A novel method to evaluate large amounts of data on Chronic Neurological Diseases and it's consequences.

MUSTAFA A. KHAN, M.D.
(10-07-2013)

Thanks to the painstaking work by large number of people, there is now available an enormous amount of information on various chronic neurological diseases. This information is usually presented as a 2-dimensions coordinate-time mapping consisting of a variable x v/s t, where 'x' is the variable and 't' is the time. Another way is to compare one variable 'x_1' against another variable 'x_2'. One of the great limitations of these ways of looking at a disease is that from the limited information obtainable from 2-dimensional mapping contradictory results about the entire disease itself can be obtained. It is not uncommon to see conclusions, which are supposed to be applicable to the entire disease, reached from x v/s t contradict conclusions reached from x_1 v/s x_2. A method to look at the entire data with as many variables as possible at once would be much better and less likely to give results that are likely to be unrealistic or applicable to the entire disease.

The method I propose is to take as many variables ${}^{n}_{D}x$ about any disease 'D' and map them on a (n+1) dimensional space with time as one of the variables. Time is a very important variable to include since one is dealing with a chronic condition. Hence for any patient 'i' one can map all the ${}^{n}_{D}x_i$ variables known against 't'. Using this information, a function ${}_{D}f_i({}^{n}_{D}x_i, t)$ can be obtained for each patient 'i'. From the collection of these ${}_{D}f_i({}^{n}_{D}x_i, t)$, a "generic" function

$_{D}F\ (^{n}_{D}x\ ,t)$ can be derived that is applicable to the entire disease 'D'. This function will be the most accurate and realistic description of the disease 'D' given the variables $^{n}_{D}x$. With time, as more information is found, i.e. more $_{D}x$ are found, the function $_{D}F$ can be further fine tuned. This function $_{D}F$ I like to call the Khan (or K) function for the disease 'D'.

For the purposes of illustrating the power and utility of this method, I like to apply it to one of my favorite chronic neurological disease, the relapsing-remitting multiple sclerosis (RRMS). It will be immediately clear that this application to one disease is not unique, but is fully general and applicable to any chronic disease. In the case of RRMS, the $^{n}_{R}x$ will refer to the EDSS, MSFS, number of T lesions supra-tentorially, number of T lesions infra-tentorially, number of T lesions within the spinal cord and so forth. The greater the number of variables considered at once the better, which is the opposite of the other methods, such as the 2-dimensional method. Then for each patient 'i' one obtains $_{R}f_{i}(^{n}_{R}x_{i},t)$ and from these we get the K function for RRMS, $_{R}F\ (^{n}_{R}x\ ,t)$. This then gives the full description of RRMS as is currently known. From this one can obtain certain very useful conclusions or consequences.

The first one is that one may find that the $_{R}F$ is not unique. It may be that the spread of $_{R}f_{i}$ is such that one has to have several $_{R}F_{i}$ suggesting several sub-types of RRMS (i.e. RRMS1, RRMS2, RRMS3,.....). This would not have been possible to find without the use of this methodology. Finding sub-types of RRMS

further has great consequences on the research and treatment of RRMS.

The second consequence is that when a patient 'i' is first given the diagnosis of RRMS (i.e at time t_o), one can predict what the ${}^{n}_{R}x_i$ for this patient will be at any other time 't'. For this we will need a "closeness" (or "distance") quantity C_t which defines how "close" this patient is to ${}_{R}F$. For the present I would recommend to define $C_t = \sqrt{\sum_{t} ({}^{n}_{R}x - {}^{n}_{R}x_i)^2}$ at time 't'. Knowing C_{t_o}, one can extrapolate or obtain ${}_{R}f_i$ for this patient by assuming that the ${}_{R}f_i$ remains parallel to ${}_{R}F$ and hence obtain ${}^{n}_{R}x_i$ for any future time 't'. The assumption of ${}_{R}f_i$ being parallel to ${}_{R}F$ is not unreasonable given that ${}_{R}F$ defines RRMS. This will help the patient and the physician what course the disease will take in this particular patient, which will be extremely helpful in the discussion that usually follows after a diagnosis of RRMS is made.

The third conclusion is that for any treatment 'T' for RRMS one can obtain a K function ${}_{R}F_T$ from the longitudinal data of the treatment clinical trial. By comparing the C_t between ${}_{R}F$ and ${}_{R}F_T$ we can immediately stratify the various treatments. An immediate corollary of this stratification is that by comparing the C_t between ${}_{R}f_i$ and ${}_{R}F_T$ one can decide at 't' which treatment will be most appropriate for the patient 'i'. Another corollary of this stratification is the lack of any need to do head-to-head clinical trials of the various treatments which tend to be both long and quite expensive.

Finally, by this time it should be quite clear that this method is not specific for RRMS only, but can be applied to any

chronic neurological disease and even to chronic non-neurological diseases. Also, this methodology would have been nearly impossible to implement before the modern computers and computer graphics. But with the current computer technology available this methodology is quite easy to implement and thereby put in our hands a powerful and also a very useful tool to grasp, visualize and finally help in understanding any chronic disease easily, accurately and realistically in a way that had not been possible before.