The Use of the Personalized Estimate of Death Probabilities for Medical Decision Making

R. Giorgi\textsuperscript{1}, J. Gouvernet\textsuperscript{1}, E. Jougl\textsuperscript{2}, G. Chatellier\textsuperscript{3}, P. Degoulet\textsuperscript{3}, M. Fieschi\textsuperscript{1}

\textsuperscript{1} Laboratoire d'Enseignement et de Recherche sur le Traitement de l'Information Médicale, Faculté de Médecine de la Timone, 27 Bd Jean Moulin, 13005 Marseille, France; \textsuperscript{2} SC8-INSERM, 44 chemin de Ronde, 78110 Le Vésinet, France; \textsuperscript{3} Service d'Informatique Médicale, Hôpital Broussais, 96 rue Didot, 75014 Paris, France

Received October 28, 1999

Data coming from the French national statistics on the cause of deaths are used to calculate the probabilities of death from pathologies. These probabilities are calculated according to age, sex and place of residence of the patient to “personalise” estimate. This individual prediction of the risk of death is proposed for pathologies for which the feasibility and the utility of prevention measures had been demonstrated. Relative risks of death according to the socioprofessional category, which are coming from the scientific literature, are used to adjust the probabilities of death in function of the patient socioprofessional category. The aim of this work is to guide a scientist towards a prevention strategy according in time and characteristics of patient. The use of computers by the scientists will make possible the diffusion of such tool of prediction to improve a personalised prevention.

INTRODUCTION

Patients must frequently make choices about the medical decisions concerning them. In general, they understand the major risk factors for the most common pathologies (1). They are probably aware of the causal relation between smoking and lung cancer, between the combination of tobacco and alcohol and cancers of the upper respiratory and digestive tracts, of breast cancer in women as of a certain age, and several other risk factors. This knowledge of causal relations does not provide patients with quantitative information on the risks of death due to a pathology according to the accumulation of risk factors. Measuring these risks of death can provide the attending physician with arguments to initiate preventive measures adapted to the patient’s case. Thus the attending physician plays an important role of prevention and providing information. It is up to the physician to warn patients about the major causes of death and the degree of risk which they are facing. Physicians can use a number of indicators to transmit this information. The relative risk is the most commonly used indicator, but it is certainly not the most pertinent for informing an individual of the risk of dying from a given pathology (2,3). Consider the example of a 64
year-old woman who does not have breast cancer but who wants to know her risk of dying from breast cancer. Telling her, for example, that her risk is 2 or 3.5 times greater than the risk of a 35 year-old woman is not what she requires. Furthermore, it is possible to die of breast cancer at 35, but this information is not expressed with the relative risk if the reference class includes people of age 35.

Dupont (2) proposes a computational method for converting relative risks into absolute risks, i.e. expressing risks in terms of probability over a given time period. Indeed, providing the probability of death is a better way of answering the patient’s question. The information is provided directly (without the need to compare it to a reference value) and is expressed for a given time period. However, it raises problems concerning how to interpret probabilities, especially when they are very low and are expressed in an unusual manner (out of 10,000 or 100,000 instead of as a percentage).

This article presents a method for using the probability of death that provides the physician with an objective way of knowing the main causes of death in a healthy patient concerning the pathologies for which preventive measures exist. The purpose is to encourage the physician to use a preventive strategy based on the evidence-based medicine (4) that is suited to the patient and complies with current scientific data.

This study is part of the ESPER (Personalized Risk Estimation) project based on rules of fact-based medicine and featuring an Internet server. The project has two main objectives: 1) provide a tool for physicians to evaluate the risks of death in a patient according to the patient’s individual characteristics, and 2) offer suggestions for preventive measures that take into account the patient’s individual risk.

MATERIALS AND METHOD

We use two types of data to calculate the probabilities of death: mortality data and population data. The former is based on French national statistics on the cause of death (Public Service 8 of the INSERM, the French National Institute of Health and Medical Research). The latter is taken from the 1990 French national census; other years are based on estimations by the INSEE (the French National Institute of Statistics and Economics). The goal of this study is to provide “personalized” estimations. The probabilities of death are calculated using all available pertinent predictive variables, i.e. the patient’s age, sex, and place of residence (region or department). As the calculation is based on different cohorts, we must assume that these cohorts have a similar profile.

We applied the INSEE computational method (5). Based on mortality data and the general population, we calculated the probabilities of dying over 1 year between the ages of \( x \) years and \( x + 1 \) years, assuming that the patient is alive at \( x \) years of age:

\[
q_x = \frac{\text{number of deaths between ages } x \text{ and } x + 1}{\text{number of people alive at } x \text{ at the start of the year}}
\]

The probability of surviving to age \( x + n \) for a person alive at age \( x \) is estimated as follows:

\[
S_{x+n|x} = (1 - q_x)(1 - q_{x+1})\ldots(1 - q_{x+n-1})
\]

Thus we derived the probability of death between ages \( x \) and \( x + n \) according to the sex and place of residence:
We calculated the probabilities of death based on mortality and population data grouped over five consecutive years (from 1990 to 1994) in order to a) reduce the impact of the use of different cohorts and thereby smooth the estimations, and b) allow the computation to cover a sufficient number of cases and thereby provide consistent estimations. Then we calculated an average probability of death by taking the average values of $q_x$ for the five consecutive years studied, as well as the average proportion of death due to a cause $C$, for a given sex and place of residence. Then we can calculate the probabilities of death between ages $x$ and $x+n$ for a cause $C$ based on the sex and place of residence after having multiplied the average probability of death between ages $x$ and $x+n$ by the average proportion of deaths for cause $C$, for a given sex and place of residence. The mortality and population data used for calculating the probabilities of death are updated over time to account for changes in the causes of death.

The patients’ socioprofessional category (SPC) is another important factor to consider in order to “personalize” the probability of death (see the Discussion section for more on this point). The SPC is adjusted by modulating the probabilities of death by the relative risks of mortality per SPC taken from scientific literature. SPCs are grouped into three homogeneous classes in terms of life expectancy (6) (“workers-employees”, “merchants-technicians-middle managers”, “top executives-professionals”) in order to limit bias due to how professions are recorded.

For this study, we used the pathologies for which preventive measures are available and whose feasibility and utility have been demonstrated. Therefore certain causes of death do not appear even if they represent a relatively important share of mortality in certain age groups (e.g. leukemia in children). These have been grouped in a class called “Other causes”. The causes of death that we used are provided in Table 1.

**PRESENTATION OF PERSONALIZED ESTIMATIONS ON THE RISK OF DEATH**

After connecting to the server, the physician in his office opposite his patient, or the patient himself from home, must enter the information required to calculate the probabilities of death. During a medical consultation, the information must be rapidly available, clearly presented, and easy to interpret, while being as complete as possible. To illustrate the features developed on the ESPER server and to demonstrate its use, we chose to present two different ways of displaying the results. The first shows a hierarchy of the major causes of death at 5 years or at 10 years for the actual patient, as well as for a simulated patient. The second shows the kinetics of the probabilities of death for a given choice.

**Probabilities of death at 5 years and at 10 years**

Based on the patient’s age, sex, and place of residence, the system determines the causes of death over a set period (5 years or 10 years). The principal causes of death are thus presented and classified in decreasing order for the prediction over the given period.

Consider the example of a 45 year-old man living in the Pas-de-Calais department (patient 1) and wishing to know the principal pathologies that may cause his death over the
next 10 years. The physician simply enters these few data points in a computer and quickly
obtains a personalized prediction of the probabilities of death for this patient (figure 1). This

TABLE 1

<table>
<thead>
<tr>
<th>Causes of deaths list</th>
<th>ICD-9 code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trachea-bronchi-lung cancer</td>
<td>162</td>
</tr>
<tr>
<td>UADTa cancer (with esophagus)</td>
<td>140-149,150,161</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>174,175</td>
</tr>
<tr>
<td>Uterus cancer</td>
<td>179,180,182</td>
</tr>
<tr>
<td>Colorectal Cancer</td>
<td>153, 154</td>
</tr>
<tr>
<td>Prostatic Cancer</td>
<td>185</td>
</tr>
<tr>
<td>Ischemic heart disease</td>
<td>410-414</td>
</tr>
<tr>
<td>Cerebro-vascular diseases</td>
<td>430-438</td>
</tr>
<tr>
<td>Pneumonia and influenza</td>
<td>480-486,487</td>
</tr>
<tr>
<td>Obstructive lung diseases (except asthma)</td>
<td>490-492,494,496</td>
</tr>
<tr>
<td>Aids</td>
<td>042-044</td>
</tr>
<tr>
<td>Cirrhosis and alcohol psychosis</td>
<td>291,303,571.0-3.5</td>
</tr>
<tr>
<td>Road accidents</td>
<td>E810-E819/E826-E829</td>
</tr>
<tr>
<td>All accidents (except road accidents)</td>
<td>E800-E809/E810-E825/E830-E849</td>
</tr>
<tr>
<td>Suicides</td>
<td>E950-E959</td>
</tr>
</tbody>
</table>

a UADT: upper aerodigestive tract

is indeed a personalized prediction, not an individual one, because the adjustment is made
using only the four known variables (sex, age, place of residence and SPC). The major causes
of death for patient 1 include pathologies for which alcohol is an important risk factor (cancer
of the upper aerodigestive tract and the esophagus, cirrhosis and alcohol psychosis). Given
this information, we will consider two cases:

In the first case, patient 1 declares that he consumes between 80 and 100 grams of alcohol
daily. The physician can then give his patient quantitative information on the probabilities he
has of dying due to chronic ethylic intoxication. The server also offers to perform a simulation
by modifying one or more parameters in the query. The physician may choose to simulate the
data for a patient of the same sex, age, and SPC, but living in a department known to have a
low alcohol consumption rate (patient 2). The server provides the main causes of death with
the corresponding probabilities of death for a person who consumes little alcohol. The
attending physician can then show patient 1 that his probabilities of dying of cancer of the
upper aerodigestive tract and the esophagus, cirrhosis and alcohol psychosis up to the age of
55 would be much lower if he drank less alcohol, i.e. as low as those of patient 2. (In both
cases, they would be nearly $\frac{1}{3}$ as high). In other words, the physician can show his patient the
excess risk in terms of probability of death due to ethylic intoxication. The statistical data is
an additional argument to help the physician convince the patient to follow his advice for
preventing the disease.

In the second case, patient 1 declares that he consumes very little alcohol. If the patient
smokes, we can calculate his probability of death from upper trachea, bronchi or lung cancer,
which would be his major cause of death. Then come the cardio-vascular pathologies, which
are also susceptible to preventive measures. As in the previous case, the physician can
simulate data for a patient living in a department or a region where cardio-vascular mortality is lower.

### Présentation des probabilités de décès à 5 ans et à 10 ans

#### Prédiction à 5 ou 10 ans

Homme de 45 ans - Toutes professions confondues - Lieu de résidence : Pas-de-Calais

<table>
<thead>
<tr>
<th>Causes de mortalité</th>
<th>Probabilités de décès pour 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Votre sujet (Patient 1)</td>
</tr>
<tr>
<td>1. cancer VADS (avec oesophage)</td>
<td>1,32</td>
</tr>
<tr>
<td>2. cirrhose et psychose alcoolique</td>
<td>1,131</td>
</tr>
<tr>
<td>3. cancer de l'axe trachée-bronches-poumon</td>
<td>0,918</td>
</tr>
<tr>
<td>4. cardiopathies ischémiques</td>
<td>0,647</td>
</tr>
<tr>
<td>5. suicides</td>
<td>0,551</td>
</tr>
<tr>
<td>6. SIDA</td>
<td>-</td>
</tr>
<tr>
<td>7. autres accidents (sauf circulation)</td>
<td>0,414</td>
</tr>
<tr>
<td>8. maladies cérébro vasculaires</td>
<td>0,254</td>
</tr>
<tr>
<td>9. accidents de la circulation</td>
<td>0,219</td>
</tr>
<tr>
<td>10. cancer colorectal</td>
<td>0,154</td>
</tr>
<tr>
<td>autres causes</td>
<td>3,157</td>
</tr>
<tr>
<td>Toutes causes</td>
<td>8,873</td>
</tr>
</tbody>
</table>

Vous pouvez sélectionner une pathologie pour obtenir la courbe de probabilité cumulée de décès de votre sujet.

**FIG. 1.** Comparison of the 10 years cumulative probabilities of death for a 45 year-old man living in the Pas-de-Calais department (high level of alcohol consumption) with those of a man living in the Alpes Maritimes department (low level of alcohol consumption). The first column list the 10 principal causes of death. The two last line are for the “others causes” and for “all the causes”. The 10 years probabilities of death (in percentage) for patient 1 and for the simulated patient (patient 2) are respectively given in the second and the third column. The bottom of the window allows to modified one or more characteristic like “Sex”, “Age”, “Department” and “Socio-professional category” (i.e. simulate data for a new patient).

**Cumulative probabilities of death**

Now we will consider a particular cause of death by studying the kinetics of the risk of death. Predictions are not limited to a 5 or 10 year period. Estimations are made for the age of the patient up to 85 years. For a given patient, we present his cumulative probabilities of dying from a cause $C$ over time, based on the patient’s age and his other characteristics. This dynamic representation shows the evolution of the probabilities of death if no preventive
action is taken. It gives the physician information or arguments concerning the usefulness of
and the time to start preventive measures, given that a certain delay is required between the
start of a preventive measure and the return on a basic risk. (It is for these reasons that
probabilities of death are not given for time periods under 5 years.)

Figure 2 shows the cumulative probabilities of death due to ischaemic heart disease for a
25 year-old man living in the Greater Paris region. At 30 years of age, his cumulative
probability for dying from this pathology is $2\,000/000$. It increases rapidly around the age of 55,
and surpasses 1% at 65 years, then increases by more than 1% every 5 years. This figure
clearly demonstrates that the turning point is in the 55-60 age range, and that thereafter the
cumulative probabilities increase sharply and regularly. This provides convincing visual
information to the patient, offering a projection of his probabilities of death due to ischaemic
heart disease.

The physician has a readily available, easy-to-use graphic tool to interpret the probabilities
of death by pathology. He can determine for his patient the pathologies to be prevented and
the best time to implement preventive measures.

**DISCUSSION**

This section will discuss the biases of our method, the modulation of death predictions
based on relative risks of death by SPC, and the advantages of our approach.

**FIG.2.** Cumulative probabilities of death due to ischaemic heart disease for a 25 year-old man living in the
Greater Paris region (in percentage). Cumulative probabilities are given until 85 years of age.
To calculate the probabilities of death, we used mortality data provided by medical death certifications. We know that physicians may make errors when filling out death certificates (7,8), either because they do not know the patient (e.g. an internal medicine on night duty), or due to other errors related to inaccurate declarations as to the cause of death. This imprecision may be due to an error in determining the main cause of death or, in the case of a tumor-based pathology for example, a coding error due to lack of precision in the anatomical location. Therefore we grouped together certain pathologies to reduce this type of bias: deaths by influenza and pneumonia were grouped together, as were deaths by cancer of the cervix and cancer of the uterus.

The available data includes so much missing data concerning SPC that we cannot use it in our predictions. Nevertheless this is an important factor to consider when predicting the risk of death. Several studies have shown the level of social inequity concerning death (6,9,10) and it has been demonstrated that this disparity is increasing (11,12). The use of relative risks of mortality by SPC provided by medical literature to modulate our individualized predictions of risk of death allows us to make an adjustment on the major variables explaining mortality. By using these relative risks, we can obtain an order of magnitude concerning the calculated probability of death, but we cannot directly calculate a probability of death. The physician should interpret the probabilities of death to his patient using these criteria.

Another limitation of this study is the inability to directly account for certain behavioral factors (such as alcohol or tobacco consumption, or eating habits) in the personalized estimation of the probabilities of death. But as our example demonstrates, the physician can vary the available adjustment parameters and thereby provide indirect assistance when making a decision.

Although physicians find this type of indicator concerning risk of death to be useful, it remains difficult for them to estimate. In the cardio-vascular domain (13,14), for example, it has been shown that the estimation of risks by physicians is subjective and biased. It is preferable to use quantitative, objective means available on different media.

The Internet is one such medium that makes it possible to provide medical information available at any time and useable on a daily basis. In France, all doctors’ offices are now required to be computerized (15). Although the major objective is for administrative purposes, these computers will allow physicians to access medical information over the Internet. The ESPER project is based on the availability of an Internet-based prediction and decision support system based on real scientific data. Recommendations are made according to the rules of evidence-based medicine. They evolve according to medical knowledge in order to provide the patient with the best recommendation for treatment available at that time. The physician can optionally keep a record of the information obtained for the patient (by creating a patient file) or can print the results and recommendations.

The individualized probabilities of death by cause can be of interest to both physician and patient. Although he knows the risk factors concerning his patient, physicians make subjective and biased estimations. This tool allows physicians to target and quantify the major risks of death for the patient. Knowing these risks, the physician can suggest preventive measures at the appropriate time that are adapted to the patient. The physician can also use the statistics to increase the patient’s awareness of the usefulness of these preventive measures, and use simulations to demonstrate the expected results.
Patients requesting medical information concerning themselves obtain personalized answers to their questions. Furthermore, probabilities of death can be more readily understood than relative risks. In any case, the preceding examples demonstrate that the probabilities of death at age \( x + n \), based on the patient’s sex and region, may be low when we know that the relative risk for developing this pathology is high. Phillips K.-A. and coll. demonstrated this for breast cancer (16). This is not self-contradictory and can be explained by the fact that risks of death are competitive, i.e. the patient may die from another cause. The information provided by probabilities of death or the relative risk of death is not exclusive. It appears preferable to give the results of these indicators so that the patient understands his major risks of death, and so that the physician can instigate preventive health care measures.

For certain pathologies, these preventive measures are clearly demonstrated and defined as medical guidelines. Physicians can receive the information they contain through various media. A study in England concerning physicians in the Cambridge and Huntingdon (17) sector identified 855 different recommendations in a paper report (60% were produced locally and 40% nationally). Although they are not all dated, the number of recommendations produced increased exponentially since 1989 (8 in 1990, 73 in 1995, 138 in 1996 and 57 for the first quarter of 1997).

It is useful to question the impact of all these recommendations on medical practice as well as on the level of scientific proof for their contents. A meta-analysis (18) evaluating the impact of printed pedagogical materials on the behavior of physicians and the treatment of patients demonstrated the inefficiency of this type of communication. Vissers and coll. (19) tried to demonstrate that a semi-active computer system had an impact on how physicians questioned their own therapeutic strategies. Another meta-analysis (20) showed that computerized checklists were effective for vaccination reminders, for reducing cardiovascular risk, and for screening for certain cancers (breast and colorectal).

Obviously the choice of the communication medium has a major influence on the results. Using a computer to present recommendations for treating a given patient according to risks objectively determined for that patient can be a useful tool for physicians in their daily practice. The evaluation of this decision-support tool in daily practice mainly concerns two areas: the ergonomics of the site, and the receptivity and pertinence of the information provided.

**CONCLUSION**

The personalized prediction of risk of death is useful for pathologies for which preventive measures are available. The indicator selected to estimate this risk must be easy to understand and interpret for both physicians and patients. We believe that using conditional probabilities of death fulfills these conditions. Finally, the generalized availability of computers in doctors’ offices provides a tool for making personalized predictions of the risk of mortality and making recommendations to patients based on those risks.

**ACKNOWLEDGEMENTS**

The authors would like to thank I. Colombet, J.C. Dufour, H. Dréau, L. Falco, D. Fieschi, M. Joubert, E. Michel, D. Sauquet, for their helpful comments. ESPER project is supported by the CANAM (the French National Health Insurance Fund for the Self-employed). ESPER project members: J.C. Dufour, D. Fieschi, M. Fieschi, R. Giorgi, J. Gouvernet, M. Joubert (Laboratoire d’Enseignement et de Recherche sur le Traitement de l’Information Médicale, Faculté de Médecine de Marseille, France), G. Chatellier, I. Colombet, P. Degoulet, H. Dréau, L. Falco, D. Sauquet (Service d’Informatique Médicale, Hôpital Broussais, Paris, France), E. Michel,
REFERENCES