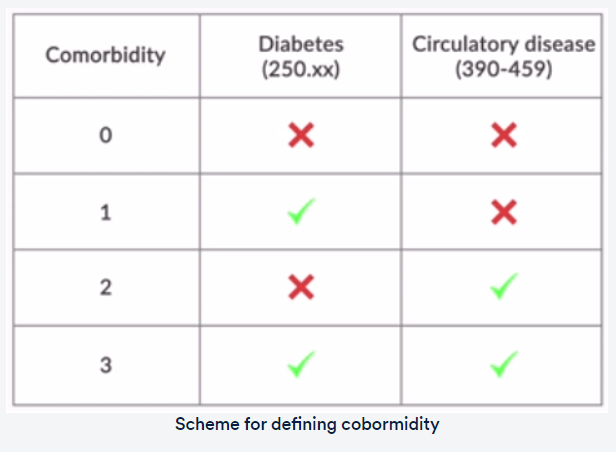
`Problem Statement-WeCare

A care management organisation called WeCare wants to identify among its diabetic patients, the ones that are at high risk of getting re-admitted to the hospital. They wish to intervene by providing some incentive to these patients that will help them improve their health. As the star analyst of this organisation, your job is to identify high-risk diabetic patients through risk stratification. This will help the payer to decide what are the right intervention programs for these patients.

**Data preparation**

1. Remove redundant variables, duplicate rows/columns
2. Check for missing values and treat them accordingly.
3. Scale numeric attributes and create dummy variables for categorical ones.
4. Change the variable 'readmitted' to binary type by clubbing the values ">30" and "<30" as "YES".
5. Create the derived metric 'comorbidity', according to the following scheme -



**Data Exploration**

1. Perform basic data exploration for some categorical attributes
2. Perform basic data exploration for some numerical attributes

**Model Building**

1. Divide your data into training and testing dataset
2. Train and compare the performance of at least two machine learning algorithms and decide which one to use for predicting risk of readmission for the patient.
3. Important feature for each model is calculated.
4. Use trained model to stratify your population into 3 risk buckets:
   * High risk (Probability of readmission >0.7)
   * Medium risk (0.3 < Probability of readmission < 0.7)
   * Low risk (Probability of readmission < 0.3)