2

Methods for comparing mortality risks from different diseases

Author : Mr. P.N. Lee

Date : 23.9.75

1. Three separate events have induced this paper to be written. The first of these is that the T.S.C. in preparing, at the E.C.'s request, a list of 6 titles on which a monograph might usefully be written, have suggested as one possibility the title "The importance of smoking - associated diseases in relation to death from other causes". The idea of this monograph, as I understand it, is to provide a reference document to answer questions such as "Is lung cancer a more serious hazard to life than motor accidents?".

2. The second event was the statements made by Mr. John Treasure and Sir. Richard Doll in the Thames Television programme "This week" on 11.9.75. These were as follows (slightly reduced for convenience).

Mr. Treasure: "30, 000 people die every year of lung cancer. Of those 30, 000 10% of them are non-smokers, therefore, it's 27, 000 deaths a year that could be caused by smoking. There are 750, 000 people die every year, so lung cancer as a cause of death is about 1 in 30 or rather less. So you have here a cause of death which is pretty small".

Sir Richard: "The figures he gave are just wrong actually. We've never had 750, 000 deaths in a year in this country, the greatest number has been just under 600, 000. I think 750, 000 refers to births not deaths, so that brings it down to one in 22. But it's not really very sensible to compare deaths with lung cancer in smokers with all deaths in smokers and non-smokers. You ought to relate it to the number of deaths in smokers and this is down to about one in nine."

Mr. Treasure: "The average age of death in this country of men is 67.3 years, the average age of death of the 30, 000 people who died of lung cancer is 66.5, so that whether or not smoking has any effect in this area it certainly does not result in a premature shortening of life on average".

Sir. Richard: "It sounds good but it's obvious why you can't make this comparison.

100443722

You see deaths from all causes include deaths in infants and you've got to live to the age of 20 or 30 at any rate before there's any chance of getting lung cancer. If you want to work out how much life you lose if you die of lung cancer, what you must say is - what is the normal expectation of life of a man of the age of someone who develops lung cancer and if you do this you find that the man that dies of lung cancer loses 12½ years on average".

3. The third event was the statement made by Dr. M.A. Kastenbaum when I visited him at the Tobacco Institute. He said that assuming smokers had no excess risk of mortality one would expect that smokers would die 11 years earlier than non-smokers.

4. Clearly, before preparing a monograph along the lines T.S.C. suggest, it is sensible to think first about what is a valid method of statistically representing the risk to life of a disease. This paper, therefore, looks at the various possible methods available and assesses their advantages and disadvantages. The aim of the paper is to provoke discussion among S.S.C. members and arrive at an agreed statistic (or set of statistics) which would be sensible to use in the preparation of the monograph.

5. In order to discuss these methods more meaningfully we shall calculate various statistics using real data. For the sake of example we shall use figures relating to lung cancer in U.K. males in 1971. Table 1 presents the basis data to be used. The colums 1-9 give values by 5 years age group of the following parameters:-

i) i age group subscript ($i = 1, \ldots 17$)
ii) $years_i$ range of ages for age group i
iii) Y_i approximate midpoint of range used in subsequent work.
iv) A_i number alive (in '000's) of age group i
v) N_i number dying of all causes in age group i
vi) NL_i number dying of lung cancer (ICD 162) in age group i ($=100\ N_i/A_i$)
vii) R_i death rate per 100, 000 from all causes in age group i
viii) RL_i death rate per 100, 000 from lung cancer in age group i ($=100\ N_i/A_i$)
ix) PS_i proportion of smokers in age group i

These figures were extracted from the Registrar General's Statistical Review, RP1 and OPC...

6. We will now go through a number of statistics (S) of possible relevance, define and calculate them and discuss their advantages and disadvantages. It is appreciated that the wideness of the age-groups used and the inaccuracies involved in assuming that all deaths occur at the centre of the age groups may result in not very precise answers. However, for our purpose, discussing methods, great precision is not necessary.

7. <u>S1 - Number of deaths from lung cancer</u>

Definition: $S1 = \sum NL_i$

1971 male UK value = 25, 762

S1 is clearly a statistic of some use in assessing the risk of mortality from lung cancer. On its own it, of course, does not tell the whole story as it does not take into account age of death. Clearly a cause of 25, 762 deaths among 20 year olds would be more serious than lung cancer.

8. <u>S2 - Average age of death from lung cancer</u>

Definition: $$S2 = \sum_{i=1}^{17} (NL_i \; Y_i) \; / S1$$

1971 male UK value = 66.67

The average age of death from all causes, $\sum_{i=1}^{17} (N_i Y_i) / \sum_{i=1}^{17} N_i$ is 67.15. These figures are very similar to those quoted by Treasure.[1] Doll criticised the comparison of these average ages of death on the grounds that early deaths were wrongly included and indeed if one excludes deaths in the age up to 34, the average age of death from all causes rises to 70.32 whereas that for lung cancer hardly changes (66.76). Although, Doll's criticism has some validity, he did not point out the major weakness of such a comparison. This is that it is eminently possible for a cause of death to have an average age of death <u>greater</u> than that from all causes. On Treasure's implied line of argument this would mean such a cause was beneficial!

9. The use of average age of death as a comparative tool for <u>people of different types</u> can also be extremely misleading. Following Kastenbaum's approach let us define KS_Z as the average age of death from all causes of smokers on the

100443724

assumption that current smokers have, at all ages greater than that of non-smokers with KNS_z as the corresponding average age of death of non-smokers is the same comparison. Mathematically

$$KS_z = \frac{\sum_{i=1}^{17} (U_i Y_i)}{\sum_{i=1}^{17} U_i} \quad \text{where } U_i = \frac{Z.\ NL_i\ PS_i}{100+PS_i\ (Z-1)}$$

$$\text{and } KNS_z = \frac{\sum_{i=1}^{17} (N_i - U_i)\ Y_i}{\sum_{i=1}^{17} (N_i - U_i)}$$

10. If one computes this for the extreme cases Z=1 (smoking has no effect) and Z=2 (smoking doubles the probability of death) the results are quite amusing, viz:

KS_1 = 66.48

KS_2 = 67.20

KNS1 = 67.56

KNS_2 = 67.08

11. In other words, despite all the deaths of infants being attributed to non-smokers, the average age of death of smokers is <u>lower</u> than that of non-smokers if smoking is assumed to have <u>no</u> effect but it is <u>higher</u> if it is assumed to have a very marked effect. These results demonstrate very clearly the dangers of using such a statistic average age at death for trying to make inferences as to the effect of smoking. (Prof. Burch please note!)

12. <u>S_3 = Loss of expectation of life of lung cancer decedents</u>

Definition: $$S_3 = \frac{\sum_{i=1}^{17} (NL_i\ e_i)}{S1}$$

where e_i, column 10 in Table 1, the expectation of life, has been estimated approximately as: ½ (average age of death from all causes of men dying after the

1004437725

end of the interval + that from those dying after the beginning of the interval) - Y_t e_{17} has been estimated using more detailed figures available of age at death by single years.

1971 male UK value = 11.26 years.

13. Again this figure agrees fairly well with that given by Doll. However, this also appears a rather deceptive statistic. Everyone who dies loses expectation of life and in fact the average loss of life expectation for all deaths is higher than this, 12.31 years. Of course again this is partly because of the younger deaths, but even ignoring all deaths before 20 years old it is still nearly as much, 10.30 years.

One would also expect this statistic, though I have not done the detailed calculations to have some of the unhelpful properties of S_2, with an increasing loss of life expectation being associated, in some circumstances, with a decreasing loss of life expectation.

14. So far we have looked at the various statistics that were used by Doll and Treasure and find that S_1 gives insufficient information and S_2 and S_3 information which can be totally misleading. The approach that seems more likely to succeed is to start by comparing the life-table for death from all causes with that for death from all causes with the particular cause of interest, here lung cancer, excluded (reduced life-table). Any characteristics of such a comparison must inevitably be in the correct direction without danger of bias. For example, one might compare age at X% survival or average age at death. Perforce all the ages must be higher in the reduced life-table than in the life-table including all causes and the stronger the force of mortality of the cause being excluded the greater the increase.

15. Table 2 gives the probability of survival, computed from the data Table 1, for the two life-tables. As can be seen from this table a man has a 70.4% chance of reaching the age of 65. Had lung cancer not existed this chance would have increased to 73.2%.

16. Given a life-table approach is the correct method there still remains a number of different statistics that could be used to compare, for example, deaths from lung cancer and deaths from motor-cycle accidents. Clearly the probability of

100443726

survival to age 25 is less affected by lung cancer than it is by motor-cycle accidents. However, in view of the greater number of deaths from lung cancer, the probability of survival at age 65, for example, will be more affected by lung cancer. There still remains the problem of selecting a single statistic that reflects the differences between the causes of death in some meaningful way.

17. Essentially, the comparison to be made between causes of death, will be between weighted sums of survival probabilities at different ages. The weights to be applied however, depend on the relative values applied to living at different ages. If any year of life is equally "valuable" (in some sense) then it would seem reasonable to assess a cause of death in terms of the increase in life expectation that would occur if this cause did not exist. However, if one is assessing value in terms of financial contribution to the nation, then loss of life in early years will weigh for more important than loss of life in later years.

18. It could be most interesting to have the views of the SSC on what particular statistics of this sort it would be worth presenting in a monograph of the type planned. My personal opinion is that the life-table itself ought to be presented, together with probabilities of survival (and/or death) over certain periods of interest (20-65, 35-65, 50-65 say) and increase in life expectation caused by removal of the cause. Number of deaths should also be given, but average age at death should only be presented if a warning paragraph were added.