A Comprehensive Introduction to Disability-Adjusted Life Years (DALY)

And

DisMod & CODEm

Rajesh Majumder
St. John’s Research Institute, Bengaluru, India

9th April 2026

A. What is GBD ?

GBD (Global Burden of Disease Study) is a large international research program that quantifies health loss due to diseases, injuries, and risk factors across countries, time, age, and sex. It is coordinated mainly by the Institute for Health Metrics and Evaluation (IHME) at the University of Washington and involves thousands of researchers worldwide.

Main Objectives

GBD aims to:

- Measure mortality and morbidity globally.
- Compare health loss across diseases, countries, and years.
- Identify major risk factors contributing to disease.
- Help policy makers prioritize health interventions.

Example

- Compare burden of **stroke vs diabetes vs road injuries** in India.
- Estimate how much **AP contributes to CVD**

Main Measures Used in GBD

Core Measures are:

- Deaths
- Incidence
- Prevalence
- Years of Life Lost (YLL)
- Years Lived with Disability (YLD)
- Disability-Adjusted Life Years (DALY)
- Healthy Life Expectancy (HALE)

DALY (Disability-Adjusted Life Year)

A **DALY represents one lost year of healthy life**.

$$DALY = YLL + YLD$$

Where,

- YLL= Years of Life Lost due to premature death
- YLD= Years Lived with Disability

Concept behind DALY

- If any health outcome has an impact on social welfare, it should be included in the burden of disease index.
- Irrespective of social, educational, economical, religious diversity; consider only country, age and sex
- Treating similar health outcomes equally (“**like should be treated as like**”)
- Use time as a unit.

Years of Life Lost (YLL)

- It’s a Measures premature mortality. Also called **Standard expected YLL**

$$YLL = \sum_{x=0}^{l} d_x L_x^*$$

- $x =$ indicator of different age groups $\forall x = 0,1,2,\ldots,l(last\ age; in\ years)$
- $d_x =$ no. of deaths at age x, due to a specific disease.
- $L_x^* =$ Expected life at each age based on some ideal standard. (Japanese females age 82.5; life-table West Level 26)
- Example: From any population let’s say at age group 50 years there was 100 deaths in a year. The standard life expectancy for any 50 years old healthy people would be 30 years more to live. So the YLL= 100×30=3000, meaning 3000 years of life lost due to premature death.

Difference between Male and Female

Question: Should the same life expectancy be used for men and women?

- From an equality perspective → should be the same
- But biologically women live longer

In reality:

- Difference ~2–3 years (in the future)
- So different values are taken in DALY

Social Value of Time Lived at Different Ages

- In all societies, people’s **social roles change with age**. Children and, in many cases, older adults are:
  - Physically dependent
  - Mentally dependent
  - Financially dependent
- Because of these changing roles and levels of dependency across age groups, it may be reasonable to **assign different values to time lived at different ages (unequal valuation)**.

- Two approaches are shown:
    - **Uniform value** → all ages are valued equally
    - **Unequal age weights** → more weight given to middle age

How Were Age Weights Determined

1) Human Capital Theory
    - Based on productivity (output). Uses wages, income, etc.
    - Leads to inequality
    - Not used in DALY
2) Social Role Perspective (Accepted)
    - Focuses on social roles at different ages
    - Based on ideas by Daniels
    - Fair over lifetime
    - Reflects social roles

Since data was limited:

Modified Delphi Method was used
(expert consensus approach)

Two options:

- Discrete weights for each age
- Continuous mathematical function

Continuous function preferred (simpler for calculation)

Age Weighting Function

The function used is:

$$W(x) = Cxe^{-\beta x}$$

Where,

- $\beta =$ constant controlling curve shape

- $x =$ age in years
- $C =$ constant, which ensures that, total disease burden remains unchanged (with or without age weighting)

Typical values:

$\beta = 0.03$ to 0.05

Used: $\beta = 0.04$ (in GBD study) and $C = 0.16243$

Non-Fatal Health Outcomes

- Public health mainly focused on **death (mortality)**.

But this misses a huge part of reality:

- People live with diseases
- They suffer disability
- They lose quality of life

So the key question is:

"How do we measure the burden of diseases that don't kill but still harm?"

The Core Challenge

We needed a way to make **non-fatal outcomes comparable to death**.

In simple terms:

Death = total loss of life

Disability = partial loss of health

So we needed a **common unit** to compare both.

(A) **Health Economics → QALY**

Focus: individual preferences

Idea: How much "quality" is lost?

Methods: trade-offs (time, risk, etc.)

Problem: Didn't fully capture social & functional aspects

(B) **Health Status Research**

Focus: measuring health dimensions

Included: Physical, Mental, Social-function

Problem: Weights were often arbitrary

(C) **WHO Framework (ICIDH)**

Focus: conceptual clarity

Chain: Disease → Impairment → Disability → Handicap

Key insight: Disease affects function, then affects society

Problem: “Handicap” depends on environment → can create inequality

(D) **The Key Decision (DALY Framework)**

To build a fair and universal measure, DALY use Disability, NOT Handicap.

Because: Handicap depends on social context; Could bias results (rich vs poor)

Goal: “Equal health loss should have equal value everywhere.”

Standardizing Disability

Challenge: How do we measure “how bad” a disability is?

Solution:

- Create 6 disability classes (mild → severe)
- Assign each a weight (0 to 1)

Interpretation:

- 0 = perfect health
- 1 = equivalent to death

Separating Two Key Ideas

- **Duration vs Severity**
    - Severity → how bad the condition is
    - Duration → how long it lasts
- These are calculated separately
- Example: Many mild cases = one severe case
- Total burden can be equal

Linking Disease to Disability

Another challenge: A disease does not directly give DALY.

- So we need: Disease → Impairment → Disability mapping
- This was developed in Global Burden of Disease (GBD)

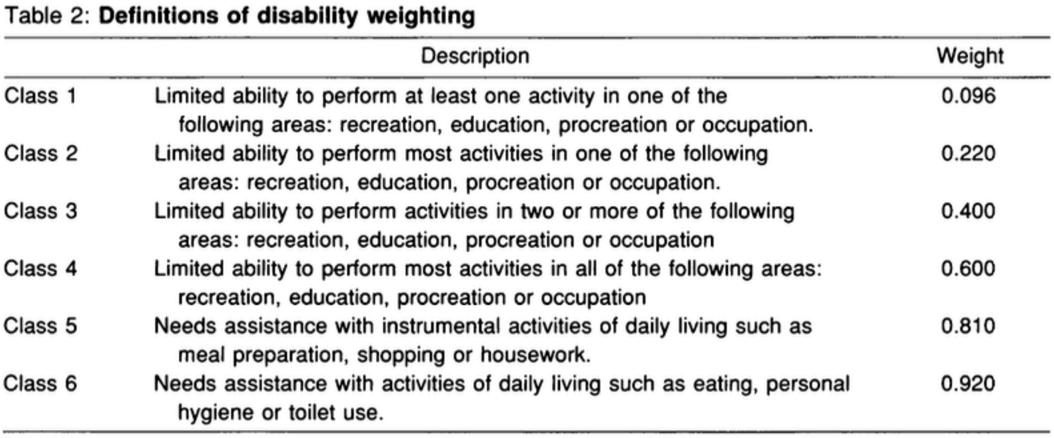


Table 2: **Definitions of disability weighting**

	Description	Weight
Class 1	Limited ability to perform at least one activity in one of the following areas: recreation, education, procreation or occupation.	0.096
Class 2	Limited ability to perform most activities in one of the following areas: recreation, education, procreation or occupation.	0.220
Class 3	Limited ability to perform activities in two or more of the following areas: recreation, education, procreation or occupation	0.400
Class 4	Limited ability to perform most activities in all of the following areas: recreation, education, procreation or occupation	0.600
Class 5	Needs assistance with instrumental activities of daily living such as meal preparation, shopping or housework.	0.810
Class 6	Needs assistance with activities of daily living such as eating, personal hygiene or toilet use.	0.920

WHO Bulletin OMS. Vol 72 1994

Time Preference in DALY

- At its simplest, time preference is an economic concept:

People value present benefits more than future benefits

- That means:

Something received today is worth more than the same thing received in the future

- Example:
  - ₹100 today vs ₹100 after 1 year → Most people choose ₹100 today
  - ₹100 today vs ₹110 after 1 year → Some people may choose ₹110 later

This shows: People are willing to wait only if compensated

What is Discounting

- In economic evaluation, We use a discount rate to convert:

Future benefits → Present value

- This helps:
  - Compare costs and benefits on the same scale
  - Perform cost-effectiveness analysis

Big Debate: What Should Be the Discount Rate?

- There is no universal agreement on:
  - Theoretical basis
  - Exact rate

Ideas of different Discounting Rates

Social Opportunity Cost of Capital
Uses market return on investment
Typical: 8%-15%
Example: World Bank, US Congress: ~10%
Problem: Market distortions (tax, policy, etc.)

Social Time Preference
Based on society’s preference for present vs future
Typically: Lower rate (1-3%)

Myopia (Short-term thinking)
People naturally prefer present
Problem:

- Not a strong argument socially
- Future generations are equally important

Growth in Consumption

- Future consumption ↑ → marginal utility ↓
- So future benefits are valued less

Problem:

- DALY measures time, not consumption
- Future life years may be equally or more valuable

Time Paradox

If we do NOT discount:

- Benefits remain same over time
- Costs decrease (due to investment returns)

Result: Always better to delay projects

Example: ₹1000 today → ₹1100 next year ⟹Every project will be postponed

To avoid this: We must discount both costs and benefits.

Conditions for Time Paradox

Occurs when:

- Future opportunities remain same
- Budget increases over time
- Returns in other sectors are higher

If no discounting:

A disease eradication program:

- One-time cost
- Infinite future benefits

Leads to: All resources should go to eradication (unrealistic)

Discounting Formula

- The fundamental discounting expression is:

$$e^{-r(x-a)}$$

- Where:
  - r= discount rate
  - a= age at death
  - x= future age ($x > a$)

This tells us: **The value of one year lived at age x, discounted back to age a**

Note that: The discount rate reflects societal time preference, not a biological parameter. So it is treated as a policy choice, not something inferred from epidemiological data.

Discounted YLL

If a person dies at age a, and would have lived up to $a + L$:

- They lose **L years of life**
- But each future year must be **discounted**

So YLL becomes: Sum (continuous) of discounted years from:

$x = a$ to $x = a + L$

$$YLL = \int_a^{a+L} e^{-r(x-a)}\ dx$$

Let: $t = x - a$. Then limits change: When $x = a$, $t = 0$; When $x = a + L$, $t = L$

$$YLL = \int_0^L e^{-rt}\ dt$$

Apply limits (0 to L): $YLL = \left[\frac{-1}{r} e^{-rt}\right]_0^L = \frac{1-e^{rL}}{r}$

- Note that: If $r = 0$: $YLL =$ L
  Meaning: **No discounting → total life lost = full life expectancy**

DALY uses a **low positive discount rate: 3% (r = 0.03)**

Reason:

- Consistent with long-term investments

- Avoids time paradox
- Reduces eradication bias

Technical Implications of Discounting

(1) Incidence vs Prevalence

- With discounting:

Incidence-based DALY < Prevalence-based DALY

- Why?

Because, Future years are discounted → less weight

(2) Mortality vs Disability Comparison

Future deaths vs present deaths → different weights

- Discounting creates this difference

Solution:

- Disability → current incidence
- Mortality → current death rate

Complete Overview of DALY

STEP 0: Start from the Original Idea

- From the early World Health Organization / Global Burden of Disease Study:

$$DALY = YLL + YLD$$

But instead of simple multiplication, they used a **continuous-time formulation**.

STEP 1: Write the General Continuous Loss Function

- Total healthy life lost is defined as:

$$DALY = \int_{a}^{a+L} Loss\,(x)\;dx$$

- Where:

a= age at onset (or death)

L= duration (or remaining life expectancy)

STEP 2: Define the "Loss Function"

Loss at time x is composed of three elements:

1. **Disability weight (severity)** → D
2. **Age-weighting** → $Cxe^{-\beta x}$
3. **Discounting** → $e^{-r(x-a)}$

So:

$$\text{Loss}(x) = D \cdot C \cdot x \cdot e^{-\beta x} \cdot e^{-r(x-a)}$$

STEP 3: Plug into the Integral

$$DALY = \int_a^{a+L} D \cdot C \cdot x \cdot e^{-\beta x} \cdot e^{-r(x-a)} dx$$

- This is the **full original DALY formula**.

STEP 4: Combine Exponentials

Exponents

$$e^{-\beta x} \cdot e^{-r(x-a)} = e^{-(\beta+r)x} \cdot e^{ra}$$

So:

$$DALY = DCe^{ra} \int_a^{a+L} x\, e^{-(\beta+r)x} dx$$

STEP 5: Solve the Integral

We solve:

$$\int x\, e^{-kx} dx$$

(using integration by parts)

$$= \frac{-(xe^{-kx})}{k} - \frac{e^{-kx}}{k^2}$$

Applying limits → final closed form:

$$DALY = \frac{DCe^{-(\beta+r)a}}{(\beta+r)^2}[e^{-(\beta+r)L}\big(1+(\beta+r)(L+a)\big)-(1+(\beta+r)a)]$$

This is the **classic WHO DALY equation (1994)**

STEP 6: Separate YLL and YLD (Interpretation Step)

- Same formula, different inputs:

YLL (death)

- $D = 1$
- $L =$remaining life expectancy

YLD (disability)

- $D < 1$
- $L =$duration

STEP 7: First Simplification — Remove Age Weighting

Criticism: Age-weighting implies some ages are more "valuable"

So remove:

$$Cxe^{-\beta x} \rightarrow 1$$

Now formula becomes:

$$DALY = \int_a^{a+L} D \cdot e^{-r(x-a)} dx$$

STEP 8: Solve Without Age Weighting

$$DALY = D\int_a^{a+L} e^{-r(x-a)}\,dx$$

Let $t = x - a$:

$$DALY = D\int_0^L e^{-rt}\,dt$$

$$= D \cdot \frac{1-e^{-rL}}{r}$$

STEP 9: Remove Discounting (Modern Shift)

Further criticism: Discounting future life is ethically controversial

So set:

$$r = 0$$

Then:

$$\lim_{r \to 0} \frac{1 - e^{-rL}}{r} = L$$

STEP 10: Final Modern DALY Formula

- From newer Institute for Health Metrics and Evaluation / Global Burden of Disease Study:

$$DALY = D \times L$$

For YLL:

$$YLL = N \times L$$

N= number of deaths

L= standard life expectancy at age of death

For YLD:

$$YLD = I \times D \times L$$

(or prevalence-based version)

B. DisMod II

DisMod II mathematically links incidence, prevalence, mortality, and duration using differential equations, survival functions, and optimization to produce internally consistent epidemiological estimates.

Conceptual Framework (Mathematical Formulation)

Describes:

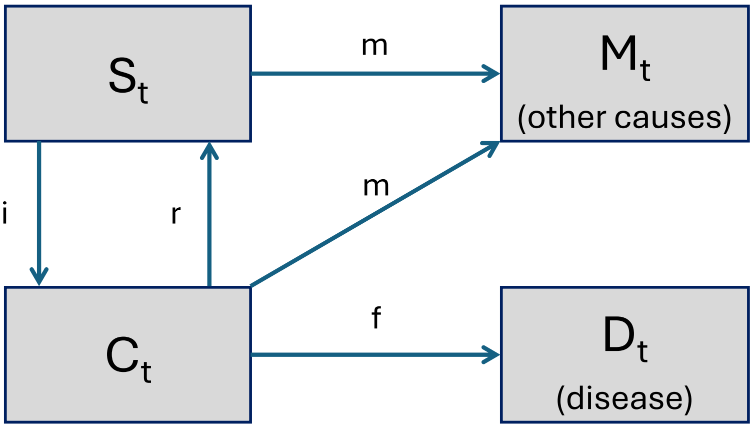
- A specific disease

- Mortality from other cause

Population divided into states:

- Healthy
- Diseased
- Dead

The Conceptual disease model



S= number of healthy people
C= number of diseased people
D= number of people dead from the disease
M= number of people dead from all other causes
t= Age (years)
Three hazards: i= incidence ; r= remission; f= case fatality
m= all other mortality

DisMod II is a **multi-state compartment model** with 3 states:

- $S(a)$: Healthy at age a
- $C(a)$: Diseased at age a
- $D(a)$: Dead due to disease at age a

Transition Hazards

- $i(a)$: Incidence rate (new cases entering disease state)
- $r(a)$: Remission rate (recovery from disease)
- $f(a)$: Case fatality rate (Death rate due to cause specific disease)
- $m(a)$: Mortality from other causes

(Initially ignoring other-cause mortality)

Core Differential Equations

The disease dynamics are governed by:

(1) Healthy population

$$\frac{dS(a)}{da} = -i(a)S(a) + r(a)C(a)$$

Loss due to incidence, gain from remission

(2) Diseased population

$$\frac{dC(a)}{da} = i(a)S(a) - (r(a) + f(a))C(a)$$

Gain from incidence, loss due to remission + death

(3) Deaths due to disease

$$\frac{dD(a)}{da} = f(a)C(a)$$

Total Population Constraint

$$N(a) = S(a) + C(a) + D(a)$$

Step 1: Start from the ODE System

We already have:

$$\frac{dS}{da} = -iS + rC$$
$$\frac{dC}{da} = iS - (r + f)C$$

Write in **matrix form**:

$$\frac{d}{da}\begin{bmatrix} S \\ C \end{bmatrix} = \begin{bmatrix} -i & r \\ i & -(r+f) \end{bmatrix}\begin{bmatrix} S \\ C \end{bmatrix}$$

Step 2: Define Matrix A

Let:

$$A = \begin{bmatrix} -i & r \\ i & -(r+f) \end{bmatrix}$$

Then:

$$\frac{dX}{da} = AX \text{where } X = \begin{bmatrix} S \\ C \end{bmatrix}$$

Step 3: Solve Using Matrix Exponential

Solution:

$$X(a+1) = e^A X(a)$$

So the problem reduces to computing:

$$e^A$$

Note that for simple differential equation:

$$\frac{dx}{dt} = kx \ (where\ k\ is\ a\ constant)$$

The solution for this would be:

$$x = Ce^{kt} \ \ (where\ C\ is\ the\ integral\ constant)$$

So, think is this manner:

$$\frac{dX}{da} = AX(a)$$

$$\mathrm{X(a) = X(0)e^{Aa}}$$

That means,

e^A= is a transition matrix

$X(a)$= Given current state

$X(a+1)$= What will be the next state

Step 4: Eigenvalues of A

Solve:

$$\mid A - \lambda I \mid = 0$$

$$\begin{vmatrix} -i-\lambda & r \\ i & -(r+f)-\lambda \end{vmatrix} = 0$$

Expand:

$$(-i-\lambda)(-(r+f)-\lambda)-ir=0$$

Simplify:

$$\lambda^2+(i+r+f)\lambda+ir=0$$

Define:

$$\lambda=\frac{-(i+r+f)\pm q}{2}$$

where:

$$q=\sqrt{(i+r+f)^2-4ir}$$

Step 5: Define Key Quantities (DisMod notation)

Let:

$$l=i+r+f$$

$$q=\sqrt{l^2-4ir}$$

Define:

$$\lambda=\frac{-l\pm q}{2}$$

So eigenvalues are:

$$\lambda_1=-\frac{l+q}{2},\lambda_2=-\frac{l-q}{2}$$

Matrix exponential using eigenvalues:

General rule:

If eigenvalues are λ_1,λ_2, then:

$$e^A=P\begin{bmatrix}e^{\lambda_1} & 0\\ 0 & e^{\lambda_2}\end{bmatrix}P^{-1}$$

Substitute eigenvalues:

$$e^{\lambda_1}=e^{-\frac{l+q}{2}}$$

$$e^{\lambda_2} = e^{-\frac{l-q}{2}}$$

To simplify writing:

$$v = e^{-\frac{l+q}{2}}, w = e^{-\frac{l-q}{2}}$$

These are eigenvalue-based exponentials

Step 6: Matrix Exponential Result

After diagonalization:

$$e^{A} = \frac{1}{q}\begin{bmatrix} (l+q)/2 - i & r \\ i & (l+q)/2 - (r+f) \end{bmatrix} v + \frac{1}{q}\begin{bmatrix} i - (l-q)/2 & -r \\ -i & (r+f) - (l-q)/2 \end{bmatrix} w$$

Step 7: Apply to Recursive Form

$$\begin{bmatrix} S(a+1) \\ C(a+1) \end{bmatrix} = e^{A} \begin{bmatrix} S(a) \\ C(a) \end{bmatrix}$$

Step 8: Final Recursive Equations

After simplification (this is the DisMod II form):

Healthy

$$S(a+1) = \frac{1}{2q}[((l+q) - 2i)vS(a) + ((l-q) - 2i)wS(a) + 2r(v-w)C(a)]$$

Diseased

$$C(a+1) = \frac{1}{2q}[2i(v-w)S(a) + ((l+q) - 2(r+f))vC(a) + ((l-q) - 2(r+f))wC(a)]$$

Step 9: Death Equation

$$\frac{dD}{da} = fC(a)$$

So:

$$D(a+1) = D(a) + \int_{a}^{a+1} fC(a)\,da$$

Substitute $C(a)$solution:

$$C(a) = Ae^{-\frac{l+q}{2}a} + Be^{-\frac{l-q}{2}a}$$

Integrate:

$$D(a+1) - D(a) = f\int_{a}^{a+1} C(a)\,da$$

$$= f\left[\frac{A}{-\frac{l+q}{2}}\left(e^{-\frac{l+q}{2}(a+1)} - e^{-\frac{l+q}{2}a}\right) + \frac{B}{-\frac{l-q}{2}}\left(e^{-\frac{l-q}{2}(a+1)} - e^{-\frac{l-q}{2}a}\right)\right]$$

Step 10: Compact DisMod Form

Using v, w:

$$D(a+1) = D(a) + \frac{f}{q}[(v-w)(iS(a) + (r+f)C(a)) + (vw)(C(a)\cdot l)]$$

Person-Years at Risk (PY)

Step 1: Definition

Person-years represent the **total time lived by individuals in a given age interval**.

We consider interval:

$$[a, a+1)$$

Step 2: Total population at time a

$$N(a) = S(a) + C(a)$$

(ignoring deaths within instant)

Step 3: Population changes over interval

At:

- Start: $N(a)$
- End: $N(a+1) = S(a+1) + C(a+1)$

Step 4: Approximate average population

Using trapezoidal rule:

$$\text{Average population} \approx \frac{N(a) + N(a+1)}{2}$$

Step 5: Person-years

$$PY(a) = \frac{N(a) + N(a+1)}{2}$$

Expand:

$$PY(a) = S(a) + C(a) + \frac{1}{2}[S(a+1) + C(a+1) - S(a) - C(a)]$$

Simplified (DisMod form):

$$PY(a) = S(a) + C(a) + 0.5\ [S(a+1) + C(a+1)]$$

Adjusts population using start and end values for better accuracy

Prevalence

Step 1: Definition

Prevalence = proportion of population that is diseased

Step 2: Use mid-interval population

$$\text{Average diseased} = \frac{C(a) + C(a+1)}{2}$$

Step 3: Divide by person-years

$$c(a) = \frac{0.5\ [C(a) + C(a+1)]}{PY(a)}$$

Adjusts population using start and end values for better accuracy

Interpretation

Uses **mid-interval approximation**
Ensures consistency with dynamic population

The equations of DisMod II are built on population-level compartment equations, not directly on prevalence. But you can also derive the exact prevalence formula by converting the system into prevalence form.

Let's define prevalence:

$$P(a) = \frac{C(a)}{N(a)}$$

Where,

$$N(a) = S(a) + C(a)$$

So, differentiate prevalence

$$\frac{dP}{da} = \frac{1}{N}\frac{dC}{da} - \frac{C}{N^2}\frac{dN}{da}$$

Now, substitute from DisMod equations:

$$\frac{dC}{da} = iS - (r + f)C$$

$\frac{dN}{da} = -fC$ (ignoring other-cause mortality)

Replace $S = N - C$

$$\frac{dC}{da} = i(N - C) - (r + f)C$$
$$= iN - iC - rC - fC$$

Plug into prevalence formula:

$$\frac{dP}{da} = \left\{\frac{1}{N}(iN - iC - rC - fC)\right\} - \frac{C}{N^2}(-fC)$$

$$= i\left(1 - \frac{C}{N}\right) - (r + f)\frac{C}{N} + f\left(\frac{C}{N}\right)^2$$

$$= i(1 - P) - (r + f)P + fP^2$$

So the final equation is:

$$\frac{dP}{da} = i(a)\big(i - P(a)\big) - \big(r(a) + f(a)\big)P(a) + f(a)P^2(a)$$

Simplified form:

Now note that, when P is small → $P^2 \approx 0$ → $fP^2 \approx 0$
And, when P is small → $(1 - P) \approx 1$

So, the final model would be

$$\frac{dP}{da} = i - (r + f)P$$

And for Steady state ($\frac{dp}{da} = 0$), we can simply write:

$$p = \frac{i}{r + f} \Longrightarrow prevalence \approx incidence \times duration$$

Note that the actual DisMod formula for Prevalence is not differ from the above derived prevalence formula.

The true (instantaneous) prevalence at age a:

$$P(a) = \frac{C(a)}{N(a)}$$ (This is a mathematical definition.)

And the DisMod formula is :

$$c(a) = \frac{0.5[C(a) + C(a+1)]}{PY(a)}$$

It is estimating prevalence over an interval $[a, a+1]$. This formula is used in DisMod because population is changing with age (dynamic system). So, we don't observe C(a) at a single instant, we observe counts across an interval. The core idea behind the prevalence formula used in DisMod is the mid-interval approximation:

$$Average\ diseased \approx \frac{C(a) + C(a+1)}{2}$$

This assumes smooth change between ages. So this connects with the true prevalence:

$$P(a) = \frac{C(a)}{N(a)} \approx \frac{Average\ diseased}{Average\ population}$$

Where C(a) → average diseased
$N(a)$ → person years

So the immediate next question would be why use person-years? Because in dynamic population people enter/exit (death, aging, etc). That means population is not constant. So,

$$PY(a) \approx averagepopulation\ in\ interval$$

So that means,

The DisMod formula :

$$c(a) = \frac{0.5[C(a) + C(a+1)]}{PY(a)}$$

Is equivalent to :

$$P(a) = \frac{C(a)}{N(a)} \approx \frac{Average\ diseased}{Average\ population}$$

Mortality Rate

Step 1: Definition

Mortality rate = deaths per person-year

Step 2: Total deaths in interval

$$\text{Deaths} = D(a+1) - D(a)$$

Step 3: Rate

$$b(a) = \frac{D(a+1) - D(a)}{PY(a)}$$

Step 4: Substitute differential equation

From:

$$\frac{dD}{da} = fC(a)$$

$$D(a+1) - D(a) = \int_a^{a+1} fC(a)\, da$$

Total deaths equal integral of disease fatality over interval

Final:

$$b(a) = \frac{\int_a^{a+1} fC(a)\, da}{PY(a)}$$

Survival in Diseased State

Step 1: Define exit hazards

$$\beta(a) = r(a) + f(a) + m(a)$$

Total rate of leaving diseased state

Step 2: Survival differential equation

Let,

d: Duration since disease onset

$$\frac{dy(d)}{dd} = -\beta y(d)$$

Rate of decline of diseased survival over duration

Step 3: Solve

$$\frac{dy}{y} = -\beta dd$$

Integrate:

$$\ln\ y = -\beta d$$

$$y(d) = e^{-\beta d}$$

Probability of remaining diseased state after duration d

Disease Duration (Second Stage)

Step 1: Expected duration

$$X(a) = \int_0^\infty y(d)\, dd$$

Expected time spent in diseased state

Step 2: Substitute survival

$$X(a) = \int_0^\infty e^{-\beta d}\, dd$$

Duration expressed using survival function

Step 3: Solve

$$X(a) = \frac{1}{\beta(a)}$$

Average duration equals inverse of exit rate

Step 4: Discrete formulation (DisMod)

$$X(a) = \sum_{k=1}^{\infty} x_{a,k}$$

Total duration as sum of yearly contributions

where:

$$x_{a,k} = \int_{k-1}^{k} e^{-\beta d}\, dd$$

Contribution of each duration interval

Optimization Framework (When Inputs Missing)

Step 1: Problem

We do NOT know:

- $i(a), r(a), f(a)$

But we may know:

- Prevalence
- Mortality

Step 2: Model prediction

Using equations:

$$\hat{c}(a), \hat{b}(a)$$

Contribution of each duration interval

Step 3: Define loss function

$$L = \sum_a w_1\,(c(a) - \hat{c}(a))^2 + w_2(b(a) - \hat{b}(a))^2$$

Measures mismatch between observed and model outputs

Step 4: Optimization

Find:

$$i(a), r(a), f(a)$$

such that:

$$L \to \min$$

Measures mismatch between observed and model outputs

Step 5: Algorithm

Uses **Downhill Simplex Method**

- Start with initial guess
- Update parameters
- Recompute outputs
- Repeat until convergence

Special Case: Chronic Disease

Step 1: Assumption

$$r(a) \approx 0$$

(no remission)

Step 2: Simplified ODE

$$\frac{dC}{da} = iS - fC$$

Change in diseased population = new cases − deaths

Step 3: Substitute $S = N - C$

$$\frac{dC}{da} = i(N - C) - fC$$
$$= iN - (i + f)C$$

Replace healthy population with total minus diseased

Step 4: Solve

$$C(a) = \frac{iN}{i + f} + Ae^{-(i+f)a}$$

Solution showing equilibrium + transient behavior

Step 5: Duration

$$\beta = f + m$$

Exit only via death (no remission)

$$X(a) = \frac{1}{f + m}$$

Duration depends only on mortality rates

Interpretation

- No recovery
- Exit only via death

C. CODEm

CODEm models cause-specific mortality by fitting multiple statistical models to log mortality rates, evaluating their predictive performance, and combining them into an optimized ensemble to produce robust mortality estimates.

Objective of CODEm

We want to estimate:

$$D_{a,s,t,l} = \text{number of deaths}$$

or equivalently:

$$m_{a,s,t,l} = \text{cause-specific mortality rate}$$

where:

- a: age
- s: sex
- t: time
- l: location

Basic Statistical Problem

We observe:

- Noisy, incomplete mortality data
- Different sources (VR, surveys, verbal autopsy)

Goal:

$\hat{m}(a, s, t, l) = f(\text{covariates,space,time})$

Transform the Outcome

Mortality rates are positive ⇒ take log:

$$y = \log{(m)}$$

This stabilizes variance and allows linear modeling

Candidate Model Classes in CODEm

CODEm does **not use one model**—it uses many.

1 Linear Mixed Effects Model

$$y_i = X_i\beta + Z_i u + \epsilon_i$$

Where:

- X_i: covariates (income, education, etc.)
- β: fixed effects
- $Z_i u$: random effects (location, time)
- $\epsilon_i \sim N(0, \sigma^2)$

2 Spatio-temporal Model

$$y_{l,t} = \mu + \alpha_l + \gamma_t + \delta_{l,t}$$

Where:

- α_l: spatial effect
- γ_t: temporal trend
- $\delta_{l,t}$: interaction

3 Gaussian Process Regression (GPR)

Smooths over time:

$$y(t) \sim GP(\mu(t), K(t, t'))$$

Where:

- K: covariance function

4 Other Models

- Age-period-cohort models
- Nonlinear regressions
- Penalized splines

Model Training

For each candidate model k:

$$\hat{y}_k = f_k(X)$$

Out-of-Sample Validation

Split data:

- Train set
- Test set

Compute error:

$$E_k = \sum(y_{\text{obs}} - \hat{y}_k)^2$$

Ensemble Step (Core of CODEm)

Instead of choosing one model:

Combine all models

Weighted Ensemble

$$\hat{y} = \sum_{k=1}^{K} w_k \, \hat{y}_k$$

Where:

- $w_k \geq 0$
- $\sum w_k = 1$

Weight Selection

Weights chosen to minimize prediction error:

$$\min_w \sum(y_{\text{obs}} - \sum w_k \hat{y}_k)^2$$

Back Transformation

$$\hat{m} = e^{\hat{y}}$$

Scaling to Total Mortality

Constraint:

$$\sum_{\text{causes}} m_c = m_{\text{all-cause}}$$

So:

$$m_c^* = \frac{m_c}{\sum_c m_c} \times m_{\text{total}}$$

Death Counts

$$D(a, s, t, l) = m(a, s, t, l) \times Population(a, s, t, l)$$

Uncertainty (Very Important)

CODEm generates multiple draws:

$$\hat{y}^{(d)} \sim \text{distribution}$$

Then:

$$\widehat{m}^{(d)} = e^{\hat{y}^{(d)}}$$

Connection to YLL

$$YLL(a) = D(a) \times L(a)$$

Full Mathematical Pipeline

Step 1: Model candidates

$$\hat{y}_k = f_k(X)$$

Step 2: Ensemble

$$\hat{y} = \sum w_k \hat{y}_k$$

Step 3: Transform

$$\widehat{m} = e^{\hat{y}}$$

Step 4: Scale

$$m_c^* = \frac{m_c}{\sum m_c} \cdot m_{\text{total}}$$

Step 5: Deaths

$$D = m \cdot Population$$

Step 6: YLL

$$YLL = D \cdot L$$

Key Difference from DisMod

Feature	DisMod	CODEm
Type	Mechanistic	Statistical
Core math	ODE system	Regression + ensemble
Output	Incidence, duration	Mortality
Goal	Internal consistency	Best prediction

Intuition

CODEm is like:

- Try many models
- Keep the best parts of each
- Combine them optimally

Final Mathematical Insight

$$\boxed{\hat{m} = \exp\left(\sum_k w_k\, f_k(X)\right)}$$

D. DisMod-MR

What is DisMod-MR?

DisMod-MR combines a mechanistic disease model (ODE system) with a Bayesian hierarchical likelihood framework, where observed epidemiological data are linked to latent parameters through likelihood functions and priors, and inference is obtained via posterior maximization or sampling.

$$\boxed{\text{DisMod-MR} = \text{Bayesian meta-regression model for disease epidemiology}}$$

It extends DisMod II by:

- Handling **multiple data sources**
- Allowing **uncertainty**
- Using **Bayesian inference**
- Producing **age–time–location estimates**

Core Idea

Same biological logic as DisMod II:

$$\text{Incidence} \leftrightarrow \text{Prevalence} \leftrightarrow \text{Mortality}$$

But instead of solving deterministically:

$$\boxed{\text{Estimate parameters probabilistically using data + priors}}$$

Goal of DisMod-MR

We want to estimate:

$$\theta = \{i(a,l,t),\ r(a,l,t),\ f(a,l,t)\}$$

given **noisy, incomplete observations** of:

- prevalence c^{obs}
- incidence i^{obs}
- mortality b^{obs}

Latent Disease Process (Mechanistic Layer)

This is the **deterministic backbone (same as DisMod II)**:

$$\frac{dS}{da} = -iS + rC$$
$$\frac{dC}{da} = iS - (r + f)C$$

Define latent state

$$X(a, l, t) = \begin{bmatrix} S(a, l, t) \\ C(a, l, t) \end{bmatrix}$$
$$\frac{dX}{da} = A(a, l, t)X(a, l, t)$$

Solve ODE

$$X(a) = \Phi(a; \theta)$$

So all epidemiological quantities are **functions of parameters**:

$$c(a, l, t) = \frac{C(a, l, t)}{N(a, l, t)}$$
$$b(a, l, t) = f(a, l, t) \cdot c(a, l, t)$$

Parameter Model (Statistical Layer)

We do NOT estimate i, r, f directly.

Instead:

Log-linear model

$$\log\, i(a, l, t) = \eta_i(a, l, t)$$
$$\log\, r(a, l, t) = \eta_r(a, l, t)$$
$$\log\, f(a, l, t) = \eta_f(a, l, t)$$

Linear predictor

$$\eta_i = X_i\beta_i + Z_i u_i + \epsilon_i$$

Similarly:

$$\eta_r = X_r\beta_r + Z_r u_r + \epsilon_r$$
$$\eta_f = X_f\beta_f + Z_f u_f + \epsilon_f$$

Components

- $X\beta$: fixed effects (covariates)
- Zu: random effects (location/time hierarchy)
- ϵ: noise

Observation Model (Likelihood)

We now link **observed data** to **model outputs**

1 Prevalence data

Let:

$$c_j^{obs} \text{ at } (a_j, l_j, t_j)$$

Assume:

$$c_j^{obs} \sim \mathcal{N}(c(a_j, l_j, t_j; \theta), \sigma_j^2)$$

2 Incidence data

$$i_j^{obs} \sim \mathcal{N}(i(a_j, l_j, t_j; \theta), \sigma_j^2)$$

3 Mortality data

$$b_j^{obs} \sim \mathcal{N}(b(a_j, l_j, t_j; \theta), \sigma_j^2)$$

Combine likelihoods

$$\mathcal{L}(\theta) = \prod_j p(\, c_j^{obs} \mid \theta) \cdot \prod_k p(\, i_k^{obs} \mid \theta) \cdot \prod_m p(\, b_m^{obs} \mid \theta)$$

Log-likelihood

$$\log \mathcal{L} = -\frac{1}{2}\sum_j \frac{(c_j^{obs} - c_j(\theta))^2}{\sigma_j^2} - \frac{1}{2}\sum_k \frac{(i_k^{obs} - i_k(\theta))^2}{\sigma_k^2} - \frac{1}{2}\sum_m \frac{(b_m^{obs} - b_m(\theta))^2}{\sigma_m^2}$$

Prior Distributions

Fixed effects

$$\beta \sim \mathcal{N}(0, \Sigma_\beta)$$

Random effects

$$u_l \sim \mathcal{N}(0, \sigma_l^2)$$
$$v_t \sim \mathcal{N}(0, \sigma_t^2)$$

Smoothness prior (age)

$$\eta(a) \sim \text{Gaussian Process}$$

or:

$$\frac{d^2\eta}{da^2} \sim \mathcal{N}(0, \tau^2)$$

Penalizes rough curves

Full Bayesian Model

$$\boxed{p(\theta \mid \text{data}) \propto p(\text{data} \mid \theta) \cdot p(\theta)}$$

Expanded posterior

$$p(\beta, u, \theta \mid \text{data}) \propto \mathcal{L}(\theta) \cdot p(\beta) \cdot p(u) \cdot p(\text{smoothness})$$

Hierarchical Structure

$$\theta_{global} \to \theta_{region} \to \theta_{country} \to \theta_{state}$$

Example:

$$u_{country} \sim \mathcal{N}(u_{region}, \sigma^2)$$

Borrow strength across locations

Computation

We need to evaluate:

$$c(\theta), b(\theta)$$

which requires:

Solving ODE:

$$X(a+1) = e^{A} X(a)$$

Then compute likelihood repeatedly

Inference

Option 1: MCMC

Sample:

$$\theta^{(1)}, \theta^{(2)}, \ldots$$

Option 2: Optimization (MAP)

$$\hat{\theta} = \arg\max \log p(\theta \mid \text{data})$$

Final Outputs

For each a, l, t:

- $i(a,l,t)$
- $r(a,l,t)$
- $f(a,l,t)$
- $c(a,l,t)$
- uncertainty intervals

Key Insight (VERY IMPORTANT)

DisMod II:

$$\min \sum (data - model)^2$$

DisMod-MR:

$$\boxed{\max[\log \mathcal{L} + \log \text{Prior}]}$$

Compact Final Form

$$\boxed{\log\, p(\theta \mid data) = -\frac{1}{2}\sum\frac{(data - model(\theta))^2}{\sigma^2} - \frac{1}{2}(\beta^T\Sigma^{-1}\beta) - \frac{1}{2}(u^T\Sigma^{-1}u)}$$

Final Interpretation

DisMod-MR is:

A Bayesian hierarchical model where epidemiological consistency is enforced via ODE constraints and parameters are inferred from noisy multi-source data.

Ultimate Insight

The **ODE system acts as a constraint inside the likelihood**

The model is essentially:

- Mechanistic model (ODE)
  -
  - Statistical model (Bayesian inference)

E. The **complete end-to-end flow of DALY calculation, Old GBD approach (GBD 1990 / early) & New GBD approach (GBD 2010+)** And explicitly showing where **DisMod II** and **CODEm** fit

Core Identity (Same for Old & New)

$$\boxed{DALY = YLL + YLD}$$

Part 1: Old GBD Approach (GBD 1990–2000)

Key features:

- Incidence-based YLD
- Age-weighting + discounting
- Heavy use of **DisMod II**
- No CODEm (simpler mortality models)

Step 1: Input Data

- Incidence (if available)
- Prevalence
- Mortality
- Population

Step 2: Use DisMod II

Purpose:

- Fill missing epidemiological parameters
- Ensure consistency

Solve:

$$i(a), r(a), f(a)$$

Generate:

$$C(a), S(a), D(a)$$

Step 3: Estimate Duration

$$X(a) = \frac{1}{r(a) + f(a) + m(a)}$$

Step 4: YLD (Incidence-Based)

$$YLD(a) = i(a) \cdot S(a) \cdot X(a) \cdot DW$$

Alternative: Derive C(a)

We can derive:

$$C(a) = i(a) \cdot S(a) \cdot X(a)$$

then,

YLD (Prevalence-Based)

$$YLD(a) = C(a) \cdot DW$$

But this was **not the primary method used**

Step 5: YLL

$$YLL(a) = D(a) \cdot L(a)$$

Step 6: Age Weighting + Discounting

Old formula:

$$YLL = \int e^{-rt} C e^{-\beta a} dt$$
$$YLD = \int DW \cdot e^{-rt} C e^{-\beta a} dt$$

Where:

- r= discount rate
- β= age-weighting

Step 7: Final DALY

$$DALY = YLL + YLD$$

Summary (OLD)

Flow:

```
Data → DisMod II → Incidence + Duration → YLD
     → Deaths → YLL
     → Age weighting + discounting
     → DALY
```

Part 2: New GBD Approach (GBD 2010+)

Major changes:

- Prevalence-based YLD
- No age-weighting
- No discounting
- Introduced **CODEm**
- Upgraded DisMod → DisMod-MR

Step 1: Input Data

- Mortality data
- Prevalence data

- Covariates
- Population

Step 2: Mortality Modeling → CODEm

Step 2.1:

$$y = \log(m)$$

Step 2.2:

$$\hat{y} = \sum w_k f_k(X)$$

Step 2.3:

$$\hat{m} = e^{\hat{y}}$$

Step 3: Deaths

$$D(a) = \hat{m}(a) \cdot N(a)$$

Step 4: YLL

$$YLL(a) = D(a) \cdot L(a)$$

Step 5: Epidemiology → DisMod (MR / II concept)

Input:

- Prevalence
- Mortality

Solve:

$$i(a), r(a), f(a)$$

Step 6: Prevalence

$$c(a) = \frac{C(a)}{N(a)}$$

Step 7: YLD (Prevalence-Based)

$$YLD(a) = c(a) \cdot N(a) \cdot DW$$

or: Cases

$$YLD(a) = C(a) \cdot DW$$

Step 8: Final DALY

$$DALY(a) = YLL(a) + YLD(a)$$

Summary (NEW)

Flow:

```
Data → CODEm → Mortality → YLL
     → DisMod → Prevalence → YLD
     → DALY
```

Side-By-Side Comparisons

Component	OLD GBD	NEW GBD
YLD type	Incidence-based	Prevalence-based
Mortality model	Basic	CODEm
Disease model	DisMod II	DisMod-MR
Age weighting	Yes	No
Discounting	Yes	No
Complexity	Moderate	High

Full Integrated Pipeline

Old System

Data → DisMod II → $(i, r, f) \rightarrow X(a) \rightarrow YLD_{incidence}$
→ $D(a) \rightarrow YLL \rightarrow DALY$

New System

Data → CODEm → $\hat{m} \rightarrow D(a) \rightarrow YLL$
Data → DisMod → $C(a) \rightarrow YLD_{prevalence}$
→ $DALY$

Final Master Equations

Old:

$$DALY = \sum[i(a) \cdot S(a) \cdot X(a) \cdot DW + D(a) \cdot L(a)]$$

New:

$$DALY = \sum[C(a) \cdot DW + D(a) \cdot L(a)]$$

Ultimate Conceptual Difference

OLD:

- “How many new cases × how long they suffer”

NEW:

- “How many people are suffering right now”

Final Insight

- **DisMod** → ensures epidemiological consistency
- **CODEm** → ensures mortality accuracy
- **DALY** → integrates both into one metric

One-line Interpretation

In the original GBD framework, DisMod II was used to estimate incidence and duration for incidence-based YLD with age-weighting and discounting, whereas the modern GBD framework integrates CODEm for mortality (YLL) and DisMod-MR for prevalence-based YLD, without age-weighting or discounting.

References:

1. Murray C. J. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization*, *72*(3), 429–445.
2. Murray, C. J., & Acharya, A. K. (1997). Understanding DALYs (disability-adjusted life years). *Journal of health economics*, *16*(6), 703–730. https://doi.org/10.1016/s0167-6296(97)00004-0
3. Barendregt, J. J., Van Oortmarssen, G. J., Vos, T., & Murray, C. J. (2003). A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. *Population health metrics*, *1*(1), 4. https://doi.org/10.1186/1478-7954-1-4
4. Peterson, H. M., & Flaxman, A. D. (2013). Meta-regression with DisMod-MR: How robust is the model? *The Lancet, 381*(Suppl 2), S110. https://doi.org/10.1016/S0140-6736(13)61364-1
5. Peterson, Hannah & Flaxman, Abraham. (2013). Meta-regression with DisMod-MR: how robust is the model?. The Lancet. 381. S110. 10.1016/S0140-6736(13)61364-1.
6. GBD 2013 DALYs and HALE Collaborators, Murray, C. J., Barber, R. M., Foreman, K. J., Abbasoglu Ozgoren, A., Abd-Allah, F., Abera, S. F., Aboyans, V., Abraham, J. P., Abubakar, I., Abu-Raddad, L. J., Abu-Rmeileh, N. M., Achoki, T., Ackerman, I. N., Ademi, Z., Adou, A. K., Adsuar, J. C., Afshin, A., Agardh, E. E., Alam, S. S., … Vos,

T. (2015). Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. *Lancet (London, England)*, *386*(10009), 2145–2191. https://doi.org/10.1016/S0140-6736(15)61340-X

7. Global Burden of Disease Study 2013 Collaborators (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)*, *386*(9995), 743–800. https://doi.org/10.1016/S0140-6736(15)60692-4
8. GBD 2013 Mortality and Causes of Death Collaborators (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)*, *385*(9963), 117–171. https://doi.org/10.1016/S0140-6736(14)61682-2
9. Kim, Y. E., Jung, Y. S., Ock, M., & Yoon, S. J. (2022). DALY Estimation Approaches: Understanding and Using the Incidence-based Approach and the Prevalence-based Approach. *Journal of preventive medicine and public health = Yebang Uihakhoe chi*, *55*(1), 10–18. https://doi.org/10.3961/jpmph.21.597

Example:

Simulation example for Calculate DALY for Diabetes.

Steps involved:

1. Realistic dummy data generator (India, state–year–sex–age)
2. Mixed data problems (incidence vs prevalence, deaths vs rates)
3. DisMod-like harmonization
4. CODEm-like mortality modeling
5. Final DALY calculation (Diabetes)

This is teaching-grade but structurally correct.

CALCULATION OF DALY FOR DIABETES

```
## STEP 0: Setup
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

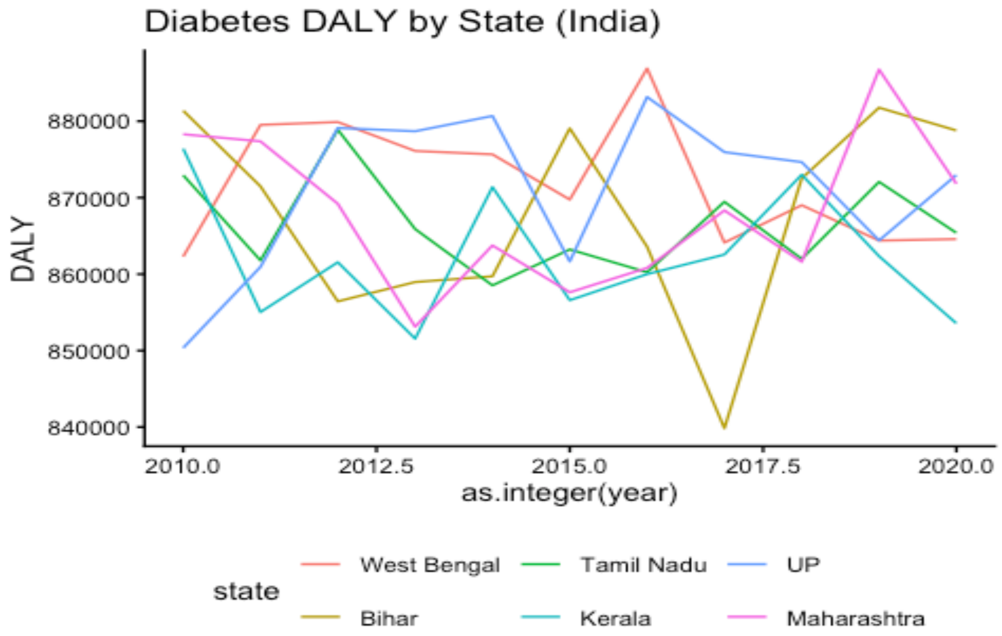
```
library(tidyr)
library(ggplot2)
set.seed(123)
#
## STEP 1: Define Dimensions
years= 2010:2020
states= c("West Bengal","Bihar","Tamil Nadu",
          "Kerala","UP","Maharashtra")
sex= c("Male","Female")
age= seq(20,80,by=10)
#
df= expand.grid(year= years,
                  state=states,
                  sex= sex,
                  age= age)
#
## STEP 2: Generate Population + Covariates
df$population= round(runif(dim(df)[1],50000,200000))
df$income= runif(dim(df)[1],500,3000)
df$education= runif(dim(df)[1],5,12)
df$urbanization= runif(dim(df)[1],0.2,0.8)
df$bmi_mean= runif(dim(df)[1],22,30)
df$sbp_mean= runif(dim(df)[1],110,150)
#
## STEP 3: Generate Mixed Epidemiological Data
# Some states → prevalence
# Others → incidence
df$epi_type= ifelse(df$state %in% c("Kerala","Tamil Nadu"),
                    "prevalence","incidence")
df$prevalence= ifelse(df$epi_type=="prevalence",
                      runif(dim(df)[1],0.05,0.15),NA)
df$incidence= ifelse(df$epi_type=="incidence",
                     runif(dim(df)[1],0.005,0.02),NA)
#
## STEP 4: Generate Mortality Data (Mixed Format)
df$mort_type= ifelse(df$state %in% c("UP","Bihar"),
                     "rate","count")
df$death_rate= ifelse(df$mort_type=="rate",
                      runif(dim(df)[1],0.005,0.02),NA)
df$deaths= ifelse(df$mort_type=="count",
                  round(runif(dim(df)[1],500,5000)),NA)
#
## STEP 5: Standardize Mortality (CODEm Pre-step)
# Convert everything → deaths
df$deaths_std= ifelse(!is.na(df$deaths),
                      df$deaths,
                      df$death_rate*df$population)
#
## STEP 6: CODEm-like Mortality Modeling
# Use predictors to smooth deaths
```

```
codem_model= lm(log(deaths_std)~age+sex+income+education+urbanization+
                  bmi_mean+sbp_mean, data= df)
#
df$deaths_codem= exp(predict(codem_model))
#
## STEP 7: Life Expectancy
df$life_expectancy= 80-age
#
## STEP 8: YLL Calculation
df$YLL= df$deaths_codem*df$life_expectancy
#
## STEP 9: DisMod-like Harmonization
# Convert incidence ↔ prevalence
#
# Assumption:
#   (1) Chronic disease → remission = 0
#   (2) Duration ≈ 10 years
df$duration= 10
df$prev_from_inc= df$incidence*df$duration
df$prevalence_final= ifelse(!is.na(df$prevalence),
                            df$prevalence,df$prev_from_inc)
#
## STEP 10: Add Disability Weight
df$DW= 0.2 # diabetes average DW
#
## STEP 11: YLD Calculation
df$YLD= df$prevalence_final*df$population*df$DW
#
## STEP 12: DALY Calculation
df$DALY=df$YLL+df$YLD
#
## STEP 13: Aggregation
# By State
state_summary=df %>%
  group_by(state,year) %>%
  summarise(
    DALY= sum(DALY),
    YLL= sum(YLL),
    YLD= sum(YLD))

## `summarise()` has grouped output by 'state'. You can override using the
## `.groups` argument.

#
# By Sex
sex_summary= df %>%
  group_by(sex) %>%
  summarise(DALY=sum(DALY))
#
## STEP 14: Visualization
ggplot(state_summary,aes(x= as.integer(year),y= DALY,color= state)) +
  geom_line() +
  labs(title = "Diabetes DALY by State (India)")+
```

```
  theme_classic()+
  theme(legend.position = "bottom")
```



AGE-STANDARDIZED DALY RATE (ASR)

We need this age-standardized when different states have different age structures. So, we standardize:

$$ASR = \frac{\sum(w_i \times r_i)}{\sum w_i}$$

Where:

- r_i= age-specific DALY rate
- w_i= standard population weight

```
## STEP 1: Create Standard Population
# (Use dummy WHO-like weights)
std_pop= data.frame(
  age= seq(20,80,by=10),
  weight= c(0.15,0.14,0.13,0.12,0.11,0.10,0.09))
#
##STEP 2: Calculate Age-specific DALY Rate
df$DALY_rate = df$DALY/df$population
#
## STEP 3: Merge Standard Population
df= merge(df,std_pop,by = "age",all.x = TRUE)
#
## STEP 4: Compute Age-standardized Rate
ASR= df %>%
  group_by(state,year,sex) %>%
```

```
  summarise(
    ASR_DALY= sum(DALY_rate*weight)/sum(weight))

## `summarise()` has grouped output by 'state', 'year'. You can override u
sing the
## `.groups` argument.

#
## STEP 5: Visualization
ggplot(ASR,aes(x= year,y= ASR_DALY,color= state)) +
  geom_line() +
  labs(title = "Age-standardized DALY Rate (Diabetes)")+
  theme_classic()+
  theme(legend.position = "bottom")
```

RISK ATTRIBUTION (BMI → DIABETES DALY)

CORE IDEA

$$DALY_{BMI} = DALY_{Diabetes} \times PAF_{BMI}$$

We assume: the Relative Risk Model

$$RR = e^{\beta(BMI - BMI_{opt})}$$

Where:

- BMI_opt = 22
- β = 0.08 (dummy)

And define Population Attributable Fraction (PAF):

$$PAF = \frac{\text{Disease burden with exposure} - \text{Disease burden without exposure}}{\text{Disease burden with exposure}}$$

For a single exposure level:

$$PAF = \frac{P(RR - 1)}{P(RR - 1) + 1}$$

Or Simply:

$$PAF = \frac{RR - 1}{RR}$$

(assuming whole population exposed)

```
## STEP 1: Define Relative Risk Model
df$BMI_opt=22
df$beta=0.08
```

```
df$RR_BMI= exp(df$beta*(df$bmi_mean-df$BMI_opt))
#
## STEP 2: Calculate PAF
df$PAF_BMI = (df$RR_BMI - 1) / df$RR_BMI
#
## STEP 3: Attributable DALY
df$DALY_BMI = df$DALY*df$PAF_BMI
#
## STEP 4: Aggregate
BMI_summary <- df %>%
  group_by(state, year) %>%
  summarise(
    DALY_total = sum(DALY),
    DALY_BMI = sum(DALY_BMI))
```

```
## `summarise()` has grouped output by 'state'. You can override using the
## `.groups` argument.
```

```
#
## STEP 5: Compare
ggplot(BMI_summary, aes(x = year)) +
  geom_line(aes(y = DALY_total, color = "Total")) +
  geom_line(aes(y = DALY_BMI, color = "BMI Attributable")) +
  labs(title = "BMI Attributable Diabetes DALY")+
  theme_classic()+
  theme(legend.position = "bottom")
```

ADVANCED (GBD-LIKE CONCEPT)

Real GBD uses:

$$PAF = \frac{\int RR(x)P(x)dx - \int RR(x)P^*(x)dx}{\int RR(x)P(x)dx}$$

Where:

- $P(x)$= observed BMI distribution
- $P^*(x)$= ideal (TMREL)

We simplified using mean BMI