



# THE RISING TREND IN DIABETES MELLITUS – A STATISTICAL REVIEW

Mamatha M.\* & Nagaraja Rao C.\*\*

## ABSTRACT

*Diabetes Mellitus (DM) is one of the most burdensome chronic disease that is increasing in epidemic proportions throughout the world. The prevalence of diabetes is increasing at alarming rate unidentified, untreated or poorly controlled diabetes can result in disabling and life threatening long term complications including blindness, hearing loss, loss of limbs, heart disease and kidney failure. This article includes Statistical implications, Statistical study of therapies, clinical issues, complications and intervention strategies. A Capture-Recapture technique for prevalence studies is provided. The findings of the sample study are also presented.*

## 1. Introduction

Diabetes Mellitus is a group of syndrome, characterized by hyperglycemia, altered metabolism of lipids, carbohydrates and proteins and an increased risk of cardiovascular diseases. DM can occur in a variety of situations. Type 1 DM occurs due to the destruction of insulin producing units in the pancreas. Viewed as an auto immune disease, it occurs because the body's immune systems goes

---

\* Dept. of Statistics, B.M.S. College for Women, Bangalore.

\*\* Dept. of Statistics, Vijaya College, Bangalore.

haywire, attacks and destroys insulin producing cells, known as beta cells, in the pancreas. Type 2 DM occurs when the cells stop responding to insulin or when the pancreas produces insulin in short supply. Gestational diabetes is an ephemeral condition associated with pregnancy. Diabetes may also develop as a secondary condition linked to another illness or medications that can interfere with the production of Insulin or its action.

The prevalence of diagnosed type 2 DM world wide has been increased dramatically over the past 40 years. Higher than optimum blood glucose ranks in the top five major determinants of world wide mortality, accounting for 3.16 million deaths a year (Lopez et al., 2006, Danaei et al., 2006). Diabetes causes one death every 10 sec. and one amputation every 30 sec. and is major cause of adult onset blindness, cardiovascular disease and renal failure. India, China and USA are the top 3 countries having largest number of people. Bangladesh, Brazil, Indonesia, Japan and Pakistan will also appear in the lists next to these top 3 countries. Asia, currently has around two thirds of the world's type 2 diabetes or around 90 million people according to Paul Zimmet, the chairman of the international obesity task force. Four out of five of the world's most diabetic populations are also in Asia, including India, China, Pakistan and Japan. The number of diabetics in Asia is set to increase to 120 million by 2010. By 2025 the number of Asian with the disease could increase to 198 million (Zimmet et al., 2001). India was having about 33 million people with type 2 DM at the end of 2005, which is major contributed by the urban population (Ramachandran, 2001). 10–12% adults in urban India suffer from DM. The prevalence of the disease is around 3-4% countryside. 50% of diabetics in India do not know that they suffer form DM (Agarwal, 2007). Indians with diabetes are 25 times more likely to develop blindness, 17 times more likely to undergo amputation, two four times more likely to develop myocardial infarction and twice as likely to suffer a stroke than non-diabetes (Pradeepa et al., 2002). Population Statistics of the current status and future estimation of DM are presented in the following tables 1 & 2.

**Table 1: Statistics of the number of people with type 2 DM**

Country	Diabetic population (in millions)		
	2000	2025	2030
World	171	300	366
India	31.7	>57	79.4
China	20.8	>37	42.3
USA	17.7	>21	30.3

Source: Trendz, The latest in medicine (2006)

**Table 2: Statistics of the estimated population of type 2 DM for the year 2025**

Country	Population Statistics (in millions)
India	57
China	37
US	21
Russia	12
Japan	08
Brazil	11
Indonesia	12
Pakistan	14
Mexico	11
Egypt	08

Source: Trendz, The latest in medicine (2006)

## 2. Statistical Implications

Statistics is the science of collection, classification, presentation and analysis of numerical data. It is used in almost all the fields of study. Various statistical tools and techniques are applied in estimation and clinical evaluation of diabetes. Efficacy of various drugs and tests are statistically studied and compared. Prevention/management of diabetes are statistically reviewed. When observations, discrete or continuous are available on the characteristics of a large number of individuals, statistics helps to condense the data as far as possible without losing any information of interest. Graphs and diagrams are used to represent these data which makes it very attractive and helps in comparison. Statistical constants enables to comprehend these data in a single effort of the significance of the whole data. The deviations of the values of the parameters from the standard values in clinical evaluation of diabetes can be measured using various measures of dispersion like coefficients of range, mean deviation, quartile deviation and standard deviation. In practice, Statistics enables to represent the relationship between two variables by simple algebraic expressions like polynomials, exponential or logarithmic functions.

Statistics is used to estimate the values of one variable which would correspond to the specified values of the other variable using regression analysis. It measures the relationship between two or more variables (diabetic population and age) using correlation coefficients. Some of the attributive data are also correlated using rank correlation coefficients. Statistics provides the intra-class correlation to know how the members of a family or group are correlated among themselves with respect to some of their common characteristics (diabetes). Statistics provides a sampling technique to study the characteristics of a large population (diabetic population) whenever complete enumeration is difficult or not possible due to various factors like time, money and labour. It provides procedures/large and small sample tests (z test, t test) in testing the hypothesis (diabetes related with heredity). Chi-square test can be applied to test the population variance. Statistics provides a very powerful test for testing the significance of the discrepancy between theory and experiment (used in clinical evaluation). It enables to find if the deviation of the experiment from theory is just by chance or is it really due to the inadequacy of the theory to fit the observed data. It is used to study whether the two attributes (incidence of diabetes and geographical factors) are independent or not. Another important statistical tool is the theory of estimation used to estimate the unknown parameters (constants) through point or interval estimation. Odds ratio measures the association between independent and dependent variables. These are some of the statistical implications to review the increasing trend in diabetes.

### **3. A Statistical Study of Therapies, Clinical Issues, Complications and Intervention Strategies**

There exists various therapies for the treatment of DM. Adequate insulin therapy prevents hyperglycemic symptoms, prevents ketoacidosis and reduces the incidence of micro vascular and neuropathic complications of diabetes. Acarbose is known to lower the LDL/HDL ratio by 26.7%. One of the studies shows that the treatment with Atorvastatin enabled a decrease in the incidence of major cardiovascular incidents by 37% ( $P=0.001$ ). There exists many complications of DM. Approximately 30% of patients with diabetes develop skin manifestations during the course of their illness. A study reveals that the risk of major congenital anomaly in an infant born to normoglycemic mother is 1-2%, whereas the risk increases by four to eight fold in infant of diabetic mother. Diabetes increases the risk of Ischaemic stroke. Diabetic Retinopathy, is a microvascular complication, which is almost inevitable in the diabetic patient. Its prevalence is around 30% to 40%. DM is associated with Non-Alcoholic Fatty Liver disease (NAFLD) including its severe form, Non-Alcoholic Steato Hepatitis (NASH). Among patients with diabetes, the risk of chronic liver disease is doubled, independent of alcoholic liver disease or viral Hepatitis. Diabetes

also increases the risk of primary liver cancers and death from liver Cirrhosis. The presence of Gestational DM during pregnancy has important consequences for both mother and child. The offspring of these women were also at increased risk for admission to the neonatal intensive care unit, large-for-gestational-age birth weight, and preterm birth before 37 weeks. People with diabetes are up to four times more likely to have cardiovascular disease (CVD) as people without diabetes.

In a population-based retrospective cohort study, the regression techniques were used to plot the relation between age (x) and cardiovascular event rates (Y), using a linear ( $Y=a+bx$ ), exponential ( $Y=ab^x$ ), or polynomial (quadratic) equation ( $Y=a+bx+cx^2$ ). The line of best fit between these variables were used to establish the average age at which men and women with or without diabetes moved from low risk to moderate risk; and from moderate to high risk for each set of outcomes. Cox's proportional hazards model were used to calculate age-adjusted and sex-adjusted hazard ratios for the rates of myocardial infarction in people with diabetes. Recent data shows that HCV infection is also a significant risk factor for the development of type 2 diabetes. The risk of type 2 diabetes is strongly influenced by inheritance. Recently, Grant and colleagues (2006) reported on the association of a common microsatellite (DG10S478) within intron 3 of the transcription factor 7-like 2 gene (TCF7L2) with type 2 diabetes. The noncoding single-nucleotide polymorphisms rs 12255372 and rs7903146 were in strong linkage disequilibrium with DG10s478 ( $r^2=0.95$ ) and ( $r^2=0.78$ ), respectively) and showed similarly robust associations with type 2 diabetes ( $p < 10^{-15}$ ). Cox regression models was examined according to genotype, intervention, and interactions between genotype and intervention as the independent variables predicting the incidence of diabetes. The population attributable risk was estimated with data from the placebo group for each ethnic group, calculated as follows:

$$1 - (1 + [p^2HR_{\text{hom}} + 2p(1-p)HR_{\text{het}} + (1-p^2)]);$$

where p is the risk-allele frequency,  $HR_{\text{hom}}$  is the hazard ratio for homozygotes and  $HR_{\text{het}}$  is the hazard ratio for heterozygotes.

The study observed a possible interaction between POPs (Persistent organic pollutants) and obesity on the risk of type 2 diabetes. Studies were conducted and tested the hypothesis that sleep deprivation results in impaired glucose tolerance, diabetes and obesity. The systematic statistical review supports the hypothesis that habitual coffee consumption is correlated with a substantially lower risk of type 2 diabetes. Researchers at Queen's University in Belfast, Northern Ireland, studied sperm from 27 men with type 1 diabetes, with an average age of 34, and 29 non-diabetic men of similar age. It was found that while semen volume was lower in diabetic men, all

other measures including sperm concentration, output, form, structure and ability to move appeared normal. But when looked at the DNA, found much more damage in the sperm of diabetic men. Sperm damage can impair male fertility and has been associated in couple with a history of miscarriages. The diabetes prevention program research group study found that lifestyle intervention delivered over 2.8 years reduced the incidence of diabetes by 58%. In the systematic review of psychological therapy was associated with a significant improvement in glycemic control in 10 studies in children and adolescents, with a pooled absolute reduction in glycosylated haemoglobin of 0.5%. A meta-regression analysis was conducted to assess the impact on glycemic control of different strategies for quality improvement (QI) in adults with type 2 diabetes. For the analysis, to avoid spurious precision in patient level outcomes, an effective sample size for each of the trials was calculated as follows:

$$N(\text{effective}) = (K+M)/(1+(M-1) \times ICC);$$

where K indicates the number of cluster, and ICC-the intraclass correlation coefficient. Two strategies were associated with statistically significant incremental reductions in HbA1c values. Interventions involving team changes reduced HbA1c values by 0.33%:  $p=0.004$ ) than did interventions without team changes and interventions involving case management reduced HbA1c values by 0.22 more (95% CI, 0.44%;  $p=0.04$ ) than did interventions that did not use case management. Many studies have conformed the useful role of yoga in the control of DM. Fasting and postprandial blood glucose level came down significantly. Good glycemic status can be maintained for long periods of time. There was a lowering of drug requirement and the incidence of long periods of time. There was a lowering of drug requirement and the incidence of acute complications like infection and Ketosis was significantly reduced. There was a fall in free fatty acids suggesting a better Insulin sensitivity and decrease in insulin resistant, with probable protective effect on beta cell function. There is a beneficial effect on the co-morbid condition like hypertension and dyslipidemia.

## 4. Statistical Model for Prevalence Studies

Capture - Recapture technique (CR): This statistical technique is used for estimating the number of individuals in a closed population. When lists of people with the characteristics of interest (for example, diabetic patients) are available, these can be used as the 'Captures'. Those people present on different lists, (the duplicates) are in effect 'recaptured'. The number of cases appearing on different lists can then be used to calculate the missing cases, total cases and subsequently the prevalence. The basic assumptions of CR method are as follows.

- (i) The study population is closed, that is there should be no significant changes in the population under study during the given time period.
- (ii) The lists are independent of one another.
- (iii) All members of the population have the same probability of being captured.
- (iv) All population members can be matched on all lists.

In diabetes studies, the common sources used depend on the structure and level of medical care. Hospital diabetic clinics and family doctor lists are potential good sources of capture (Bruno et al., 1992), but there may be problems with both these sources. Hospital in-patient data notoriously under-record actual numbers of diabetic cases, sometimes by up to 40-60%. District diabetic registers are good sources for diabetic capture, usually with relatively high ascertainment levels. Type 1 diabetes provides an easier model, because of lower numbers and ease of diagnosis compared with type 2 diabetes. Local diabetic association membership lists are source of capture that gives variable ascertainment but confidently of these lists sometimes makes them impossible to use. In general, however, to increase accuracy and reduce the effect of dependence, more than two sources should ideally be used. A combination of full name, gender and age is usually sufficient, though the addition of a postcode (if available) is useful. National identification numbers are obviously ideal, and have been used. The accuracy of the identifier cannot be over-emphasized: in the absence of a single national number, different identifiers may give different results. Difficulties such as these have limited the effectiveness of CR techniques in developing countries. Matching of the patient identifiers is generally not a problem. The numbers involved in most lists make manual inspection laborious and insufficient, but many computer software packages such as SPSS can do the job efficiently.

Two-list CR technique: Two-list CR is a modified version of original technique. The illustration is provided in the venn diagram as shown in figure 1.

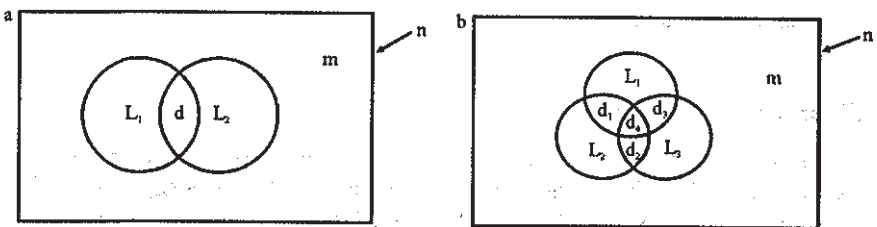


Fig. 1 : Models of CR technique



(a) Two-list, (b) Multiple-list (in this case three sources), In-list  $n$ ;  $d$ -duplicates;  $m$ -missing cases;  $n$ -total population; Multiple-list capture-recapture.

The box represents the total population required to be measured ( $n$ ), and inside are the two overlapping lists ( $L_1$  and  $L_2$ ), and surrounding them are the missing cases ( $m$ ). The total cases,  $n$ , can be calculated by Chapman's formula (Chapman, 1954), which is as follows:

$$n = \frac{(L_1 + 1)(L_2 + 1)}{(d + 1)} - 1;$$

where  $L_1$  and  $L_2$  are the numbers in the two lists, respectively, and  $d$  is the duplicates. Similarly, 95% CIs can be calculated as

$$n \pm 1.96 = \sqrt{\frac{(L_1 + 1)(L_2 + 1)(L_1 - d)(L_2 - d)}{(d + 1)^2(d + 2)}} - 1$$

Although two list CR is simple, it is vulnerable to serious error if the assumptions previously discussed are not validated. In particular, independence of lists is very important. If the lists are positively dependent, then this will result in an underestimation of the population size. Conversely, if the lists are negatively dependent, then an overestimation of population size will result. Use of three or more lists allows greater accuracy and tends to reduce the difficulties of list dependence or independence. The calculations with multiple-list CR are highly complex, and usually involve Poisson log-linear modeling (Cormack, 1989). This gives an estimate of missing cases, and also provides information on the interdependence of sources. Multiple-list CR has been widely used and is now the preferred method (Laporte et al., 1993). The complex statistics are greatly eased by the use of computer software such as 'GLIM'. As mathematics become increasingly complex beyond three lists, and recent studies have demonstrated that three good non-independent lists are optimal.

### CR for type 2 diabetes

CR has mostly been applied to type 1 diabetes, but there is some limited experience of CR in type 2 diabetes prevalence estimation. Bruno et al., (1992) conducted a large population-based prevalence study of type 2 diabetes. They used the following sources : (i) diabetic clinic and general practioner lists; (ii) hospital discharge data; (iii) drug prescription records (insulin and oral agents); and (iv) prescription



reimbursement requests. The first capture included both hospital and GP lists because of significant interdependence. An overall diabetes prevalence of 2.8% was found by CR calculations, with 95% CIs of 2.4-3.1%. For type 2 diabetes only, the prevalence was 2.7% (CI 2.6-2.8). Studies in Liverpool, UK, used a total of six separate lists namely (i) hospital adult diabetic clinic register; (ii) children's hospital diabetic clinic list; (iii) hospital discharge data; (iv) hospital retinal clinic list; (v) research database for stroke admissions; and (vi) general practitioner lists. Lists (i) and (ii) were combined as first list, (iii), (iv) and (v) were combined as second list, but (vi) was left as a single dataset. This provided three large lists with very little dependence. An overall CR calculated diabetes prevalence of 3.1% (95% CI 3.0-3.2%) was obtained, compared to a figure of 1.5% (95% CI 1.4-1.5%) by electronic data linkage using the same datasets. The specific type 2 diabetes CR adjusted prevalence rates in the study showed rates varying between 3.2% and 6.7% in different areas of the population, and there was a strong relationship between the prevalence of type 2 diabetes (but not of type 1), and indices of social deprivation. Although CR technique has many advantages, it has a few drawbacks. Some of them are the following.

- (i) It is said to be rapid and inexpensive, but no detailed comparative cost-effectiveness assessments have been made.
- (ii) Results comparing CR with 'gold standard' population surveys are lacking.
- (iii) It is used primarily to assess the incidence of type 1 diabetes than type 2 diabetes
- (iv) It must be meticulously applied, as otherwise factors such as source interdependence and identifier failures can lead to seriously misleading results.

## 5. Sample Study

A small sample study was conducted in Bangalore South on the topic "A Statistical assessment of risk factors of Diabetes Mellitus" during the month of Nov / Dec 2007 by considering 100 diabetic patients as the sample using purposive sampling. The outcomes of this sample study reveals the following information.

1. It was noted that 43% diabetes are employed and 57% are non-employed. Out of 43% diabetic employees, 25%, 9% and 9% are distributed as servicemen, business men and others respectively.
2. It was found that that mean diabetic age = 46.125 yrs, S.D = 12.11 yrs and coefficient of variation = 26.25. It was observed from the following figure

(fig. 2) that concentration of diabetes was high for the age group (30-60) yrs and very low for the age groups (0-30) and (60-75) yrs. Hence (30-60) yrs age group was considered as the modal diabetic age group.

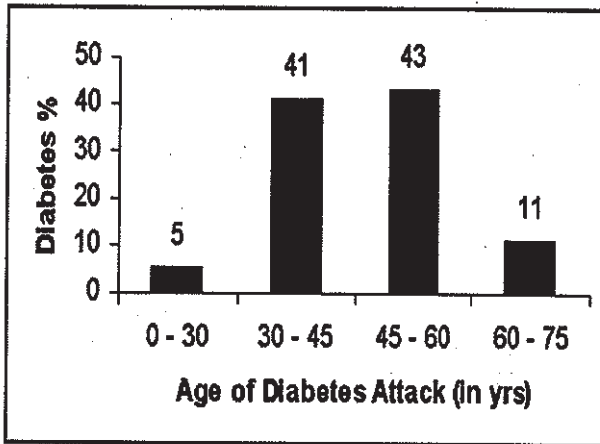


Fig. 2 : Bar diagram showing the % diabetics according to attack

3. Stress related to diabetes was equally probable for both the sexes. (Chisquare (cal) = 1.4685 < Chisquare (table) at 1 d.f. both at 5% and 1% level of significance).
4. There was no significant difference in occurrence of various stress factors for male and female diabetics (Chisquare (cal) = 12 < Chisquare (table) at 4 d.f. at 5% level of significance).
5. It was noticed that B.P, poor/blind vision and heart problems are comparatively high after diabetes attack. Majority of the diabetes suffer from B.P. and least from poor/hearing loss.
6. It was observed from this study that 37% diabetes have a family history of diabetes.
7. Out of various intervention strategies, 38% diabetics adopt exercise / walk and only 6% adopt homeopathic treatment to control / manage diabetes. It can be concluded that daily exercise/walk has become the major intervention strategy is controlling / managing diabetes.

8. It was found that odds ratio for diabetic complications according to age was 0.4607. Hence it can be concluded that diabetic complication is 0.4607 times as frequent among those diabetes of the age group (0-45) years than among those of age group >45 years.
9. It was noted that diabetic complications are comparatively less for the people practicing exercise (Odds ratio calculated = 0.2613). Hence it can be concluded that the occurrence of diabetic complication is 0.2613 times as frequent among those who exercise than who do not exercise in the study population.

## Discussion

Diabetes mellitus is a world wide epidemic of the 21<sup>st</sup> century. It is a serious and economically devastating illness that has been reached epidemic proportion in both industrialized and developing countries, poses a major threat to public health in the 21<sup>st</sup> century. In conjunction with genetic susceptibility, DM is brought on by environmental and behavioral factors such as a sedentary lifestyle overly rich nutrition, obesity etc. Age and gender also influence the prevalence of DM. Out of various methods, CR technique has been widely used for determination of type 1 DM. The following gives the recent developments for controlling /managing diabetes.

- (i) Scientists hope that the insulin produced in the genetically modified cow could be on the market in the next couple of years. There are about 200 million diabetics world wide and the Argentina Scientists said that just 25 insulin producing cows could be enough for Argentina's 1.5 million diabetics (TOI, Dt. 5-6-07).
- (ii) The medical breakthrough, which shows that stem cells taken from the umbilical cord of newborns can be engineered to produce insulin, offers promise to cure diabetes in the future (TOI, Dt. 28-5-07).
- (iii) Researchers pointed to new biological basis for the diabetes by identifying seven new genes connected to the most common form of diabetes. (TOI, Dt. 28-4-07).
- (iv) Scientists recently created a pill from insulin (oral insulin pill) which could be inhaled instead of repeated injections making it the first new method to get that particular hormone into the body, since it was discovered in 1921. The inhalable treatment option is now undergoing trials in India (TOI, Dt. 25-6-07).

The following lists some of the future work to be initiated:

- (i) Further work is initiated on introduction of National register system to obtain accurate statistics on incidence, complications, mortality and hence optimum strategies for management of diabetes be undertaken effectively.
- (ii) Further statistical research should be undertaken to study about new genes which are connecting to a common form of diabetes.
- (iii) Future research should focus more on new strategies (stem cell therapy) to produce insulin.
- (iv) Statistical modeling on causes and spread of DM among different groups of people may be initiated.
- (v) Multiple correlation on various factors of DM (Variables) may also be undertaken.

## References

1. Agarwal, Dream 2047, April, e mail : dryatish @ yahoo.com (2007).
2. Bruno G, Bergero G, Vuolo A, Pisue, Pagano G. A population-based prevalence survey of known diabetes in northern Italy *Diabetologia*; 35:851-6 (1992).
3. Chapman. G. The estimation of biological population. *Ann Math Stat*; 25:1-15 (1954).
4. Cormack R.M., Log-linear models for CR *Biometres*; 45: 395 – 413 (1989).
5. Danai G, Lawes CMM, Vander Hoorn S, Murry C.J.L, Ezzati & regional mortality from Ischaemic heart disease and stroke attributable to higher than- optimum blood glucose concentration comparative risk assessment, *Lancet*; 368: 1651-59 (2006).
6. Laporte R. E., Mccarty D., Bruno G., Tajima N., Baba S. Counting diabetes in the next millennium, application of CR technology *Diabetes care*; 16: 528-34 (1993).
7. Lopez A.D., Mathers C.D., Ezzati C.D., Ezzati M., Jamison D.T., Murray C.J., Global and regional burden of disease and risk factors 2001; Systematic analysis of population health data, *Lancet*; 367: 1747- 57 (2006).
8. Pradeepa R., Deepa, R., Mohan V. Epidemiology of diabetes in India current perspective and future projections. *J Indian Med Assoc*, 100 (3): 144-8 (2002).
9. Ramachandran A, Snehalatha C., Kapur A., Vijay V., Mohan V., Das A.K., Rao P.V. Diabetes Epidemiology study group in India, *Diabetologia*; 44: 1094-101 (2001).
10. TOI-Times of India, April 28, Saturday (2007).
11. TOI-Times of India, May 28, Monday (2007).

12. TOI-Times of India, June 5, Tuesday (2007).
13. TOI-Times of India, June 25, Monday (2007).
14. Trendz- the latest in medicine, Vol. IV/issue-4/ October-December (2006).
15. Zimmet P, Albery K.G, M.M. Shaw. J. Global and societal implications of the diabetes epidemic. *Nature*, 414(6865). 782-787 (2001).