The Laboratory for Quantitative Medicine
Massachusetts General Hospital/Harvard Medical School
James Michaelson PhD

DATA RESOURCES AND APPLICATIONS
Databases on Patients,
Mathematics of Cancer Outcome, Screening and Treatment
Analysis of Medical Usage and Cost among Cancer Patients
Communication of Cancer Outcome (The CancerMath.net web calculators)

RESEARCH SUMMARY:
For the past decade, the work of my group has been concerned with building very large databases on patients, using this information to understand disease, and especially cancer, and its treatment, using mathematics to answer practical questions about health, and building web-based tools for communicating this information to patients and medical professionals (the CancerMath.net calculators).

DATABASES:
We have built a number of very large databases on patients. Last June, we were tasked by the MGH Cancer Center to build a database on all of the 173,301 Massachusetts General Hospital cancer patients, 167,814 of whom were diagnosed between 1968 and 2010. The database contains 559,921 pathology reports, 575,204 discharge reports, 10,938,444 encounter notes, 304,211 operative reports, 22,009,527 procedure notes, 9,159,232 radiology reports, ~1,700,000 aggregated medical bills, and ~250,000 images. The database contains all-cause survival information from the Social Security Administration Death Master File (which provides information on all deaths of persons issued social security numbers since 1937), and cause-of-death information from the Massachusetts Death Certificate Database (which contains international classification of disease cause of death information on 1,984,790 people who died in the state of Massachusetts between 1970 and 2008). The database is linked to the MGH SNAPSHOT gene sequence dataset, thus providing a great wealth of genetic data on a large number of patients.

As far as we are aware, in terms of the total mass of data, this database is the largest source of clinical information on cancer in the world.

Our group also maintains a great wealth of other data on patients, including the largest multi-institutional database on the breast cancer in the world, with data on more than 33,000 breast cancer patients (and all the varieties of information noted above). We have also built a similar multi-institutional data on melanoma patients. We also possess the Van Nuys Breast cancer database. We also have our own MS SQL Server re-build of the SEER national database with, for example ~325,000 breast cancer patients for who there is full tumor size and survival information. We also possess the SEER/Medicare dataset for breast cancer. Finally, we also have build a number of specialized databases, including our database on all MGH cancer patients who were treated by liver resection.

MEDICAL USAGE AND COST:
Over the past year, we have served on the MGH Cancer Center's lung cancer and colon cancer redesign committees, providing detailed and actionable data on medical use and cost for cancer patients. For example, our analysis of the lung cancer patients has revealed that the total cost of treating lung cancer patients at the MGH is ~$59 million/year, but that standardization of care has the potential of saving ~$30 million/year.

Studies are underway for generating data on treatment and cost for patients with many types of cancer.

A MATHEMATICAL APPROACH TO CANCER LETHALITY:
We have developed a mathematical framework for comprehending cancer lethality: the binary biological model of metastasis. This mathematics has proven to provide highly specific method for estimating the risk of death for individual breast carcinoma patients, with an accuracy of 1%. This mathematics also provide estimates of the risk of death for melanoma, renal cell carcinoma, sarcoma, and head and neck squamous cell carcinomas, and the applicability of this mathematics in analysis of survival for other cancers is under analysis. This math also provides a basis for estimating the risk of node positivity for breast carcinoma and melanoma patients, as well as providing insight into the nature of the events of spread that underlie cancer lethality, as well as providing a way to address a whole range of questions in oncology.

Our binary-biological mathematics has also formed the core of a computer simulation model of cancer screening, which has made it possible to derive biologically plausible and testable estimates of the reduction in cancer death that can be expected from screening various patients at various intervals. This work has been accompanied by a whole range of studies on the operational details of the usage of breast cancer screening.

Our binary-biological mathematics has been used to create a series of web-based CancerMath.net calculators, which provide patients with breast carcinoma, melanoma, and renal cell carcinoma with information on their likely outcomes. For breast carcinoma, we also provide a CancerMath web-calculator that shows the benefit that they can expect from the various adjuvant chemotherapy agents available to them. Analysis of the use of these CancerMath.net calculators has taught us that they are very widely used, being consulted by 1-in-5 USA breast carcinoma patients, 1-in-2 renal cell carcinoma patients, and a large number of melanoma patients. More details can be found below.
For the last decade, my research has concerned the development of a mathematical framework for predicting survival for cancer patients, together with the impact that various treatment choices will have on that outcome and the creation of very large databases on patients so that this mathematics can be accurate.

One of the applications of this work has been the creation of a series of Cancer Math.net web calculators, which appear to have become the most widely used decision aids used by cancer patients (www.CancerMath.net). These calculators are used by ~1-in-5 breast carcinoma patients, ~1-in-2 renal cell carcinoma patients, and large numbers of patients with other cancers. This has happened quite spontaneously, without publicizing these tools.

The goal of this work has been to provide patients and their physicians with highly accurate, disinterested, information on their survival expectations, together with information on the impact that can be expected from the various treatment options that are available to them.

This math also has applications in other fields, such as: generating accurate estimates of the benefit of cancer screening; generating accurate estimates of life expectancy for the insurance industry; generating accurate estimates for legal professionals of the harm caused (if any) in the detection and treatment of cancer.

We have also created a web-calculator, PreventiveMath.net (www.PreventiveMath.net), which provides individuals with a list of the class A US Preventive Services Task Force recommendations, prioritized by the benefit that they can expect, so that people can see the benefit, and choose those steps that will give them the greatest possible extension in life.

The current tools available at CancerMath.net website include:
1) Breast Cancer Outcome Calculator (Provides information on survival expectation [see below for definition*], at the time of diagnosis, assuming standard of care therapy)
2) Breast Cancer Therapy Calculator (Provides information on survival expectation*, and the impact which various adjuvant chemotherapy options can be expected to have on that outcome)
3) Breast Cancer Conditional Survival Calculator (Provides information on survival expectation*, assuming standard of care therapy, for patients who have remained disease free 2-15 years after diagnosis)
4) Breast Cancer Nodal Status Calculator (Provides information on the likelihood of cancer spread to the local lymph nodes)
5) Breast Cancer Nipple Involvement Calculator (Provides information to the surgeon on the likelihood that nipple sparing surgery can be carried out without leaving cancer behind)
6) Melanoma Outcome Calculator (Provides information on survival expectation*, at the time of diagnosis, assuming standard of care therapy) (Also includes Conditional Survival Information)
7) Renal Cell Carcinoma Outcome Calculator (Provides information on survival expectation*, at the time of diagnosis, assuming standard of care therapy) (Also includes Conditional Survival Information)
8) Colon Cancer Outcome Calculator (Provides information on survival expectation*, at the time of diagnosis, assuming standard of care therapy) (Also includes Conditional Survival Information)
9) Head & Neck Cancer Outcome Calculator (Provides information on survival expectation*, at the time of diagnosis, assuming standard of care therapy) (Also includes Conditional Survival Information)
10) Sarcoma Outcome Calculator (Provides information on survival expectation*, at the time of diagnosis, assuming standard of care therapy) (Also includes Conditional Survival Information)

Notes:
1) * survival expectation measures provided by the CancerMath calculators are: risk of death, for each of the first 15 years after diagnosis: 1) to cancer; 2) to causes of death other than cancer; 3) to all causes.
   Also provided is the life expectancy with cancer, life expectancy without cancer, and the reduction in life expectancy that is caused by cancer.
   For the therapy calculator, the impact of the various breast cancer adjuvant chemotherapy regimens on these measures are given.
2) The outcome calculators also provide Cancer Stage.
The effect of close or positive margin status in breast cancer in a mammography population.


43. Cady B, Michaelson JS Chung MA, The “Tipping Point” for breast cancer mortality decline has resulted from size reductions due to mammographic screening. *Annals of Surgical Oncology* in Press 2011


Laboratory of Quantitative Medicine Technical Reports
(Available at http://www.lifemath.net/cancer/about/techreports/index.php)
1. Technical Report #1 - Mathematical Methods (March 9, 2009)
2. Technical Report #2 - Equation Parameters (March 9, 2009)
3. Technical Report #3 - Validation: SizeOnly Equation (June 24, 2008)
4. Technical Report #4 - Validation: Size+Nodes Equation (June 26, 2008)
6. Technical Report #6 - Comparisons with AdjuvantOnline (July 7, 2008)
12. Technical Report #11 - Comparative Effectiveness Calculators For Predicting Melanoma Death (August 19, 2009)