Risk factor analysis of cancer types, BRCA1 and BRCA2 using the mathematical models: a review

Abstract

In order to determine the exact risk of cancer resulting from a mutation, genetic testing for the breast and ovarian cancer susceptibility genes BRCA1 and BRCA2 can help the clinical management of individuals with family history (FH) of the disease, by identifying individuals at highest risk. Those individuals can then be offered screening from an earlier age, prophylactic mastectomy or oophorectomy or potentially chemoprevention. It is important that genetic testing is targeted towards the individuals who are most likely to prove positive. Many of the models have been developed using the prior data to identify the genes responsible for the cancer at different stages. In this paper it is made to review the various methods followed to understand the disease causing genes, data analysis for the effective data modelling and the risk factors of the cancer. Name of the models have been discussed in the current publication with the advantages and limitations of each model.

Keywords: BRCA1, BRCA2, genetic models, mutation, cancer

Introduction

BRCA1 and BRCA2 associated hereditary breast and ovarian cancer syndrome (HBOC) is characterized by an increased risk for female and male breast cancer, ovarian cancer which also include fallopian tube and primary peritoneal cancers. To a lesser extent other cancers viz., prostate cancer, pancreatic cancer and melanoma are associated primarily in the individuals with a BRCA2 pathogenic variant. Women with harmful mutations in either BRCA1 or BRCA2 have a risk of breast cancer that is about five times the normal risk, and a risk of ovarian cancer that is about ten to thirty times the normal. Generally, having a high-risk mutation does not guarantee that the woman will develop any type of cancer, or imply that any cancer that appears was actually caused by the mutation. Screening for cancer causing mutations and predicting the genes has become widespread, and more than 750 protein-truncating mutations in these genes having been identified. Some women found to carry such mutations generally tend to undergo prophylactic mastectomy and/or oophorectomy, because their cancer risk is extremely high. The following are the few models discussed based on the various studies carried out and the data produced.

Claus model

The Claus model was developed from data that was obtained from the Cancer and Steroid Hormone Study, a multicenter, population-based, case-control study conducted by the CDC. The data set consists of 4,730 histologically confirmed breast cancer cases aged 20-54 years and of 4,688 controls. Cases were considered in Atlanta, Detroit, San Francisco, Seattle, the four urban counties of Utah, and the states of Connecticut, Iowa, and New Mexico. Cases described in five year categories of age and corresponding geographical region, without any previous history of breast cancer were used. The model also used the family history of only first degree white relatives-colored cases were not included. The original model proved the existence of a rare autosomal dominant allele for increased susceptibility to breast cancer, one that conferred an even higher risk at young ages. The overall susceptibility to breast cancer by those women who carried the allele was around 92%. Women who were non-carriers were proved to have the reported lifetime risk for the general US population is 10%. Hence, lifetime risk tables for most combinations of affected first-degree and second-degree relatives were published in order to enable easier use in clinical settings.

Later versions of this model also include analysis regarding increased risk of women to breast cancer if a relative is affected with ovarian cancer. Use of the Claus model is limited to specific combinations of affected relatives (eg, two first-degree relatives, mother-maternal aunt, etc). These tables do not include the commonly encountered mother-maternal grandmother pair. In this situation, a combination with the same degree of relatedness (eg, mother-maternal aunt, instead of mother-maternal grandmother) is the best approximation. Hence similar such combinations can be used to determine the risk factor. The Claus model has three major drawbacks that affect its regular use:

a. First, the model does not include any non-hereditary risk factors (eg, hormonal or reproductive factors).

b. Second, the Claus lifetime risk tables reflect risks for North American women in the 1980s, which are known to be lower than the current incidence of breast cancer in North America and in most of Europe.

c. Third, the published tables and computerized versions of the model appear to give different results: The Tables 1–4 give consistently higher risk figures than the computer model. An explanation for this may be that while the tables make no adjustments for unaffected relatives, the computerized version is able to reduce the likelihood of a germline mutation in an individual with an increasing number of unaffected women.
The Claus Extended Model or the Jonker Model is a genetic model to predict breast cancer risk based on the family history of breast and ovarian cancers. This model is an extension of the Claus model combined with the BRCAPRO model. It seeks to explain familial clustering of breast and ovarian cancers by using three genes: BRCA1, BRCA2, and a hypothetical third gene called BRCAu. The hypothetical gene was modelled to explain all familial clustering of breast cancer that was not accounted for by the BRCA1 and BRCA2 genes. The model parameters were estimated using published population incidence and relative risk estimates. The Jonker model does not include data on personal risk factors for breast cancer. The Claus extended model has been criticized for two major limitations:

a) Inability to estimate risk in women with complex family histories.
b) Its validation in individual families rather than in an independent series.

**Couch model**

The Couch Model considers family history of breast and ovarian cancer. It is a logistical regression model that uses univariate and multivariate analytical methods to predict the probability of a BRCA mutation, with respect to various factors considered. These include unilateral breast cancer, bilateral breast cancer, ovarian cancer, combined breast and ovarian cancer, the number of women at risk in a family (those over 20 years of age), the average age at diagnosis of breast cancer, the average age at diagnosis of ovarian cancer, and Ashkenazi Jewish ancestry. It is particularly useful to distinguish between risk factors for Ashkenazi Jew families and non-Ashkenazi Jew families.6,7

DNA samples from 263 unrelated women with breast cancer were analyzed, from 1993-1995. Of these, 169 women had been referred to breast-cancer clinics because of a familial risk factor for breast cancer, and reported 1 to 11 cases of breast cancer per family. The remaining 94 women were identified in general-oncology practices because of the diagnosis of breast cancer before the age of 40; some of these women also had a family history of breast cancer. Twenty-five families reported Ashkenazi Jewish ancestry. Of the 169 women with breast cancer and a familial risk factor, 27 (16 percent) had a BRCA1 mutation. Mutations were identified in 13% of the women in whom breast cancer was diagnosed before the age of 40. Mutations were also identified in 7% with members with breast cancer without ovarian...
cancer, 40% with members with both breast and ovarian cancer, and 67% with a single member with both breast and ovarian cancer. The frequency of BRCA1 mutations among Ashkenazi Jewish women was 26%. The Couch studies proved that the presence of breast cancer alone (without ovarian cancer in the family) is infrequently associated with mutations in the coding region of BRCA1 or BRCA2. Thus, there remain a large number of families in which breast cancer may be associated with mutations in noncoding regions of BRCA1 or other susceptibility genes. The risk calculated is the family’s probability of a BRCA1 mutation and applies to all affected (diagnosed) family members. The probability of a mutation in an unaffected first-degree relative is half of the family’s probability for carrying a mutation. Ex. If the Couch mutation probability is 10% for a family, the daughter of the affected individual in the family is 5% chance, as she has a 50% chance of inheriting the mutated allele from her mother (Table 5).

Table 5 Probability of detecting BRCA1 mutations in families

<table>
<thead>
<tr>
<th>Families with breast cancer and ovarian cancer</th>
<th>Ashkenazi Jewish families with breast cancer and ovarian cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average age at diagnosis of breast cancer</td>
<td>Average age at diagnosis of breast cancer</td>
</tr>
<tr>
<td>Predicted probability (%) (95% CI)</td>
<td>Predicted probability (%) (95% CI)</td>
</tr>
<tr>
<td>&lt; 35 years</td>
<td>&lt; 35 years</td>
</tr>
<tr>
<td>55.0% (27.2%-80.0%)</td>
<td>84.20%</td>
</tr>
<tr>
<td>35-39 years</td>
<td>35-39 years</td>
</tr>
<tr>
<td>43.5% (22.4%-67.2%)</td>
<td>77.1% (40.1%-94.4%)</td>
</tr>
<tr>
<td>40-44 years</td>
<td>40-44 years</td>
</tr>
<tr>
<td>32.7% (17.0%-53.5%)</td>
<td>67.90%</td>
</tr>
<tr>
<td>45-49 years</td>
<td>45-49 years</td>
</tr>
<tr>
<td>23.4% (11.4%-42.1%)</td>
<td>57.2% (24.9%-84.3%)</td>
</tr>
<tr>
<td>50-54 years</td>
<td>50-54 years</td>
</tr>
<tr>
<td>16.2% (6.7%-34.2%)</td>
<td>45.70%</td>
</tr>
<tr>
<td>55-59 years</td>
<td>55-59 years</td>
</tr>
<tr>
<td>10.8% (3.5%-28.8%)</td>
<td>34.7% (10.8%-70.0%)</td>
</tr>
<tr>
<td>&gt; 59 years</td>
<td>&gt; 59 years</td>
</tr>
<tr>
<td>7.1% (1.7%-24.5%)</td>
<td>25.10%</td>
</tr>
</tbody>
</table>

*Based on families with an average of 4.0 breast cancers per family

CI, Confidence interval

Limitations

This model provides risk estimate to be of only BRCA1. In clinical practice this is modified by multiplying the couch probability factor by 1.33 to include BRCA2. It does not consider other types of cancer associated with the BRCA gene, aside from breast and ovarian cancer. It does not include male breast cancer, but only Caucasian women were included in the study. The updated version of this model is PENNII. The output of PENNII model includes the consultant’s probability of having a deleterious BRCA1 or BRCA2 gene mutation and the probability of a deleterious gene mutation in an affected family member, i.e., one with breast or ovarian cancer. If the proband has a diagnosis of breast or ovarian cancer, the risk prediction will be the same for individual and family. If the consultant is unaffected, then the family risk reflects the probability in an affected member and the consultant risk is reduced according by Mendelian logic to whether they are first or second degree relatives of the affected member. The details of this model’s development have not yet been published.

Myriad model

The Myriad mutation prevalence tables, of Myriad Genetics Laboratories are derived from Frank Model. This method, developed by Frank et al, involved testing 10,000 individuals over a 3 year period to identify Mutations anywhere in the BRCA1 and BRCA2 genes. Three specific Ashkenazi Jewish founder mutations include the factors like personal and family history of cancer, ancestry, invasive versus noninvasive breast neoplasia and sex (both males and females were considered) were included. The largest proportions of individuals included in the study were Northern/Western Europe [41%]. Other ancestries of the individuals tested (in descending order) were Ashkenazi [30%], Central/Eastern European [10%], Latin American/Caribbean [2.3%], Native American [2.2%], African [1.6%], Asian [1.1%], and Near Eastern/Middle Eastern [0.9%]. A total of 1,775 individuals (18%) did not specify ancestry.

Overall, 1,731 deleterious mutations in BRCA1 and BRCA2 were identified in the 10,000 individuals tested. Other types of mutation like frameshift, missense and nonsense mutations were also identified and classified. (53%) and (43%) of the distinct mutations occurred in the largest exons (exon 11) of each BRCA1 and BRCA2. The prevalence of mutations generally increased with earlier age of onset of breast cancer. In contrast, there was no consistent increase in the prevalence of mutations seen among women whose ovarian cancer was diagnosed before age 50 years compared with those diagnosed at a later age. A unique characteristic of this study is that it includes the largest series of men with breast cancer analyzed for BRCA1 and BRCA2 to date. BRCA1 and BRCA2 mutations were identified in 28% of men with breast cancer, of which a substantial proportion occurred in BRCA1. The Myriad Mutation Prevalence tables are updated frequently and made to be used in a clinical setting. The tables (where are the tables??) represent observations of deleterious mutations by Myriad Genetic Laboratories through its clinical testing service.

Advantages

These tables are categorized by Ashkenazi Jew versus Non-Ashkenazi Jew ancestry. They are updated frequently and are easily accessible.

Disadvantages

These tables don’t take into account specific age of onset, number of affected relatives, bilateral breast cancer, and other BRCA associated cancers. It is dependant on family history provided by health forms, which are subject to error. Hence the probabilities may be overestimated or underestimated. There is no distinction made between first-degree and second-degree relatives, or maternal versus paternal affected relatives.

BOADICEA model

“BOADICEA” stands for Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm. This model is a general model of breast cancer susceptibility that provides estimates of cancer risks and carrier probabilities to women with a FH of breast/ovarian cancer and was developed using complex segregation analysis of breast and ovarian cancer based on a series of combination of families identified through population-based studies of breast cancer and families with multiple affected individuals who were screened for BRCA1 and BRCA2 mutations. 16-17 BRCA1 and BRCA2 mutations have similar population frequencies in BOADICEA, with BRCA2 being slightly more predominant. Refinement of the BOADICEA model by is in progress by refitting the model using additional family data as several improvements can be incorporated into the present model. An obvious deficiency of the current model is that the genotype-specific incidence rates are computed for broad age categories and putting together pathological information into the model could improve the accuracy of carrier prediction. In principle, the model can also be extended to incorporate non genetic risk factors, like that of age at menopause, parity and breastfeeding. 15 Although, this requires precise estimates of these effects in BRCA1 and BRCA2 carriers. The model can be extended to account for the effects of other susceptibility genes other than BRCA as well. The polygenic component in BOADICEA represents the combined effects of low-penetrance BRCA1 and BRCA2. The latest version of the model was based on 2785 families, of which 537 segregated BRCA1 and/or BRCA2 mutations.

Assumptions made by the existing model include a fixed set of duration incidences applied to all cohorts, when breast cancer incidences have been increasing over time. Moreover, the incidences were assumed to change in 5-year intervals when in reality they change smoothly with age. 14 It was also estimated the BRCA1 and BRCA2 breast and ovarian cancer risks, but these were based on a relatively small number of mutation-carrying families and were therefore imprecise. 14 The model took into account only the incidence of a first breast or a first ovarian cancer and the risks of second or subsequent cancers were ignored. 14 BOADICEA assumes that the residual familial clustering of breast cancer is explained by a polygenic component (a large number of genes each of small effect on risk) with a variance that shows a linear decrease with age. 14

BOADICEA has been validated in a large series of families from UK genetics clinics. Cancer incidences from recent times (1992–2010) are available and have been included in the updated version of BOADICEA. The BOADICEA has recently been extended to include breast tumour pathology information, where breast cancer subtypes are treated as distinct disease end points. In particular, oestrogen receptor status, triple-negative status (oestrogen, progesterone and HER2 negative) and expression of basal markers (CK5/6 and CK14) are taken into account additionally. 17 BOADICEA, the world’s first polygenic breast cancer risk model, remains the only one available to the global healthcare community. It was the first breast cancer risk model to incorporate pathology data and cancer incidences specific to populations in the risk calculations. Similarly, the BOADICEA Web Application was the first breast cancer risk model interface to be enforced on the Web, to include pedigree building for families of arbitrary size and structure beyond degree 2, to check that all family members within an uploaded pedigree have familial connection to the index, to include Web-based batch processing, and also to generate processing report PDFs. 17

BRCA PRO model

BRCA PRO is a program, developed at the Institute of Statistics and Decision Sciences by G Parmigiani, DA Berry and O Aguilar, Duke University, USA, which is used to calculate the probability of one particular family member carrying a germ line mutation of the BRCA1 and BRCA2 genes. 15 Though the method is applicable to both women and men, the Proband is usually female. BRCA PRO model has been widely used and extensively evaluated, mostly in non-Hispanic white populations. This model was developed on the basis of mutation rates and penetrance observed mostly in women of European ancestry and Ashkenazi Jewish. The model makes use of Bayes’ theorem to calculate the probability of an individual carrying a mutation in the BRCA1 or BRCA2 gene based on the family history of breast and/or ovarian cancer. 19 An intermediary step in the calculation of this probability includes the calculation of a likelihood ratio for each individual in the family. These feasibility ratios are based on the probability that a specific cancer history would be observed whether the individual (With breast/ovarian cancer or without) were a carrier of the mutation or not. The final probability is calculated using the frequency of the alleles of BRCA gene mutations in the relevant population and breast and ovarian cancer incidence rates for BRCA gene mutation carriers and non-carriers. The calculation of the final probability incorporates age and cancer information for all first- and second-degree relatives on the basis of Mendelian inheritance of an autosomal dominant gene. 19

Advantages

It makes use of two forms of input data: proband-specific and population-based. Proband-specific data include the proband’s first- and second-degree family history of breast and ovarian cancer. 19 It uses the following information for each proband and each first- and second-degree relative (including those not affected by cancer)-sex, exact relationship to the proband, any diagnosis of breast cancer, second primary breast cancer, and ovarian cancer, and age at each diagnosis, current age or age at death. Whether the proband is Ashkenazi Jewish (AJ) is also accounted for in the model. It is especially accurate in predicting testing results when the carrier probability is less than 70%. 18

Disadvantages

But no provision is made for inbreeding, monoyzogous twinning, or for other relatives. If part of the family history information was missing, a conservative approach was taken. BRCA PRO may meet its goal and yet show up poorly when compared with testing results. Sometimes individuals who carry a mutation and whose BRCA PRO probabilities are seen to be close to 100% will test negative. 18

The clinical application and importance of BRCA PRO depends on the individual’s carrier probability, but it also depends greatly on the individual’s attitudes. If she does not make changes in her life as a result of learning her carrier probability, then the information has no value. Nevertheless, some women will make decisions on the basis of the level of risk associated as in, a probability less than 5% might not prompt any changes in lifestyle or health care decisions, whereas a probability in the range of 5% to 10% might prompt the individual to seek genetic counselling and a high probability might require a prophylactic surgery. An individual with a 10% probability might base subsequent health care treatment options on the outcome of genetic testing, opting for a surgical solution if the test finds a deleterious mutation, and avoiding surgery otherwise. The benefit of
testing depends on the individual’s attitudes and carrier probability, and for some it may result in additional life expectancy. For individuals who present themselves at genetic counselling clinics, BRCAPRO provides an accurate assessment of probability of carrying a deleterious mutation of BRCA1 and BRCA2. An accurate estimate of the probability of testing positive is found by multiplying the probability by the sensitivity of the test method used. Moreover, even though an individual’s carrier probability is affected by the estimates of penetrance and prevalence used, the accuracy of BRCAPRO is not dependent on having accurate estimates of these quantities. The performance of the model was seen to be the best in Hispanics and the worst in African Americans and was seen to be intermediate for the other minority groups, mainly Asian-Americans and Native Americans. However, the difference between ethnic groups was not statistically significant. Data suggest that BRCAPRO performs reasonably well and is applicable to all minorities but the sample size that was taken may not suffice for subgroup comparisons. Real differences possibly exist in BRCAPRO performance across ethnic minority groups, because the same penetrance and non-Ashkenazi white allele frequencies were used in the calculation. These genetic parameters likely vary across ethnicity. Better BRCAPRO performance in Hispanics may be because Hispanics are genetically closer to other white populations. This model has been found to be superior to other risk assessment models and comparable to the results from experienced genetic counsellors. BRCAPRO was also found to be a useful tool to determine the need for testing in male breast cancer patients with no significant family history in a study.

The latest versions of BRCAPRO model developed include version 5.0 and 6.0:

a. Version 5.0: The older BRCAPRO model requires computer access and is not easily accessible during genetic counseling, being rather cumbersome for data entry, whereas this version of the model represents a helpful tool in management of MBC cases with a negative FH. Indeed, when a cut-off value x 10% was considered, BRCAPRO 5.0 was seen to have fully succeeded in identifying mutation carriers with no FH in the MBC series.

b. Version 6.0: This model uses a Bayesian model, which takes into account data of proband, affected and unaffected family member and the population. The variables included in the model were a personal history of cancer, first- and second-degree family history, age at cancer diagnosis, current age or age at death, ethnicity, Ashkenazi Jewish ancestry and history of risk-reducing surgery. The population-based data in BRCAPRO includes mutation prevalence and disease penetrance.

Gail model

The Gail Model was developed by Dr. Mitchell Gail and his colleagues with the Biostatistics Branch of NCI’s Division of Cancer Epidemiology and Genetics. It is a statistical breast cancer risk assessment algorithm which. It was developed following a huge screening study of 280,000 women between the ages 35 and 74 years. The Gail model has shown to be a rational tool for evaluating breast cancer risk in white women, and researchers have eventually supplemented the model to provide accurate risk assessments for African American, Hispanic, and Asian women. The Gail model looked at women’s personal medical history, familial history, and reproductive history. Specifically, it considered the number of foregoing biopsies and the presence of any kind of atypical hyperplasia on them. It also took into consideration the age of start of menstruation and age at first live birth of a child, and also the incidence of breast cancer among first degree female relatives. These variables were then adjusted according to increasing age brackets and associated higher risk for older women.

Limitations

It does not include all the risk factors for breast cancer development. Like, certain established risk factors such as increased estrogen levels, adiposity, and the use of hormone replacement therapy, a sedentary lifestyle, and alcohol abuse are absent. This model has not been validated in all population subgroups, such as women with genetic predispositions for breast cancer or younger women. It does not take into account ethnic or racial difference, genetic variants of BRCA, tamoxifen use and excluded who had already had a confirmed diagnosis of either ductal or lobular breast carcinoma in situ. According to some researchers, there is an indication that this model underestimates the risk of breast cancer development in women with a typical breast hyperplasia already existing. Gail model cannot be used if the patient’s family history is the main source of risk. It gives the average risk for a group of women with similar risk factors and hence, cannot predict whether an individual is prone to breast cancer risk. So, it is not clear what this risk means for any one woman. Even with it’s current limitations, this breast cancer risk assessment tool can be very useful. However, it is not considered the best with respect to finding out an individual’s risk for cancer.

Tyrie-Cuzick or international breast cancer intervention study model

This relatively new risk assessment model uses Bayes theorem to predict the risk of BRCA mutations. It involves a two locus genetic model, with one of the genes being either BRCA1/2, or the other being a hypothetical gene that causes only breast cancer, whose theoretical prevalence is 11% in a population where the lifetime risk is 24%. This helps to account for unknown genes whose mutations may result in cancer, if a family history doesn’t exist. These are called the Bayesian variables. The factors that it incorporates include family history, endogenous hormonal factors, benign disease, risk factors such as age and body mass index and genetic factors (including BRCA, all into a single statistical model. The Tyrie-Cuzick model in its independent studies has shown that it is considerably accurate as compared to other risk assessment models. The model has been incorporated into a computer programme, and with the help of the IBIS software it gives a personalised risk estimate. The program assumes that there is a gene that is prone to breast cancer in addition to that of the BRCA genes. The Woman’s family history would help to predict the likelihood of her carrying the adverse gene, which can help to calculate the likelihood of her developing breast cancer. This is the only model that accounts for both personal and extensive family history risk factors.

Risk factors that are taken into consideration are:

i. Advancing age
ii. Older age at first live birth (>30y)
iii. Younger age at menarche (<12y)
iv. Older age at menopause (>55y)
v. Hormone replacement therapy
vi. Increased number of breast biopsies
vii. Increased breast density
viii. Birth control
ix. Radiation exposure  
xi. Decreased physical activity  

xii. Alcohol use

**Tyre-Cuzick model to calculate lifetime risk for breast cancer**

a) Step 1: Download. Free software can be downloaded only for PC computers only at http://www.ems-trials.org/risk evaluator/.

b) Step 2: Begin. Click “evaluate” and then enter the name and identification number for the consultand. A consultand is the person for whom a risk calculation has to be done. This person is identified on the pedigree with an arrow.

c) Step 3: Personal factors have to be entered. Enter the consultand’s age, age at menarche, and height and weight measurements. Next, enter parity and then history of any breast disease of the consultand. If the consultand has never been diagnosed with any breast abnormalities, no benign disease box has to be marked. If she has been diagnosed with hyperplasia, unknown benign disease, a typical hyperplasia, or ovarian cancer, mark in those boxes which are provided. Next, enter the consultand’s menopausal status. If she had surgical menopause, enter the age she was at the time of surgery in the “age at menopause” box. The final series of questions in the personal factors section are related to hormone therapy (HRT). If the woman is currently using HRT, a box will ask for intended length of use. Finally, if the patient is of Ashkenazi Jewish heritage (Eastern European Jewish), mark the corresponding box because this population is at increased risk for carrying mutations in the BRCA1 and BRCA2 genes. The final task in the personal information section is to check the competing mortality box. With this box marked, the calculations will include the possibility that someone might suffer from other causes even before a breast cancer could manifest. Leaving this box unchecked will result in a higher lifetime risk calculation.

d) Step 4: Family history; this model is dependent on entering the family history to provide the best results. As each step is completed, the pedigree drawing will expand. An arrow on the pedigree indicates which family member is the consultand. The order in which this section is completed matters. The overall objective with the family history section is to check which would influence the woman’s risk for breast cancer. For sisters, aunts, and daughters, the number of women in each category needs to be entered before their cancer history. The program will accept cancer history for up to 5 relatives in each of the categories.

**Limitations**

This model accounts only for hereditary breast and ovarian cancer (HBOC). HBOC is caused by a mutation in the BRCA1 and BRCA2 genes. But, there are 27 other hereditary breast cancer syndromes that exist which can increase a woman’s risk for developing breast cancer. The different type of syndromes and their affected genes include TP53 mutations in Li-Fraumeni syndrome, PTEN mutations in Cowden syndrome and STK11 mutations in Peutz-Jeghers syndrome. If a woman presents with any of these syndromes, a genetic counsellor should be consulted, as the Tyre-Cuzick model cannot account for other conditions other than the information regarding the family history. According to the NCCN, women should be referred to a genetic counsellor if they have a family history of more than 3 of the following: “breast, pancreatic cancer, prostate cancer, melanoma, sarcoma and many more.”

**Kaplan-Meier estimation**

Kaplan-Meier estimate is one of the best options to measure the fraction of subjects living for a certain amount of time even after the treatment. In clinical trials, the effect of improvement of the medical condition is assessed by measuring the number of subjects survived or saved over a period of time. It is the simplest way for analysing the survival over time in spite of all the difficulties faced. It involves computing probabilities of occurrence of an event at a certain point of time and multiplying these successive probabilities by any earlier computed probabilities to give the final estimate. This can also be calculated for two groups of subjects. The Kaplan-Meier survival curve is defined as the probability of surviving in a given length of time while considering time in small intervals. There are three assumptions used in this analysis:

a. *Firstly*, we assume that the patients who have been examined at the same time have the same survival prospects as those who continue to be followed.

b. *Secondly*, we assume that the Probabilities of Survival are the same even if the subjects have been listed either early or late in the study.

c. *Thirdly*, we assume that the event happens at the time specified.

This creates a problem in certain conditions about when the event would have been conducted at a regular examination. All we know is that the event happened between two examinations. Estimated survival rates can be calculated more accurately when it is carried out in shorter time intervals. The Kaplan-Meier estimate is also called as “product limit estimate”. Since, this method involves multiplying of the successive probabilities with that of the earlier computed probabilities

\[
S(t) = \frac{\text{Number of subjects living at the start} - \text{Number of subjects died}}{\text{Number of subjects living at the start}}
\]

The Kaplan Meier plotter is capable to assess the effect of 54,675 genes on survival using 10,293 cancer samples. These include 5,143 breasts, 1,648 ovarian, 2,437 lung and 1,065 gastric cancer patients with a mean follow-up of 69/40/49/33 months. Primary purpose of the tool is a meta-analysis based biomarker assessment.

**How does it work?**

The database is controlled by a PostgreSQL server, which combines gene expression and clinical data simultaneously. To predict the value of a particular gene, the patient samples are split into two groups by various quantile expressions proposed by the biomarker. All the biomarkers can be analysed simultaneously. The two patient cohorts are compared by a Kaplan-Meier survival plot. Multiple genes can also be used.

For Kaplan-Meier estimator to be generated two things are a must-

i. The status of the last observation at the event( examination) that has occurred and

ii. The time taken for the event to occur.
Hosmer and Lemeshow describe the Kaplan–Meier estimator as follows. “This estimator incorporates information from all of the observations available, both uncensored and censored, by considering survival to any point in time as a series of steps.

**Improvements**

A reliable prognostic method uses condition monitoring (CM) which can deal with any number of indicators and need not be specific about selecting the most important ones as many other methods prefer. It does not depend on any thresholding strategies to separate normal or abnormal values of condition indicators. It uses both age and CM data as inputs to estimate the remaining useful life (RUL). The key idea behind this method is based on Kaplan–Meier as a time driven estimation technique and logical analysis of data as an event-driven diagnostic technique to reflect the effect of operating conditions. The performance of the estimated RUL is measured in terms of difference between the actual and predicted value.27

**Advantages**

This method can be taken into account even when right censoring occurs, that is when a patient withdraws from the study or is alive without the event occurrence. On the plot, small vertical tick-marks indicate individual patients whose survival times have been right-censored.

**Limitations**

The Kaplan–Meier estimator may be useful to examine RECORD rates, the probability of death, and the effectiveness of treatment but it’s only limited to estimate survival given by observations of an event, parametric Survival models.

**Cox model**

This model involves survival analysis also known as time-to-event data analysis, is widely used in oncology since we are usually interested in studying the delay from the time a person has been diagnosed with cancer or started with their treatment. It is a semi-parametric survival model which relies only on fewer assumptions compared to parametric models (assumes finite set of parameters), which is the proportionality of the hazard function.28 The Cox proportional hazards (PH) model helps to describe the survival time as a function of multiple prognostic factors. The main assumption in this model are that the predictors act multiplicatively on the hazard function but does not assume that the hazard function as a constant. This model helps to find out the percentage of risk whether the risk is decreased or increased. For interpretation of a Cox model the coefficients need to be examined for each explanatory variable. A positive regression coefficient having high positive values on a particular variable means that they are major hazards for the patient. Conversely, a negative regression coefficient with high values of that variable implies a better prognosis for patients. In most of the situations the true form of the hazard function is unknown or complex.

**Advantages**

It does not require any assumptions concerning the shape of a hidden survival distribution. In comparison with the equally common model of logistic regression, the hazards model makes it possible to consider survival time and occurrence of truncated variables.

**Acknowledgments**

None.

**Conflicts of interest**

Authors declare that there is no conflict of interest.

**References**

7. https://books.google.co.in/books?id=AaAmzCCQ-N0C&pg=PA11&lpg=PA11&dq=couch+model+brca1&source=bl&ots=xp7mG5gHuA&sig=48KHUG7BZ4QVfGZpA9Gw=onepage&q=couch%20model%20brca1&f=false
11. https://books.google.co.in/books?id=FQ5POnjhmC&pg=PA87&lpg=PA87&dq=couch+model+brca1&source=bl&ots=8XKXZm5AK&sig=8cU4PlGF-MY9quo- Wij=2jOuGk1=eneksuW=6r&ved=0ahUKEw-67o7RAhUSR4KHUG7BZ4Q6AEIQTAH#v=snippet&q=couch%20model%20brca1&f=false


