Multiple Primary Cancers: Associations & Network Link Analysis

By Elena Manilich, Ph.D.[1], Arshiya Mariam, B.S.[1], Zachary Zinda, B.S.[1], & Saima Hanif, M.D.[2]

[1]John Carroll University, Cleveland, Ohio; [2]SUNY Medical Center, New York, New York

Introduction

Cancer survivors are considered to be at an increased risk of developing new primary cancers. According to recent estimates, approximately 17% of cancer diagnoses are of a second or higher-ordered primary—a number expected to increase over time.[1]

Most of the literature on multiple primary cancers (MPCs) focuses on their general incidence. Second and higher-ordered primaries can be therapy induced, syndrome related, or resulting from shared etiological factors, such as a genetic disposition or an environmental factor.[2,3]

The purpose of this study is to examine the role of lower-order primaries in determining higher-order ones, by using data science algorithms to analyze the complex relationships between MPCs. The patterns discovered in sequences of MPCs may indicate that certain patient populations are at a higher risk of developing specific primaries. Further research in this area is needed. It is our hope that this analysis will direct future research by providing the basis for deeper, more contextualized investigations.

Methodology

Data Retrieval

Data was retrieved from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. The SEER Research Data encompasses patient records from 1973-2015 and is one of the most comprehensive databases of cancer records to-date. At the onset of our analysis, the SEER Research Data had 10,050,814 records. Records were filtered as malignant according to the SEER Behavior Record for Analysis. Records were included so long as the recorded tumors were classified as malignant under the ICD-O-3 coding scheme. Records were then filtered according to the total number of tumors per patient. 89.53% of patients had only 1 recorded malignancy, 10.44% of patients had 2-4, and the last 0.03% had 5+. Only patients with 2-4 recorded malignancies were included in our analysis.

Multiple malignancies can be divided into two categories: synchronous and metachronous. SEER defines a synchronous primary as occurring within 60 days of an initial primary, and a metachronous primary as one manifesting after this two-month window. Each non-index tumor in our analysis was classified as either synchronous or metachronous according to this two-month window. Records were included only if all non-index tumors were confirmed to be metachronous. Thus, if a patient had 4 recorded malignancies, but only the second was metachronous, the record was not included in our final analysis. This filtering narrowed down the dataset to 688,892 patient records.

Analysis

After the initial preprocessing, we approached the data from two angles: network link analysis and association rule learning. First, we used a form of Google's PageRank algorithm to differentiate the relative importance of second and higher-order primaries. Relative importance is an indication of the number of cancers that lead to that malignancy. The relative importance of each cancer type is determined by two factors: the number of malignancies leading to that cancer type and the relative importance of each of those lower-order (incoming) malignancies. Next, we employed the Parallel FP-Growth algorithm introduced by Li et al. in 2008 to identify common relationships between lower and higher-order tumors. Data analysis and computations were done on Google Cloud servers using Apache PySpark and the R statistical programming language.

Network Analysis

Figure 2: Network Diagram of the Top Cumulatively-Ranked Cancer Types.

The overall ranking of relative importance is represented by the radius of the malignancy node; and the width of each edge represents the total number of patients exhibiting that relationship.

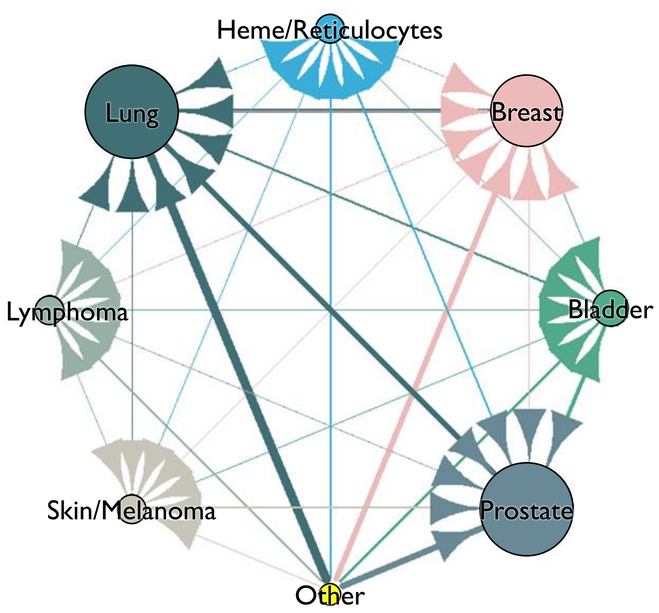


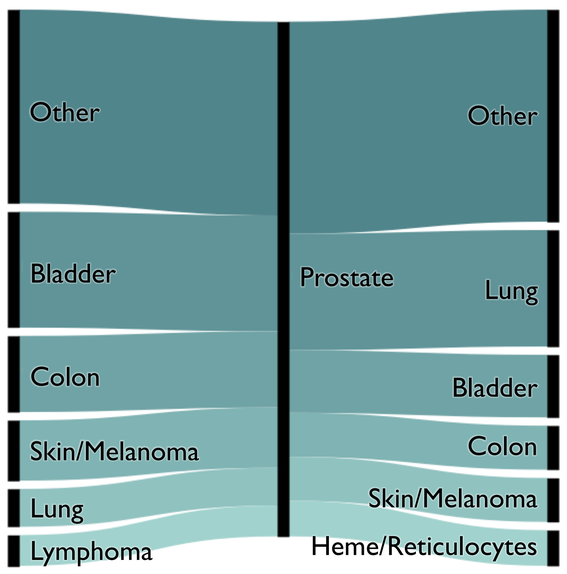
Table 1 – Top 7 Highest Ranked Cancers Based On Relative Importance

Rank	Type	N Cancers	N Cases
1	Lung	102	108,151
2	Prostate	83	59,303
3	Breast	72	42,327
4	Bladder	61	41,746
5	Heme/ Reticulocytes	50	32,432
6	Lymphoma	54	27,823
7	Skin/ Melanoma	48	24,703

Association Rule Learning

Prostate Cancer

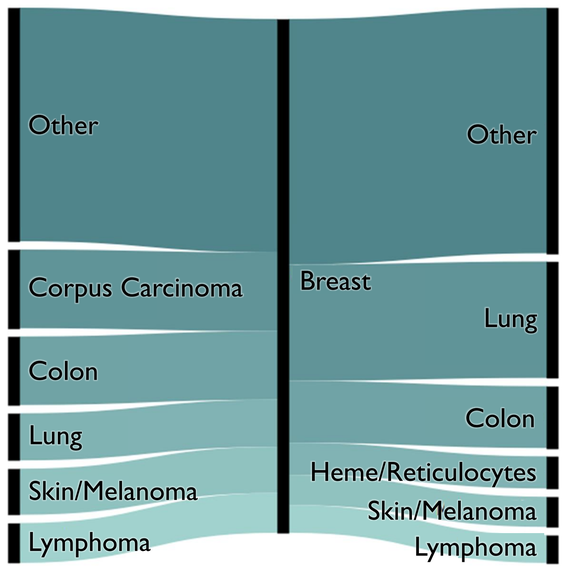
Figure 1.1: Top Most Frequent Malignancies Leading to and from a Secondary Prostate Malignancy.



Prostate cancer appeared most frequently in our dataset, with 187,442 patients diagnosed at least once. Bladder cancer was the leading antecedent to prostate cancer with a 21.5% frequency rate. Lung cancer was the leading consequent at a 21.04% frequency rate.

Breast Cancer

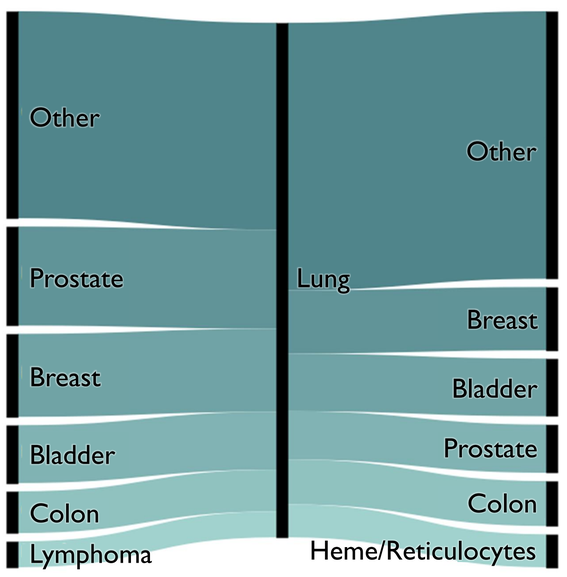
Figure 1.2: Top Most Frequent Malignancies Leading to and from a Secondary Breast Malignancy.



Breast cancer appeared second-most frequent in our dataset, with 169,647 patients diagnosed at least once. Uterine cancer was the leading antecedent to breast cancer with a 16.24% frequency rate. Lung cancer was the leading consequent at a 18.70% frequency rate.

Lung Cancer

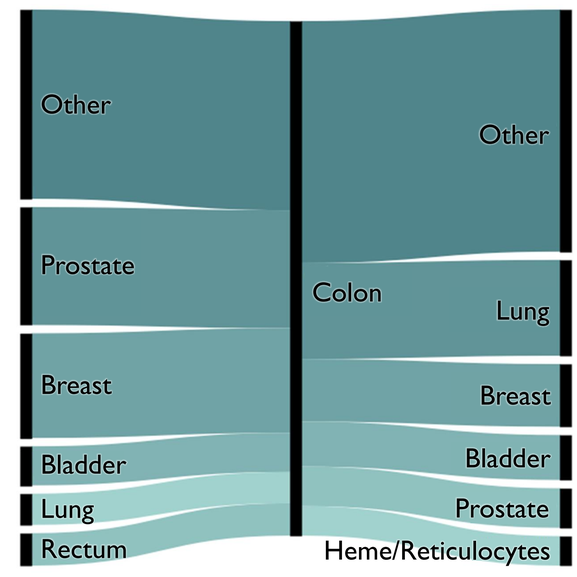
Figure 1.3: Top Most Frequent Malignancies Leading to and from a Secondary Lung Malignancy



Lung cancer appeared third-most frequent in our dataset, with 150,205 patients diagnosed at least once. Prostate cancer was the leading antecedent to lung cancer with a 21.54% frequency rate. Breast cancer was the leading consequent at a 12.01% frequency rate.

Colon Cancer

Figure 1.4: Top Most Frequent Malignancies Leading to and from a Secondary Colon Malignancy



Colon cancer appeared fourth-most frequent in our dataset, with 104,664 patients diagnosed at least once. Prostate cancer was the leading antecedent to colon cancer with a 24.31% frequency rate. Lung cancer was the leading consequent at a 18.15% frequency rate.

Table 2 – Top 5 Most Frequent Cancers Seen Together

Malignancies	N
Bladder, Prostate, Lung	1315
Bladder, Prostate, Colon	764
Colon, Prostate, Lung	676
Corpus Carcinoma, Breast, Colon	393
Colon, Breast. Lung	377

Discussion

Large scale data analysis reflecting chronological associations between different primaries opens up a new avenue for further inquiry into the development of metachronous cancers. Genetic, host, and environmental risk factors as well as chemotherapy regimens should be further analyzed for associations between multiple primaries—particularly for cancers with high relative importance. Clinically, this analysis may be used for formulating guidelines and timelines for screening tests for potential metachronous cancers.

Bibliography

[1]Vogt A, Schmid S, Heinimann K, et al. Multiple primary tumours: challenges and approaches, a review. ESMO Open. 2017;2:e000172. doi:10.1136/esmoopen-2017-00017.

[2]Sakellakis M, Peroukides S, Iconomou G, Boumpoucheropoulos S, Kalofonos H. Multiple primary malignancies: a report of two cases. Chinese Journal of Cancer Research. 2014;26(2):215-218. doi:10.3978/j.issn.1000-9604.2014.02.15.

[3]Takalkar U, Asegaonkar BN, Kodlikeri P, Asegaonkar S, Sharma B, Advani SH. An elderly woman with triple primary metachronous malignancy: A case report and review of literature. International Journal of Surgery Case Reports. 2013;4(7):593-596. doi:10.1016/j.ijscr.2013.03.032.