chemotherapy drugs, subject to the included practice types.
— A claim with a primary cancer ICD-10 code has been paid for at least one of the specified specialised oncology drugs, subject to the included practice types.

In terms of the four definitions below, the monthly rate is equivalent to "point prevalence" and the exposure-weighted average monthly rate over a given time period is similar to "period prevalence" except that we are counting beneficiary exposure months in the numerator and not "whole" beneficiaries. Using beneficiary exposure months avoids over-stating prevalence due to beneficiaries exiting the scheme either through death or membership termination.

— Scheme/Plan Cancer Prevalence (across all cancers), per month:
  – **Numerator:** The sum of the number of eligible lives by scheme/plan for prevalence (registered and claiming in the last 12 months) and the number of eligible lives for incidence who haven't claimed yet (registered for the first time in last xx months), in the given month.
  – **Denominator:** The number of active beneficiaries by scheme/plan in the given month
— Scheme/Plan Cancer Prevalence (across all cancers), exposure-weighted average per month for specified reporting period:
  – **Numerator:** The sum of the number of eligible lives for prevalence (registered and claiming in the last 12 months) and the number of eligible lives for incidence who haven't claimed yet (registered for the first time in last xx months), for all months in the specified reporting period (i.e. sum of beneficiary-exposure months for prevalence and non-claiming incidence lives).
  – **Denominator:** The sum of the number of active beneficiaries by scheme/plan for

all months in the specified reporting period (i.e. sum of beneficiary-exposure months for scheme/plan lives).

— Prevalence by Scheme/Plan and Cancer Type, per month:
  - **Numerator:** The sum of the number of eligible lives by scheme/plan and cancer type for prevalence (registered and claiming in the last 12 months) and the number of eligible lives by scheme/plan and cancer type for incidence who haven't claimed yet (registered for the first time in last xx months), in the given month.
  - **Denominator:** The number of active beneficiaries by scheme/plan in the given month

— Prevalence by Scheme/Plan and Cancer Type, exposure-weighted average per month for specified reporting period:
  - **Numerator:** The sum of the number of eligible lives by scheme/plan and cancer type for prevalence (registered and claiming in the last 12 months) and the number of eligible beneficiaries by scheme/plan and cancer type for incidence who haven't claimed yet (registered for the first time in last xx months), for all months in the specified reporting period (i.e. sum of beneficiary-exposure months for prevalence and non-claiming incidence lives).
  - **Denominator:** The sum of the number of active beneficiaries by scheme/plan for all months in the specified reporting period (i.e. sum of beneficiary-exposure months for scheme/plan lives).[30]

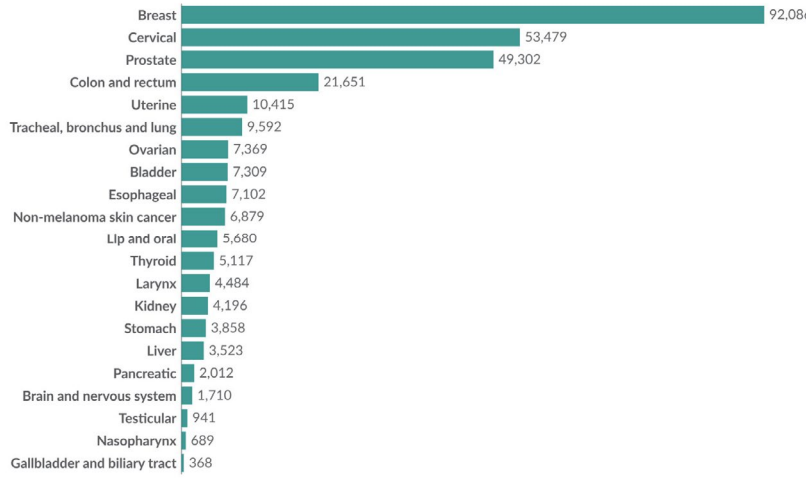


FIGURE 4. Number of people with cancer by type, South Africa 2021(estimated number of people with each form of cancer)

Data source: IHME, Global Burden of Disease (2024), OurWorldinData.org/cancer|CC BY

30 Institute for Health Metrics and Evaluation (IHME). GBD Results. Seattle, WA: IHME, University of Washington, 2024. Available from https://vizhub.healthdata.org/gbd-results. (Accessed 21 September 2024)

2.6.5 Each type of cancer has different incidence rates, prevalence rates and survival rates and should be reported accordingly. Further, within each type of cancer there are relevant classifications, such as HER2 positive breast cancer. As cancer treatment becomes more and more personalised, and if the data is available, we will see cancer prevalence reporting based on these definitions.

3. REVIEW OF DISEASE PREVALENCE DEFINITIONS

3.1 SRM (REF) criteria under the HUASR requirements from the CMS

3.1.1 Background to the Risk Equalisation Fund

According to Murove & Khumalo (2015), the REF is a Risk-Equalised Performance Indicator, which involves remuneration of private healthcare funders based on performance against various cost and quality criteria and non-financial incentives. This incentive model takes into account age, gender, chronic diseases and other factors (including the level of benefits available) for each patient when setting the target cost of treatment.

According to McLeod & Grobler (2010), the intention of the Risk Equalisation Fund (REF) was that medical schemes would no longer compete on the basis of risk selection but on the basis of cost-effective delivery of healthcare. If the REF was implemented, medical schemes that were successful at reducing the cost of delivery would retain that benefit for their members while those that were not successful would need to charge members for the difference.

According to McLeod et al. (2004: 42), the principles for the choice of risk factors in the REF formula were as follows.

Table 9. Risk factors in REF formula

Characteristic	Explanation
Validity	The risk factors should predict the need for medical care and define a system of adjustment in which the cells are relatively homogeneous.
Reliability	The risk factors should be measured without measurement errors.
Availability	The risk factors should preferably be data items that are already collected by medical schemes or that are readily available in the industry.
Feasibility	Obtaining the risk factors for all beneficiaries should be administratively feasible without undue expenditure of time or money.
Measurable and auditable	The risk factors need to measurable, objective, repeatable and auditable.
Invulnerability to manipulation	The risk factors should not be subject to manipulation by medical schemes, managed care organisations, administrators, providers, intermediaries or the beneficiaries.
No perverse incentives	The risk factors should not provide incentives for inefficiency or low quality care.
Legislative consistency	The use of the risk factors needs to be consistent with provisions in the Medical Schemes Act, the National Health Act and the Constitution of South Africa.
Privacy	The risk factors should not conflict with the right to privacy of the beneficiary and healthcare provider.

3.1.2 Inclusion of chronic diseases as a factor in the REF formula

The rationale for the inclusion of chronic diseases included in the Chronic Disease List as risk factors in the REF formula is described in McLeod et al. (2004: 71): "The Determination of the Formula for the Risk Equalisation Fund in South Africa".

Early in the consultative process it was strongly felt that using age and gender (or deliveries) would be insufficient for risk equalisation and thus some measure of chronic disease burden would be necessary.

Osburn & McLeod (2003) reported that various studies have shown that major improvements can be achieved by extending the set of risk adjusters with measures of prior utilisation or measures of chronic health status. Indirect measures of health status may perhaps be measured more reliably than direct indicators of health status such as the presence or absence of certain chronic conditions. They considered various health proxies, including prior costs, chronic health indicators, in-patient diagnostic information and self-reported health status.

According to McLeod et al. (2004), there were several over-riding considerations in South Africa at the time of constructing the REF formula, most notably the lack of or poor status of coding of health events and the need for a predominantly prospective approach that would incentivise efficiency. The use of prior costs, actual expenditure on prescription medicine or previous hospitalisations, as well as in-patient diagnostic information were thus ruled out for consideration. Self-reported health status was not considered a reliable instrument and it was feared this could be subject to manipulation. The use of diagnosis related groupings (DRGs) was contemplated. However, the coding required for analysis by DRGs was not fully implemented at the time of constructing the REF formula.

Even if the coding was available, agreeing on which DRG model to use would have been challenging. Using a DRG model could have been subject to manipulation because South Africa does not have a central DRG model.

According to McLeod et al. (2004), with the introduction in 2004 of the CDL conditions as part of prescribed minimum benefits, attention was focused on these chronic diseases. It was resolved to explore the possibility of using the numbers of beneficiaries with the 25 CDL conditions and this is what became part of the REF Formula. HIV was also added as a key measure of disease burden.

According to Grobler et al. (2003: 30), cancer was considered but excluded since it was felt that the chronic pre-authorisation data used to identify the diseases was not an optimal source for identifying cancer patients. As a consequence, oncology prevalence was not included in the REF formula. Section 2.6.4. includes a discussion of how to calculate oncology prevalence.

3.1.3 Methodology to determine numbers with chronic disease

According to McLeod & Grobler (2008), the risk factors in the formula are predominantly prospective and are as follows:

— Age last birthday on 1 January, summarised into age bands Under 1, 1–4, 5–9, 10–14, …, 75–79, 80–84, 85+.
— Gender (recommended for inclusion from 1 January 2007 but not yet implemented).
— The 25 PMB CDL conditions. Where a beneficiary has more than one chronic condition the fund may select the most expensive of the conditions.
— HIV/AIDS provided the beneficiary is receiving antiretroviral therapy according to national guidelines.
— An additional factor for multiple chronic conditions with provision for 2, 3, or 4+ simultaneous chronic conditions.
— A retrospective factor for maternity events, defined as the delivery of a single/multiple foetus, either stillborn or alive.

It was considered critical that there was a trusted and fair way to determine the numbers with chronic disease and this resulted in a comprehensive manual of Entry and Verification. There are two elements to the criteria:

— The diagnosis of a particular disease, which includes specification of applicable ICD-10 codes and limitations on the practitioners that may diagnose complex conditions. There may be mandatory tests to differentiate between diseases and results must be retained by the fund; and
— a proof of treatment element which is based on paid claims data. Claims for medicine benefits for at least two of the three calendar months prior to the month of submission are typically required in order to demonstrate proof of treatment. The applicable medicines that can be used as proof are classified using the anatomical therapeutic chemical (ATC) classification and payment must have been made from the risk pool, not personal medical savings accounts.

According to McLeod & Grobler (2008: 6), rules for multiple chronic conditions within disease groups were created to deal with substantial "code-creep" that seemed to be occurring. There were concerns that there was not enough difference between the diagnoses for certain conditions. The approach adopted was similar to the concept of hierarchical condition categories used by Ellis (2007). Rules for multiple diseases within disease groups were established which require that only one of the diseases be selected in each group.

3.1.4 Challenges experienced with inclusion of chronic diseases as a factor in the REF formula

According to McLeod & Grobler (2008), the adage that "you get what you incentivise" has been much in evidence in the collection of data on chronic disease during the shadow period. It is sobering to reflect that the behaviour of administrators and funds with respect to submissions has been on the promise of future risk equalisation transfers as no money has yet changed hands.

According to McLeod & Grobler (2008), the initial design of the formula was based on 41% of industry beneficiaries but only two of the nearly 30 administrators. The first disease data for the entire industry, in respect of the first quarter of 2005, showed massive over-reporting of as much as 800% for the most expensive chronic diseases. Over-reporting of Under 1s and under-reporting of the age 85+ group were also found. In all cases the direction of over-reporting was in the direction of age groups and diseases with greater payment from REF.

- Over the first year, tools and skills were rapidly developed to identify data that seemed unusual for the age profile of each fund. The data problems were found to be linked to particular administrators and differences persisted, despite the introduction of the entry and verification criteria.
- The two administrators that had participated in the first study in 2002 had gained substantial knowledge about REF data collection and their data were usually clean on submission. The other administrator groups did not implement the criteria at the same time and have data patterns that are much more irregular.
- One administrator added in the REF Study 2005 had made use of an "auto-chronic" process to identify chronic lives. In a full auto-chronic process a proxy diagnosis for a CDL disease is made using the medicines dispensed and an in-house crosswalk from medicine codes to diagnosis codes used to obtain the relevant disease. A less extreme form of this is to use diagnosis codes submitted with hospital and other claims data to determine the disease. Following a detailed study, a decision was taken that auto-chronic definitions using proxy diagnosis from medicines or ICD-10 codes derived from claims were no longer acceptable. The only chronic definition acceptable is where there is a granted authorisation for a CDL disease as part of a chronic medicine review process.
- In order to be counted as a chronic patient for the purposes of REF, a patient has to meet the "treated patient" criteria but this has created some perverse incentives, including paying managed care organisations additional fees to find more chronic patients. Many funds have introduced automatic postal delivery of chronic medicine to attempt to prove regular claims, irrespective of whether all the prior medicine has been used, and this increases wastage.
- The definition of treatment has been developed using only allopathic drugs and this has created an unintended barrier to other forms of treatment. In particular, it has disincentivised wellness programmes and life-style modification for chronic conditions. Complementary medicine and African traditional medicine do not currently count towards the "treated patient" criteria.

According to Armstrong et al. (2004), the verification criteria were developed with the emphasis on the verifiability of cases and are used to ensure that gaming of the REF is identified and addressed.

According to the "Report of the International Review Panel to the Risk Equalization Task Group" (Armstrong, 2004: 30, 31), the panel recommended that, all other things being equal, CDL should be used as a factor for equalisation. However, it was viewed as impractical for the following reasons:

— The "cross-over" algorithms that the Formula Consultative Task Team needed to develop for the purposes of considering CDL as a factor suggest that the definitions used currently may be imprecise.
— There are no established criteria or protocols for a review of the CDL.
— Many medical schemes are currently unable to report the CDL that apply to their beneficiaries.
— The assignment of CDL to beneficiaries may be subjective and not clearly defined.
— The criteria for monitoring that the assignment of CDL to beneficiaries continues to be valid over time are not clear. For example, childhood asthma may ameliorate with age.
— Providers and schemes may have a financial incentive to up-code their categorisation of beneficiaries.
— The auditing of CDL categorisations is likely to prove problematic.

The Panel recognised the empirical evidence of a link between the CDL and healthcare expenditures, and the underlying rationale that the CDL reflects morbidity characteristics of beneficiaries. However, until such a time that the issues raised above are resolved, the Panel advised that the CDL is not yet usable to its full extent as a basis for equalisation. The Panel suggested that medical schemes were encouraged to collect accurate information on assignment of the CDL to beneficiaries and recommended a gradual phasing-in of the CDL component in the calculation of the REF, starting with a weight of 10% in the first instance.

The National Health Insurance Policy Brief 3[31] published in 2009 included an evaluation of the CDL Disease Prevalence: Diagnosis and Treated. The analysis compared revised prevalence after the application of the multiple disease rules.

Data from four administrators was used for the analysis: Discovery Health, Medscheme, MHG and Old Mutual Healthcare, which collectively represented 63.4% of the beneficiaries in the industry. [32] The analysis was based on data on prevalence and PMB expenditure for calendar 2005. The REF Entry and Verification Criteria v2 was evaluated, which was effective from 1 January 2007 for determining diagnosis and treatment.

The analysis used the final REFCT2007 for the order of diseases for multiple rules. Diagnosed Cases Prevalence were originally published as "CASES" and Treated Patient Prevalence was originally published as "TREATED".

31 https://slideplayer.com/slide/10753591/
32 https://www.medicalschemes.com/files/Circulars/Circular_12_of_2007_Completion_Of_Ref_Study_2005.pdf

The conclusion was that for some diseases there was a very large number of patients who have been diagnosed with a disease but are not meeting the "treated patient" criteria. This means they are not receiving medication from the risk pool in the medical scheme, which, according to the REF verification criteria, for most diseases is two out of every three months.

3.1.5 HUASR entry and verification criteria

For the annual Scheme Risk Management (SRM) submission, medical schemes define prevalence as diagnosed and treated prevalence as defined by the SRM process.

Medical schemes use the HUASR Guidelines for the Identification of Beneficiaries with Risk Factors in Accordance with the Entry and Verification Criteria v15.1 (CMS, 2022)[33] to identify each chronic disease case.

According to Cairncross et al. (2019) in the CMS report the terminology "treated prevalence" is used to describe the criteria for inclusion, which include proof of diagnosis as well as proof of treatment for each CDL condition and HIV. In addition, a hierarchy is applied when counting chronic conditions for the annual SRM submission.

The HUASR Guidelines for the Identification of Beneficiaries with Risk Factors in Accordance with the Entry and Verification Criteria v15.1 (CMS, 2022)[34] includes the CDL ranks on page 12 and the chronic condition assignment rules on page 12 and 13:

- For count purposes, only one of the following chronic respiratory diseases can be assigned to the same patient: chronic obstructive pulmonary disease, bronchiectasis and asthma.
- For count purposes, only one of the following cardiovascular diseases can be assigned to the same patient: cardiomyopathy and cardiac failure, coronary artery disease, dysrhythmias; and hypertension.
- For count purposes, only one of chronic renal disease or hypertension may be assigned to the same patient.
- For count purposes, only one of the following gastrointestinal conditions can be assigned to the same patient: Crohn's disease or ulcerative colitis.
- For count purposes, only one of the following psychiatric conditions can be assigned to the same patient: bipolar mood disorder or schizophrenia.
- For count purposes, only one of the following neurological/psychiatric conditions can be assigned to the same patient: multiple sclerosis, bipolar mood disorder, or epilepsy.

33 https://www.medicalschemes.co.za/wpfd_file/guidelines-for-the-identification-of-beneficiaries-with-risk-factors-in-accordance-with-the-entry-and-verification-criteria-version-15-1/#:~:text=Guidelines%20for%20the%20Identification%20of%20Beneficiaries%20with%20Risk

34 https://www.medicalschemes.co.za/wpfd_file/guidelines-for-the-identification-of-beneficiaries-with-risk-factors-in-accordance-with-the-entry-and-verification-criteria-version-15-1/#:~:text=Guidelines%20for%20the%20Identification%20of%20Beneficiaries%20with%20Risk

— For count purposes, only one of the following auto-immune conditions can be assigned to the same patient: systemic lupus erythematosus or rheumatoid arthritis.
— Diabetes mellitus type 1 and type 2 cannot co-occur.

3.1.6 Practical considerations emanating from the recent HUASR submissions

Some anomalies have been detected, for example, in the recent HUASR submissions, the reported haemophilia prevalence was observed to be much lower than anticipated based on registrations and claims. According to Benson et al. (2018), haemophilia is usually an inherited bleeding disorder most commonly caused by deficiencies of the coagulation factors VIII (haemophilia A) and IX (haemophilia B). Factor VIII and factor IX plasma concentrations determine bleeding tendency, which is classified as mild, moderate, or severe. The definition of mild haemophilia (factor levels >0.05–0.40 IU/mL; 5–40% of normal) is broad compared with that of the moderate (0.01–0.05 IU/mL; 1–5% of normal) and severe (<0.01 IU/mL; <1% of normal) forms.

Since haemophilia is an X-linked recessive genetic disorder, it predominantly affects males and occurs in about 1 of every 5,000 male births. Haemophilia A is about three to four times as common as haemophilia B, and about half of those affected have the severe form.[35] Treatment for haemophilia is very expensive and haemophilia is top of the list on disease ranking on page 12 of the HUASR Guidelines for the Identification of Beneficiaries with Risk Factors in Accordance with the Entry and Verification Criteria v15.1 (CMS, 2022)[36] However, the diagnosis and treatment criteria in Table 18 on page 33 of the Entry and Verification Criteria v15.1 are quite restrictive, as shown in Figure 5.

As shown in Figure 6, McLeod et al. (2005: 4) in the Definitions of Entry Criteria for Determining the REF Grids paper suggests that the original intention was to include mild haemophilia. Based on the spreadsheets supporting the REF Study 2005, it is evident that the TREATED prevalence was 0,014 per 1000, 0.026 for males and 0.0002 for females. This compares to diagnosis prevalence of 0,028 per 1000, 0.044 for males and 0.012 for females.

As shown in Section 3.2, the prevalence of haemophilia reported across all schemes in 2017 was 0.02 per 1000. General prevalence of haemophilia under the relaxed definition was 0.07. Haemophilia is a rare disease and reported prevalence is likely to be volatile, particularly in the smaller medical schemes. However, treated prevalence recently submitted for a large scheme was well below the expected treated prevalence. It appears that the criteria may have been applied in too restrictive a manner. The algorithms applied need to be reviewed to ensure that cases are correctly identified and reported on. The expected prevalence may need to be revised as it was based on data from 2005.

35 https://www.cdc.gov/hemophilia/about/index.html
36 https://www.medicalschemes.co.za/wpfd_file/guidelines-for-the-identification-of-beneficiaries-with-risk-factors-in-accordance-with-the-entry-and-verification-criteria-version-15-1/#:~:text=Guidelines%20for%20the%20Identification%20of%20Beneficiaries%20with%20Risk

Diagnosis-related information				Proof of Treatment	
Provider code of the diagnosing provider	AND	ICD-10 Codes (any of the following)	AND	Evidence of payment of claims for any product included in the ATC categories below, for services/ treatment that was provided in one calendar month in the three calendar months preceding the current month:	
Any registered medical practitioner		D66 D67		B02AA02 B02BD02 B02BD03 B02BD08	B02BD04 B02BD06 H01BA B02BX06
		AND			
		Laboratory evidence of Factor VIII or IX levels lower than or equal to 5%			

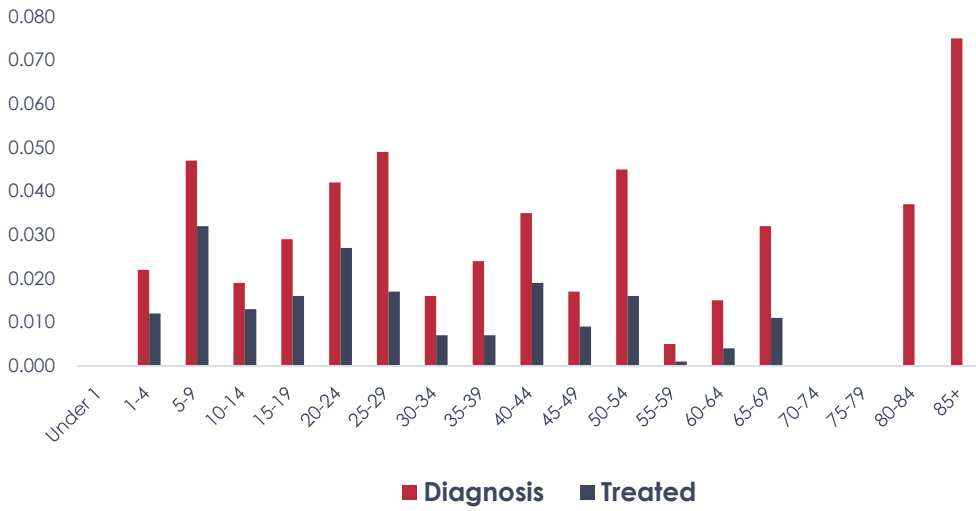
FIGURE 5. Diagnosis and treatment criteria for haemophilia in Table 18 of the HUASR guidelines for the identification of beneficiaries with risk factors in accordance with the entry and verification criteria v15.1

PMB-CDL Algorithm Definition: The definitions for Haemophilia A and B are summarised in the following table:

	Mild	Moderate	Severe
Haemophilia A	Factor VIII 5-40% of normal	Factor VIII 1-5% of normal	Factor VIII < 1% of normal
Haemophilia B	Factor IX 5-25% of normal	Factor IX 1-5% of normal	Factor IX < 1% of normal

REF Entry Criteria: Haemophilia A: Factor VIII < 40% of normal | Haemophilia B: Factor IX < 25% of normal

FIGURE 6. REF entry criteria for haemophilia according to McLeod et al. (2005: 4)



Treated Prevalence: Males 0,026 | Females 0,002 | Combined 0,014

FIGURE 7. Haemophilia reported under REF criteria (rate per 1000 beneficiaries in each age band (the prevalence of chronic disease in each age band is taken from the REF Study 2005)

3.2 Annexure A prevalence definition in CMS report

According to Cairncross et al. (2019) in the CMS report, an alternative definition of prevalence was introduced.

Annexure A of the CMS report defines prevalence in a more relaxed manner without entirely being guided by the entry and verification criteria referred to in Section 3.1.5.

Under the relaxed definition, prevalence is defined as the count of beneficiaries who have had at least one claim for a specified CDL condition during the year (rather than the two out of three in the last three months). The relaxed definition data were collected from all medical schemes for the first time in 2015.

Table 10. Average prevalence per 1000 beneficiaries across all schemes in 2017

Chronic condition		HUASR Entry and verification criteria	General prevalence (relaxed definition)
DM2	Diabetes Type 2	31.28	51.82
HAE	Haemophilia	0.02	0.07
HIV	HIV	25.12	40.27
HYP	Hypertension	90.64	160.62

3.3 CMS definition

The CMS report entitled "Quality of Care in Medical Schemes" (2024)[37] notes that the numerator must correspond to the denominator.

The report notes that coverage ratios are subject to measurement errors in both the numerator and the denominator. They also do not account for those beneficiaries who are utilising tests even though they are not registered on Disease Management Programmes (DMPs). If utilisation is not driven by the membership to a DMP, then we need to expand the scope of the analysis to include beneficiaries who are not registered but utilising tests and understand what the underlying drivers for utilisation are.

The two variables needed to calculate coverage ratios are the number of beneficiaries who utilised standard tests/procedures (numerator) and beneficiaries registered on Mos (denominator). The former is extracted from the count of utilisation claims for tests/procedures this year and the latter is extracted from the count of beneficiaries who submitted a chronic claim and are registered on a DMP this year.

3.4 Adjusted Clinical Groups methodology

The Johns Hopkins Adjusted Clinical Groups (ACG®) System[38] uses diagnosis, pharmacy and pathology data and the tools produce statistically valid, actionable information and

37 https://www.medicalschemes.co.za/download/3738/research-note-2024/28789/quality-of-care-in-medical-schemes

38 https://www.hopkinsacg.org/applications/

insights that can help medical schemes improve quality of care, better predict resource utilisation and reduce costs and inefficiency. The ACG® System is built to handle the complexities of health care information flows, disparate data sources and diverse coding standards. It uses data from individual patients' primary and secondary care records, which makes it suitable for use in a wide range of settings.

One of the ACG® System applications is population profiling. Adjusted clinical groups allow a medical scheme to:

— measure the morbidity distribution, disease prevalence and medication adherence of a patient population,
— identify population risk factors that may contribute to hospitalisation, psychosocial conditions, frailty, high costs or care coordination risk,
— forecast health care utilisation for the population by cost and type of hospitalisation, and
— stratify the population based on their patterns of disease and resource use.

ACGs can be used to calculate apparent prevalence of a disease for a particular population.

According to Orueta et al. (2012) codes for diagnoses and prescriptions were collected for all patients in the Basque Country over 14 years of age (n=1,964,337) for a 12-month period. A range of different inputs were included: hospital diagnoses, primary care diagnoses, primary care prescriptions and combinations thereof. Data were collapsed into the morbidity groups specified by the Johns Hopkins ACG® Case-Mix System. The prevalence of 12 chronic conditions was estimated and compared.

3.5 Effective coverage framework

Under the Effective Coverage Framework[39] disease prevalence is defined as beneficiaries registered on the relevant disease management programme at some point within the relevant calendar year. Coverage proportions are then calculated as a proportion of this total count of beneficiaries. Under this framework, medical schemes are required to estimate the predicted or expected population disease prevalence rates. According to the Effective Coverage Framework document, this will help to identify the potential undiagnosed disease burden in the scheme. Expected prevalence is provided in spreadsheet format for:

— hypertension prevalence rates by age, and
— diabetes prevalence rates by age.

3.6 Health quality assessment prevalence definitions

Health Quality Assessment (HQA)[40] performs an annual assessment of clinical quality in health care offered by medical schemes through the use of health care quality indicators.

39 https://www.bhfportal.co.za/bhfglobal/resources/Effective_Coverage_Mannual.pdf
40 https://www.hqa.co.za/about-us

The aim of such assessments is to assist decision-makers, such as trustees and managers of medical schemes, to evaluate and improve the quality of health care received by their members. HQA is a not-for-profit company, established in 2000, and is governed by a board including representatives from the Board of Healthcare Funders of Southern Africa (BHF) and the SA National Consumer Union (SANCU). The CMS is also an active participant and enjoys permanent observer status.

Under the HQA quality measures framework, disease prevalence is based on chronic registrations and is calculated for COPD, ischaemic heart disease, diabetes, asthma, HIV, bipolar, schizophrenia, cardiac failure, hypertension, depression, rheumatoid arthritis and hypothyroidism. For diabetes, for example, three measures are included.

Table 11. Disease prevalence

Indicator description	Numerator	Denominator
Proportion of beneficiaries registered for diabetes	Average number of beneficiaries registered on a chronic disease management programme for diabetes in the past year	Average number of beneficiaries in the past year
Proportion of beneficiaries registered for diabetes with additional chronic diseases	Number of beneficiaries with one or more additional chronic conditions, registered on a chronic disease management programme for diabetes in the past year	Average number of beneficiaries registered on a chronic disease management programme for diabetes in the past year
Proportion of beneficiaries with diabetes registered for Hypertension	Number of beneficiaries registered on a chronic disease management programme for diabetes and hypertension in the past year	Average number of beneficiaries registered on a chronic disease management programme for diabetes in the past year

Accurate data relating to disease prevalence is fundamental to the HQA measures as it impacts the prevalence measures in Table 11, the prevalent population for all other measures and the risk adjustment applied to some of the measures.

3.7 Case definitions for research studies

This section focuses on how medical scheme data is used to determine case definitions for diseases in research studies.

According to Johnson et al. (2023) in the "Model-based approach to estimating the prevalence of disease combinations in South Africa" study, nationally representative prevalence estimates for the population aged 15 and older were obtained from a number of sources. The sources included data from national surveys (that included measurements such as blood pressure and HbA1c tests as well as self-reported diagnosis or symptoms), disease registers, pharmacy and laboratory data. The medical scheme (private sector) data relied on chronic medicine registration and care plan data for identifying most conditions and included relevant ICD-10 codes linked to claims for mental health disorders and chronic kidney disease (≥2 relevant ICD-10 codes ≥100 days apart).

For the IeDEA[41] studies the following are examples of case definitions that have been used.

— Ruffieux et al. (2023) Life years lost associated with mental illness: A cohort study of beneficiaries of a South African medical insurance scheme.
  – Mental health case definition:
  ~ One ICD-10 code was used to identify cases in the main analysis to avoid biasing the sample toward more severe and well-treated mental health patients. This is because patients with milder conditions or those less engaged in treatment might only have a single code recorded. This approach accepts the inclusion of false positives to ensure that a broader spectrum of cases is captured. In the sensitivity analysis, two ICD-10 codes were required, which may miss some true cases (those with only one code) but reduce the likelihood of including false positives.

— Fernández Villalobos et al. (2024) Cervical pre-cancer and cancer incidence among insured women with and without HIV in South Africa.
  – HIV case definition:
  ~ HIV-related ICD-10 diagnoses (B20-24, F02.4, O98.7, R75, Z21)
  ~ HIV-related laboratory tests (positive HIV test, HIV RNA viral load measurement, CD4 cell count measurement)
  ~ ATC codes for antiretroviral therapy (ART)
  ~ Registration in the Aid for AIDS (AfA) disease management programme
  ~ To increase the specificity of the definition, a positive HIV status was assigned to women with two or more HIV indicators and women with only one HIV indicator were excluded from the main analysis. A negative HIV status was assigned to women with no HIV indicator.

3.8 Case definition used in the South Africa Demographic and Health Survey 2016

Statistics South Africa (Stats SA), in partnership with the South African Medical Research Council (SAMRC), conducted the South Africa Demographic and Health Survey 2016 (SADHS, 2016)[42] at the request of the National Department of Health (NDoH). Technical assistance was provided through the DHS Programme. Timely information about the health of the nation is essential for monitoring and evaluation. Survey data collection took place from 27 June 2016 to 4 November 2016. The primary objective of the SADHS 2016 was to provide up-to-date estimates of basic demographic and health indicators.

Both the Woman's and Man's Questionnaires included a module on adult health that captured information on use of tobacco, alcohol, and codeine-containing medications; consumption of fat, salt, sugar, fruit, and vegetables; health care-seeking behaviours;

41 https://www.iedea.org/
42 https://dhsprogram.com/pubs/pdf/FR337/FR337.pdf

and self-reported prevalence of a variety of noncommunicable diseases. The module was administered to all men age 15 and older and to all women age 15 and older in the subsample of households selected for the male survey and biomarker collection.

The Biomarker Questionnaire was used to record data on biomarkers (anthropometry, anaemia testing, blood pressure measurement, HbA1c testing, and HIV testing) collected from respondents by nurses. In addition, for adults age 15 and older, information on prescribed medications was recorded.

According to the SADHS (2016: 271), an HbA1c value of ≥6.5% is used to classify an individual as having impaired glucose homoeostasis, an indicator of diabetes and an HbA1c value between 5.7% and 6.4% classifies an individual as being pre-diabetic, as defined by the American Diabetes Association (ADA, 2010).[43]

The case definition for diabetes is consistent with the case definition used by Medscheme described in section 2.2.2. above in that both use an HbA1c value of ≥6.5% as an indicator of diabetes. The SADHS (2016) measurements show that 46% of women and 44% of men have hypertension. These findings include 9% of women and 6% of men who have blood pressure in the normal range but are taking medication to control their blood pressure. 13% of women and 8% of men aged 15 and older have an adjusted HbA1c level of 6.5% or above, indicating that they are diabetic. Very high proportions of women (64%) and men (66%) are pre-diabetic (adjusted HbA1c level of 5.7%-6.4%).

3.9 International definitions of disease prevalence

The Global Burden of Disease[44] defines prevalence as the total number of cases in the population.

Table 12. Disease burden definition

Measure	Number	Percent	Rate
Prevalence	Total number of cases in the population	Proportion of total cases of a particular cause relative to cases from all causes	Total cases per 100,000 population

The Centers for Disease Control and Prevention[45] definition of prevalence is as follows.

> The number of cases of a disease, number of infected people, or number of people with some other attribute present during a particular interval of time. It is often expressed as a rate (for example, the prevalence of diabetes per 1,000 people during a year).

43 https://diabetes.org/about-diabetes/diagnosis
44 https://www.healthdata.org/sites/default/files/2024-05/IHME_GBD_2021_A3_MEASURE_METRIC_DEFINITIONS_Y2024M05D15.XLSX
45 https://www.cdc.gov/nchs/hus/sources-definitions/prevalence.htm

The SEER definitions for oncology were covered in section 2.6.3.

4. CONCLUDING REMARKS

Disease prevalence is essential to measuring the underlying disease burden in a population, which can be viewed as a risk factor that is a key determinant of claims. This paper has provided some insights into various definitions of prevalence and engenders a deeper appreciation of the complexities and intricacies of determining prevalence. The definition of prevalence forms a foundation for the construction of many quality outcomes measures, which are central to a value-based care framework.

ACKNOWLEDGEMENTS

I acknowledge the contribution of all my colleagues at Medscheme, our clients and external actuaries and clinicians who constantly challenge the status quo and ensure continuous improvement in the realm of value-based care.

I am indebted to Heather McLeod for providing extensive documentation on the construction of the REF.

Finally, I thank Adam Lowe for reviewing the paper and for his invaluable comments and guidance.

ABBREVIATIONS AND ACRONYMS

ACC	American College of Cardiology (US)
ACG®	Johns Hopkins Adjusted Clinical Groups
ADA	American Diabetes Association (US)
AfA	Aid for AIDS
AHA	American Heart Association (US)
aPMBs	Ambulatory Prescribed Minimum Benefits
ART	Antiretroviral Therapy
ATC	Anatomical Therapeutic Chemical
BHF	Board of Healthcare Funders
CD4	Cluster of Differentiation 4
CDC	Centers for Disease Control and Prevention (US)
CDL	Chronic Disease List
CMS	Council for Medical Schemes
COPD	Chronic Obstructive Pulmonary Disease
CUPs	Contracting Units for Primary Health Care p2
DBP	Diastolic Blood Pressure
DMPs	Disease Management Programmes
DRGs	Diagnosis Related Groupings
FCTT	Formula Consultative Task Team
GBD	Global Burden of Disease
HQA	Health Quality Assessment
HbA1c	Haemoglobin A1c / glycated haemoglobin / glycohaemoglobin
HIV	Human Immunodeficiency Virus
HUASR	Health Utilisation Annual Statutory Returns
ICD-9 CM	International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10	International Classification of Diseases, Tenth Revision
IDF	International Diabetes Federation
IHME	Institute for Health Metrics and Evaluation
JNC7	Joint National Committee (US)
MCO	Managed Care Organisation
MHG	Metropolitan Health Group
NHI	National Health Insurance
PMBs	Prescribed Minimum Benefits
REF	Risk Equalisation Fund
REFTG	Risk Equalisation Fund Task Group
RNA	Ribonucleic Acid
SADHS	South Africa Demographic and Health Survey
SAMRC	South African Medical Research Council
SANCU	SA National Consumer Union
SASH	South African Stress and Health
SBP	Systolic Blood Pressure
SEER	Surveillance, Epidemiology, and End Results
SEMDSA	Society for Endocrinology, Metabolism and Diabetes of South Africa
SRM	Scheme Risk Management
Stats SA	Statistics South Africa
WHO	World Health Organization

REFERENCES

Armstrong, J et al. (2004). Report of the International Review Panel to the Risk Equalization Task Group [unpublished].

Benson, G, Auerswald, G, Dolan, G, Duffy, A, Hermans, C, Ljung, R, Morfini, M & Šalek, SZ (2018). Diagnosis and care of patients with mild haemophilia: practical recommendations for clinical management. *Blood Transfus*. 2018 Nov; **16**(6):535–44. DOI: 10.2450/2017.0150-17

Byrne, J, Kessler, LG & Devesa, SS (1992). The prevalence of cancer among adults in the United States: 1987. *Cancer* **69**(8):2154–2159

Cairncross, C et al. (2019). Prevalence of chronic diseases in the population covered by medical schemes in South Africa, Council for Medical Schemes' Research and Monitoring Unit. https://www.medicalschemes.com/files/Research%20Briefs/Prevalenceofchronicdiseasesinthemedicalschemespopulation.pdf

Capocaccia, R & De Angelis, R (1997). Estimating the completeness of prevalence based on cancer registry data. *Stat Med* **16**(4):425–440

Chawla, A et al. (2016). Microvasular and macrovascular complications in diabetes mellitus: Distinct or continuum? *Indian J Endocrinol Metab*. **20**(4):546–51. DOI: 10.4103/2230-8210.183480

Coldman, AJ, McBride, ML & Braun T (1992). Calculating the prevalence of cancer. *Stat Med* 11:1579–1589

Colman, I et al. (2011). Predictors of long-term prognosis of depression. *CMAJ*. **183**(17):1969–76. DOI: 10.1503/cmaj.110676

Connelly, RR, Campbell, PC & Eisenberg H (1968). Central registry of cancer cases in Connecticut. *Public Health Rep* 83: 386–390.

Duncan, I (2005). Paper 6: An actuarial method for evaluating disease management outcomes [unpublished] https://www.soa.org/globalassets/assets/Files/Research/Projects/Paper6-Actuarial-Methodology-for-Evaluating-DM.pdf

Duncan, I et al. (2007). Disease Management Practice Note prepared by the Disease Management Work Group of the Health Practice Council of the American Academy of Actuaries [unpublished] https://www.actuary.org/sites/default/files/files/publications/Practice_Note_Disease_Managemen_dec2007.pdf

Ellis, RP (2007). Risk adjustment in health care markets: concepts and applications, Paper prepared for a special edited volume on International Health Care Financing, sponsored by the Institute of Health Economics (IHE). https://people.bu.edu/ellisrp/EllisPapers/2007_Ellis_Riskadjustment25.pdf

Fernández Villalobos, NV et al. (2024). Cervical precancer and cancer incidence among insured women with and without HIV in South Africa. *Int J Cancer*. **154**(2):273–283. DOI:10.1002/ijc.34707

Folb, N et al. (2016). Socioeconomic and modifiable predictors of blood pressure control for hypertension in primary care attenders in the Western Cape, South Africa. *S Afr Med J*. **106**(12):1241–6. DOI: 10.7196/SAMJ.2016.v106.i12.12005

Gaziano, TA et al. (2017). Cardiometabolic risk in a population of older adults with multiple co-morbidities in rural South Africa: the HAALSI (Health and Aging in Africa: longitudinal studies of INDEPTH communities) study. *BMC Public Health*. **17**(1):206. DOI: 10.1186/s12889-017-4117-y

Gebara, NY et al. (2019). HIV-1 elite controllers: an immunovirological review and clinical perspectives. *J Virus Erad*. **5**(3):163–166. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6816117/pdf/jve-5-163.pdf

Gershman, ST, Flannery, JT, Barrett, H, Nadel, RK & Meigs, JW (1976). Development of the Connecticut Tumor Registry. *Conn Med* 40: 697–701

Global Burden of Disease (GBD) (2017) Disease and injury incidence and prevalence collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018 10;392(10159):1789–858. DOI: 10.1016/S0140-6736(18)32279-7

Grobler, P et al. (2003). Technical report: Risk equalisation in South African Medical Schemes [unpublished]

Gupta, RD et al. (2021) Prevalence and associated factors of hypertension among South African adults: findings from the Demographic and Health Survey 2016. *J Public Health* (Berl.) 31, 967–977. https://doi.org/10.1007/s10389-021-01607-w

Habibzadeh, F et al. (2022) The apparent prevalence, the true prevalence, *Biochem Med* (Zagreb).

Hallberg, SJ et al. (2018). Effectiveness and safety of a novel care model for the management of type 2 diabetes at 1 year: an open label, non randomized, controlled study. *Diabetes Ther* 9:583 612

Johnson, LF et al. (2023). A model-based approach to estimating the prevalence of disease combinations in South Africa. *BMJ Global Health.* http://dx.doi.org/10.1136/bmjgh-2023-013376)

Jongen, VW et al. (2019). Hypertension in a rural community in South Africa: what they know, what they think they know and what they recommend. B*MC Public Health*. **19**(1):341. DOI: 10.1186/s12889-019-6642-3

Jutel, A (2009). Sociology of diagnosis: A preliminary review. *Sociology of Health and Illness*. **31**(2):278–299. [PubMed] https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9566.2008.01152.x

Mariotto, A, Warren, JL, Knopf, KB & Feuer, EJ (2003). The prevalence of colorectal cancer patients under care in the U.S. *Cancer* 98: 1253–61

McLeod, H et al. on behalf of the Formula Consultative Task Team (2004). The determination of the formula for the risk equalisation fund in South Africa, prepared for the Risk Equalisation Fund Task Group. https://www.medicalschemes.com/files/Risk%20Equalisation%20Fund/REF_Task_Group_Jan_2004.pdf

McLeod, H et al. and the Risk Equalisation Technical Advisory Panel (RETAP) (2005) Definitions of entry criteria for determining the REF grids,.RETAP Recommendations Report No. 2 of 2005

McLeod, H & Grobler, P (2010). Risk equalisation and voluntary health insurance: The South African experience. *Health Policy*. **98**(1):27–38. DOI: 10.1016/j.healthpol.2010.06.008

McLeod, H. & Grobler, P. (2008). The experience of using chronic disease for risk equalisation in South Africa. Prepared for the Risk Adjustment Network meeting in Dublin. [unpublished]

Murove, C & Khumalo, N (2015). Managed care – the role of actuaries, presented at the Actuarial Society of South Africa's 2015 Convention. [unpublished]

Noakes, T et al. (2019). Diabetes is reversible: What does a high road scenario look like and how do we get there, presented at the Actuarial Society of South Africa's 2019 Convention. [unpublished]

Oparil, S et al. (2018) Hypertension. *Nat Rev Dis Primers*. 22;4:18014. DOI: 10.1038/nrdp.2018.14

Orueta, JF et al. (2012) Monitoring the prevalence of chronic conditions: which data should we use? *BMC Health Serv Res*. 12:365. DOI: 10.1186/1472-6963-12-365

Osburn, L & McLeod, HD (2003). Risk equalisation methodologies in a South African context. Centre for Actuarial Research, University of Cape Town, Discussion Paper for the Risk Equalisation Fund Task Group. April 2003. [unpublished]

Pheiffer, C et al. (2022). Prevalence of Type 2 diabetes in South Africa: A systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* 2021, 18, 5868. https://doi.org/10.3390/ijerph18115868

Riddle, MC et al. (2021) Consensus report: Definition and interpretation of remission in type 2 diabetes. diabetes care. Published online August 30, 2021. DOI:10.2337/dci21-0034

Ruffieux, Y et al. (2023) Life years lost associated with mental illness: A cohort study of beneficiaries of a South African medical insurance scheme. *Journal of affective disorders* Vol. 340 (2023): 204–212. https://doi.org/10.1016/j.jad.2023.08.013

Seedat, YK, et al. (2014). Hypertension guideline working group, South African hypertension practice guideline 2014. *Cardiovasc J Afr.* **25**(6):288–94 DOI: 10.5830/CVJA-2014-062

Sharp, PM, et al. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med.* **1**(1):a006841. DOI: 10.1101/cshperspect.a006841

Souri, S, et al. (2017) Identification of validated case definitions for chronic disease using electronic medical records: a systematic review protocol. *Systematic Reviews* **6**(38). DOI: 10.1186/s13643-017-0431-9

St. Clair, P, et al. (2017) Using self-reports or claims to assess disease prevalence: It's complicated. *Med Care*. **55**(8):782–788. DOI:10.1097/MLR.0000000000000753

Tomlinson, M, et al. (2009). The epidemiology of major depression in South Africa: results from the South African stress and health study. *S Afr Med J*. **99**(5 Pt 2):367–73. PMC3195337

Tonelli, M, et al. (2015) Methods for identifying 30 chronic conditions: application to administrative data [published correction appears in *BMC Med Inform Decis Mak*. 2019 Sep 4; **19**(1):177. DOI: 10.1186/s12911-019-0900-2]. *BMC Med Inform Decis Mak*. 15:31. DOI: 10.1186/s12911-015-0155-5

Williams, B, et al. (2018) ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. *J Hypertens*. **36**(10):1953–2041. DOI: 10.1093/eurheartj/ehy339