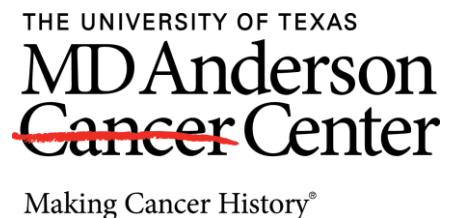


tMN Patients vs. Cancer Patient - Lab Result Elements Preliminary Analysis Overview

Zijun Wu
Data Patterns and Trend Visualization Reports

Nov 6, 2023



Agenda

Problem Statement

Data Profile

Data Processing

Data problems and difficulties & solutions

19 features introduction

General Tendency and Individual Images

Proposed Solution and Model Selection

Conclusion & Future Steps

Problem Statement

tMN prediction is a race of time



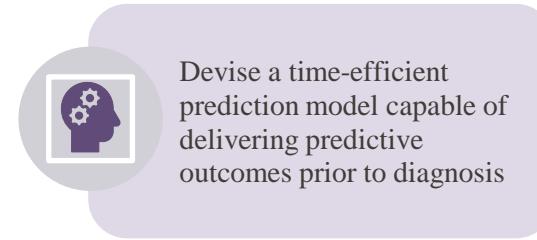
0.5%-1% cancer will become tMN patients, the longest latency could be 7 years.



Find out how to target these tMN patients before the



Therapy related leukemia progresses rapidly, and patients may face a life-threatening situation within a month.



Devise a time-efficient prediction model capable of delivering predictive outcomes prior to diagnosis

- tMN is a race of time to predict the acute disease before
- As the potential patients approaching to the critical transition point, more signs are showing on their body, but they becoming more and more vulnerable to the transition to tMNs.
- The primary objective of analyzing this dataset is to understand the hematological profiles of patients, particularly in relation to the diagnosis of tMN.
- This involves exploring trends, patterns, and potential indicators within the hematological parameters leading up to the diagnosis date.
- The lab data provides a high-level overview of the data, including patient information, hematological parameters, and temporal aspects, setting the stage for further exploration and analysis.

Data Profile



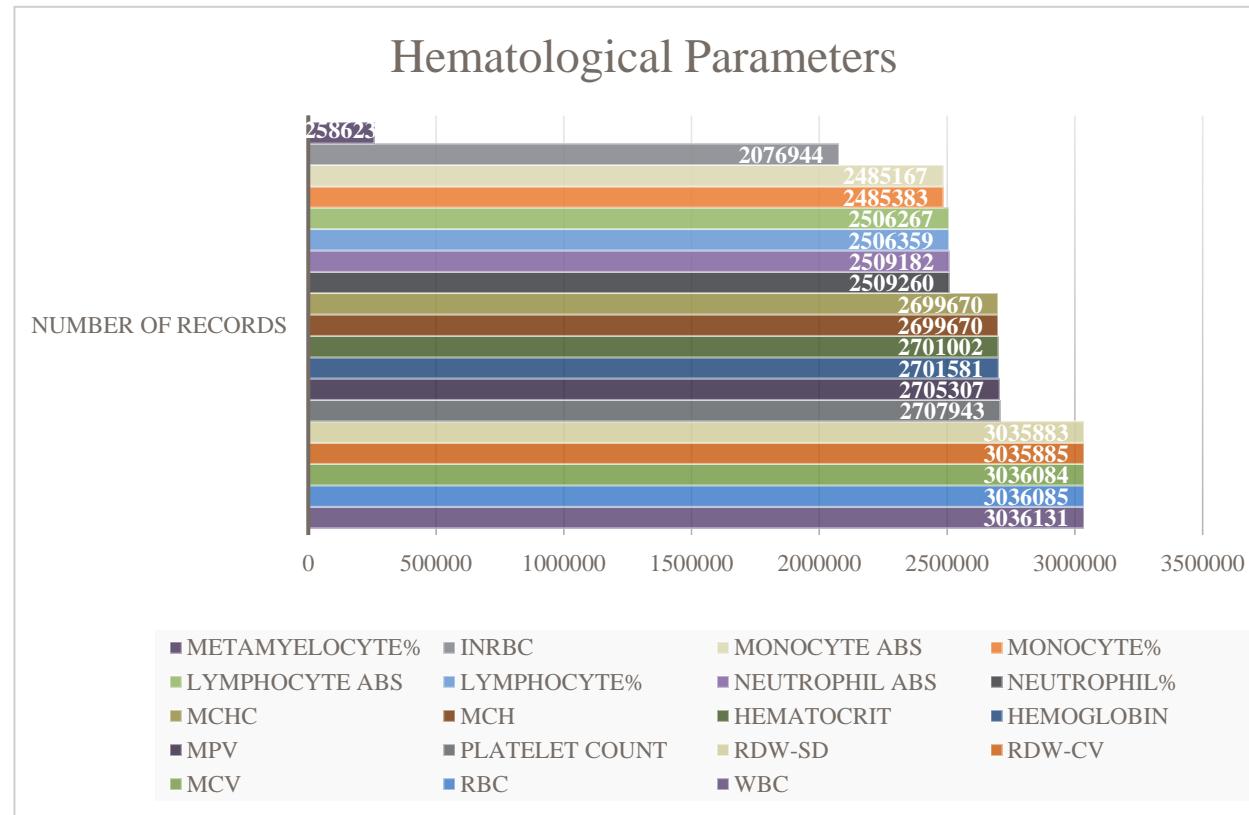
The dataset comprises 48,732,425 records capturing diverse hematological parameters across numerous patient encounters.



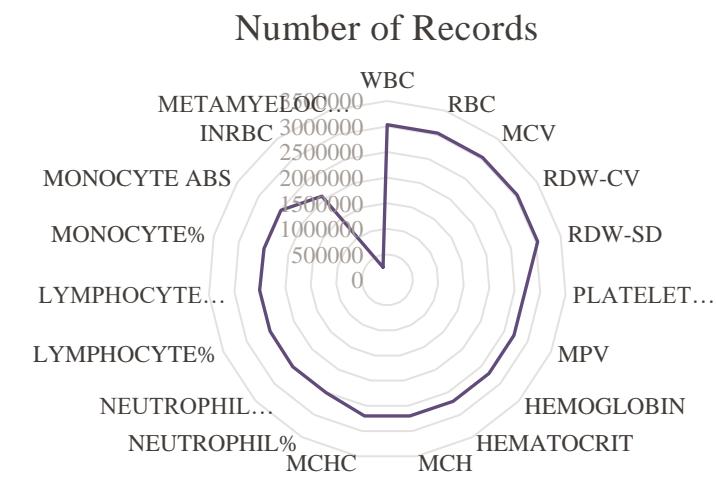
Each record provides insight into a patient's health status based on hematological measurements during a specific encounter.



Time-based information allows for the exploration of how hematological parameters relate to the diagnosis date and the time intervals leading up to the diagnosis.



Data Dimension	Descriptions
Data Variables	from 97,000 cancer patients, including 217 tMN patients
Data Size	5 million records after cleaning duplicates
Data Type	Time-based numerical data, categorical data
Target	19 categories hematological parameters



Pre-Data Processing

Data Problems:

1. No CHIP category, filtering out using indicators or other features
2. Don't have much before and after chemotherapy indicator differences or comparison (lack of relevant key features)
3. Dates of each leukemia treatment



Multi-layer filtering to get tMN patients.



No classification of tMN patients. Filtering out based on conditional filtering on 514600 diagnosis records.



Get 217 out of



Create relevant key features

Difficulties

Limited Temporal Data:

- Rapidly progressing diseases often have a short timeframe between diagnosis and critical events (e.g., death).
- This results in a limited amount of temporal data available for model training and challenges to capture the dynamics of disease progression adequately.

Imbalanced Data:

- The dataset may be imbalanced, with a small number of instances representing the critical events (e.g., death) compared to non-critical events.
- This imbalance can affect model performance, as the algorithm may be biased towards predicting the majority class.

Dynamic Nature of the Disease:

- Rapidly progressing diseases can exhibit dynamic and unpredictable patterns, making it difficult to model and predict the disease trajectory accurately.
- Changes in patient condition may occur rapidly, and capturing these nuances in the data is challenging.

Limited Feature Stability:

- Features that are relevant early in the disease progression might lose their significance as the disease advances.
- Identifying stable features that provide consistent predictive power throughout the progression becomes a challenge.

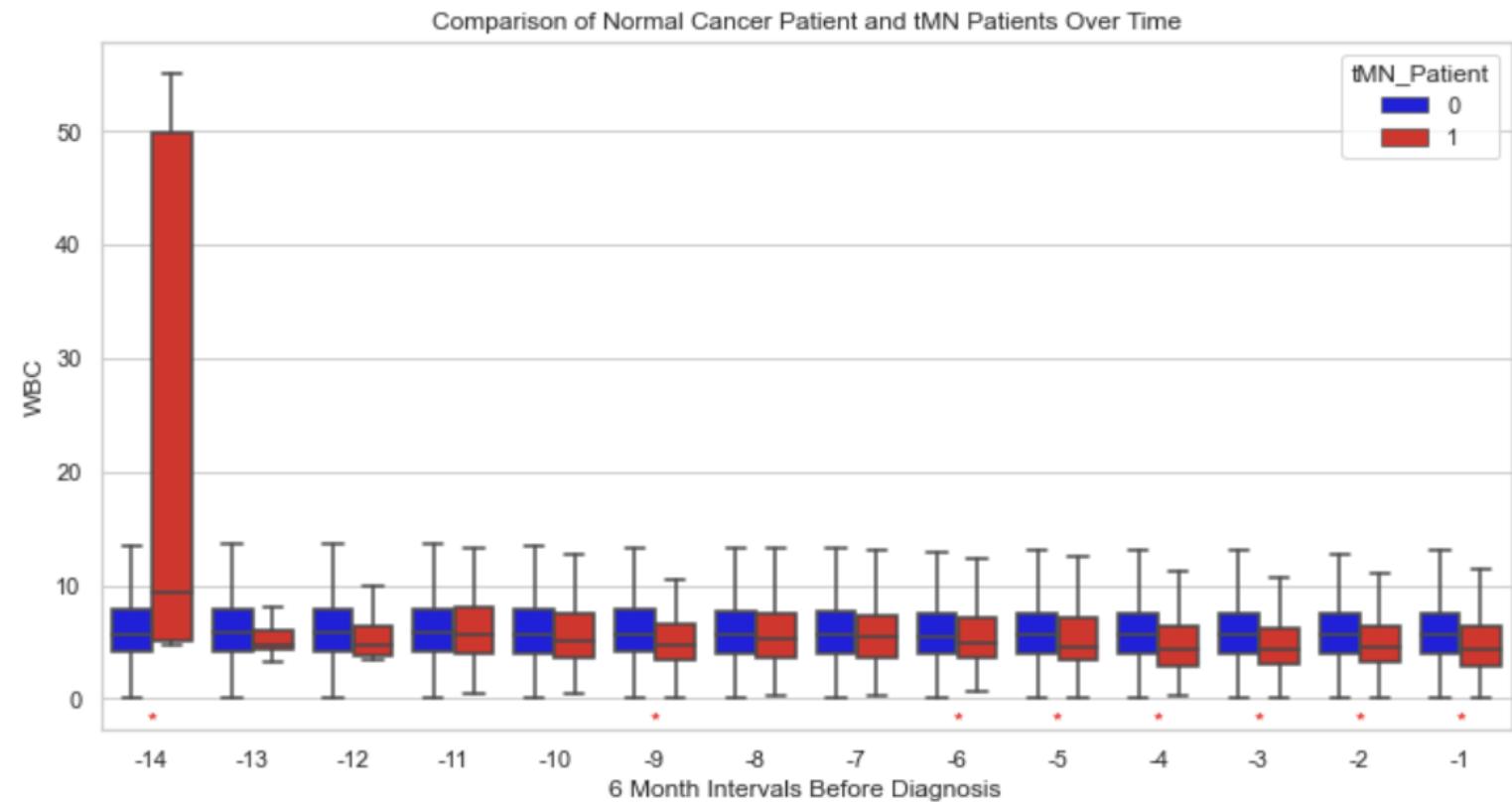
WBC

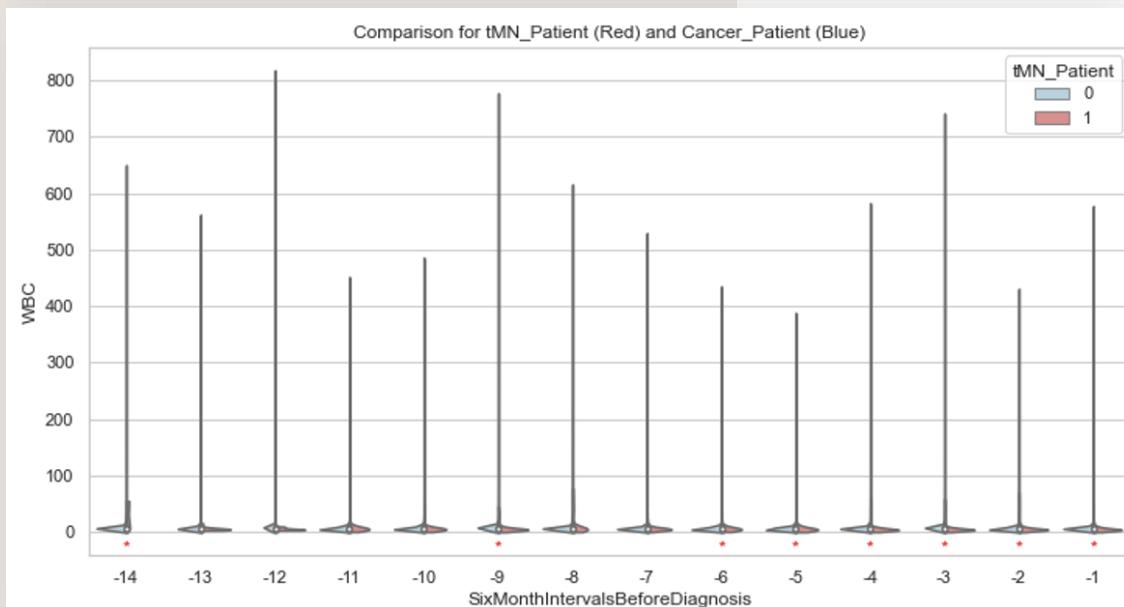
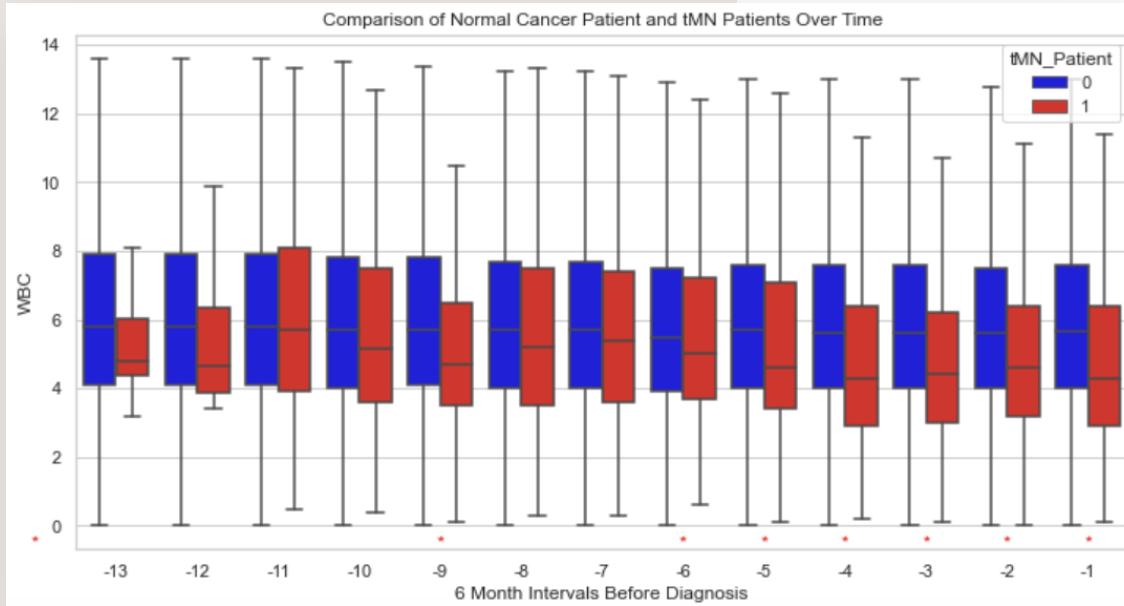
	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.460717e+09	6.999095e-53	0.00	significant
1	-2	1.374649e+09	5.219775e-30	0.00	significant
2	-3	1.071687e+09	3.413916e-47	0.00	significant
3	-4	9.438457e+08	3.783094e-36	0.00	significant
4	-5	8.816900e+08	2.948820e-12	0.00	significant
5	-6	9.008833e+08	1.015405e-03	0.00	significant
6	-7	5.405661e+08	6.544447e-02	0.07	not significant
7	-8	4.001064e+08	9.711118e-02	0.10	not significant
8	-9	3.478225e+08	7.045210e-07	0.00	significant
9	-10	2.584217e+08	7.032755e-02	0.07	not significant
10	-11	1.872571e+08	5.866058e-01	0.59	not significant
11	-12	1.571420e+07	4.364958e-01	0.44	not significant
12	-13	2.141682e+07	4.443902e-01	0.44	not significant
13	-14	1.506459e+07	4.861077e-02	0.05	significant

WBC

- Highly skewed in the -14 time interval, 7 years ago
- For time interval -14, significant but in the wrong direction

7 years before diagnosis



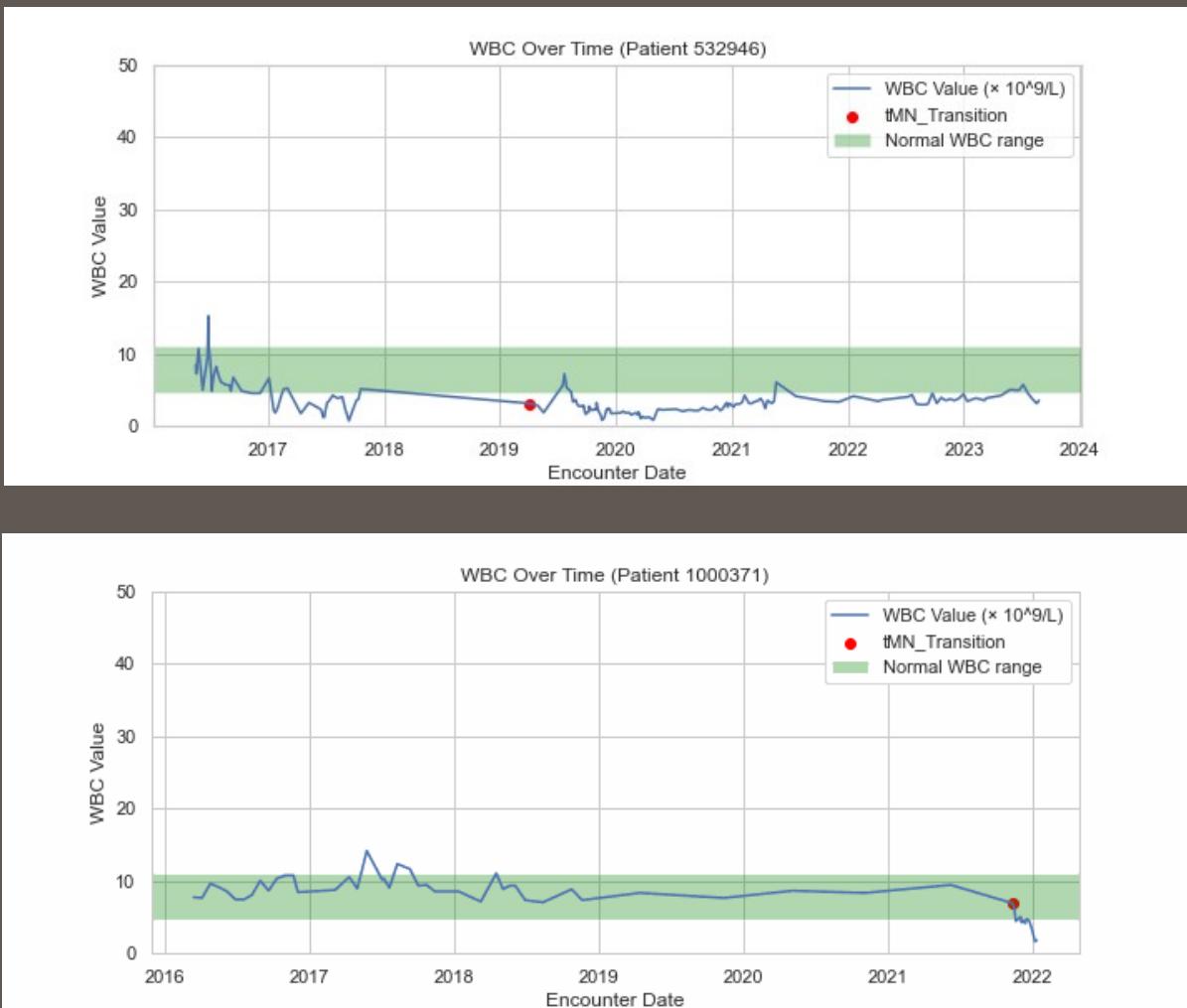


WBC – General Trend

tMN patients:

- WBC level is lower and consistently decreasing dated back to 3 years ago.
- Highly skewed in the -14 time interval, 7 years ago
- For time interval -14, significant but in the wrong direction
- Cut-off insignificant data before -6 time interval

WBC Individual Graphs – All Patients



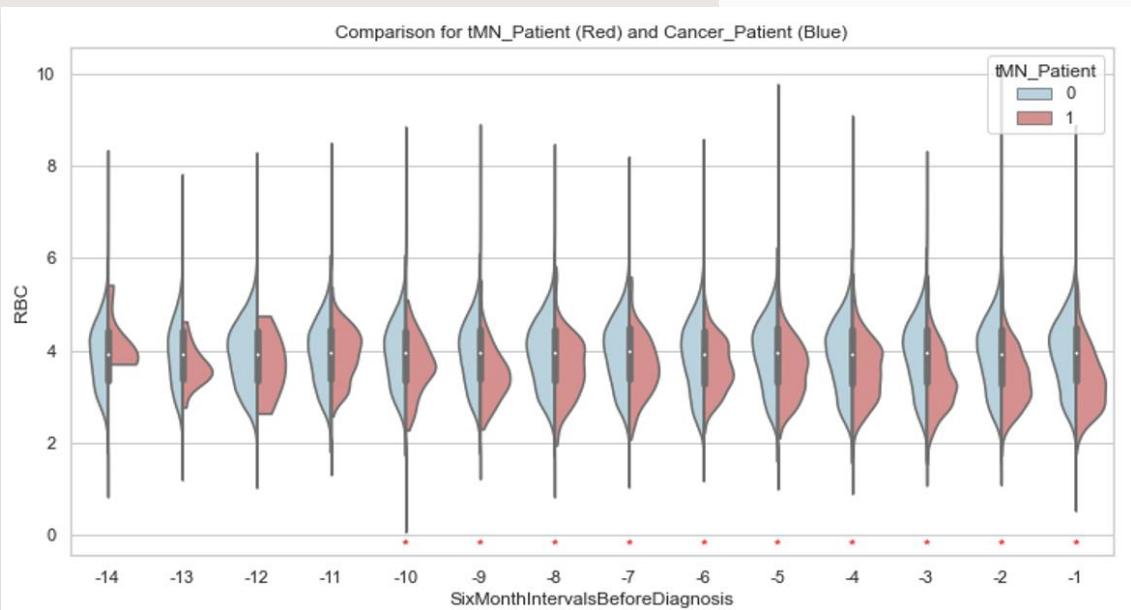
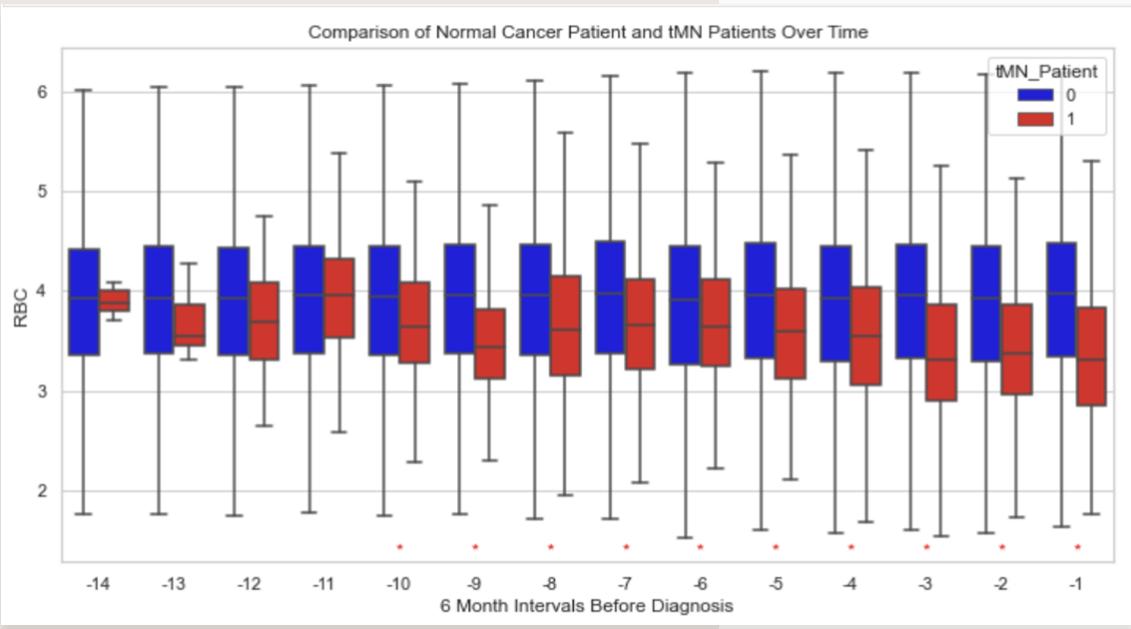
- Overall trend:
 - Statistically meaningful smaller values and decreasing trend dated back to 3 years before the tMN Transition

- Fluctuations:
 - Changes in White Blood Cell (WBC) commonly could be an immune system response to an infection or inflammation.

- Conclusion:
 - Cannot capture the majority cases, overall, not a good predictor by itself.

RBC

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.158532e+09	7.118081e-138	0.00	significant
1	-2	1.069781e+09	2.673230e-105	0.00	significant
2	-3	8.896615e+08	2.981203e-99	0.00	significant
3	-4	9.255359e+08	3.022547e-40	0.00	significant
4	-5	7.697118e+08	1.542758e-32	0.00	significant
5	-6	7.766125e+08	5.202337e-19	0.00	significant
6	-7	4.644012e+08	2.322988e-10	0.00	significant
7	-8	3.293987e+08	7.786053e-11	0.00	significant
8	-9	2.738366e+08	8.497748e-24	0.00	significant
9	-10	2.270334e+08	8.268789e-06	0.00	significant
10	-11	1.822480e+08	9.837058e-01	0.98	not significant
11	-12	1.482715e+07	2.834379e-01	0.28	not significant
12	-13	1.866487e+07	1.195921e-01	0.12	not significant
13	-14	1.134674e+07	7.229548e-01	0.72	not significant



RBC – General Trend*

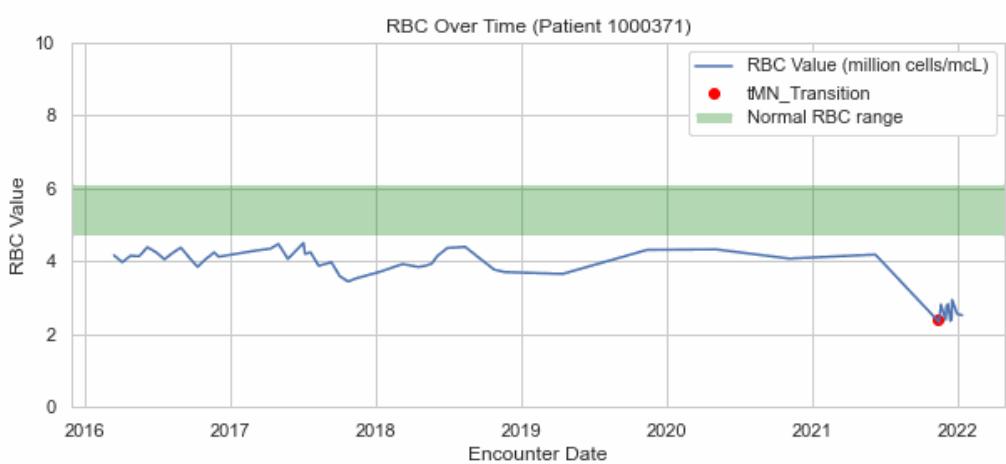
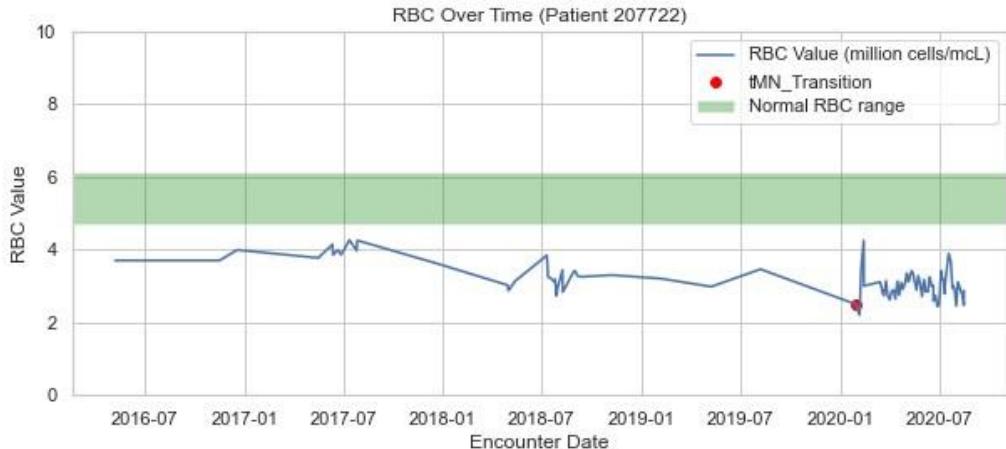
tMN patients:

- RBC level is lower and consistently decreasing dated back to 5 years ago
- Comparatively strong indicators

Significance Fluctuations:

- Relatively smaller sample size before time interval -10 and contains outliers noises

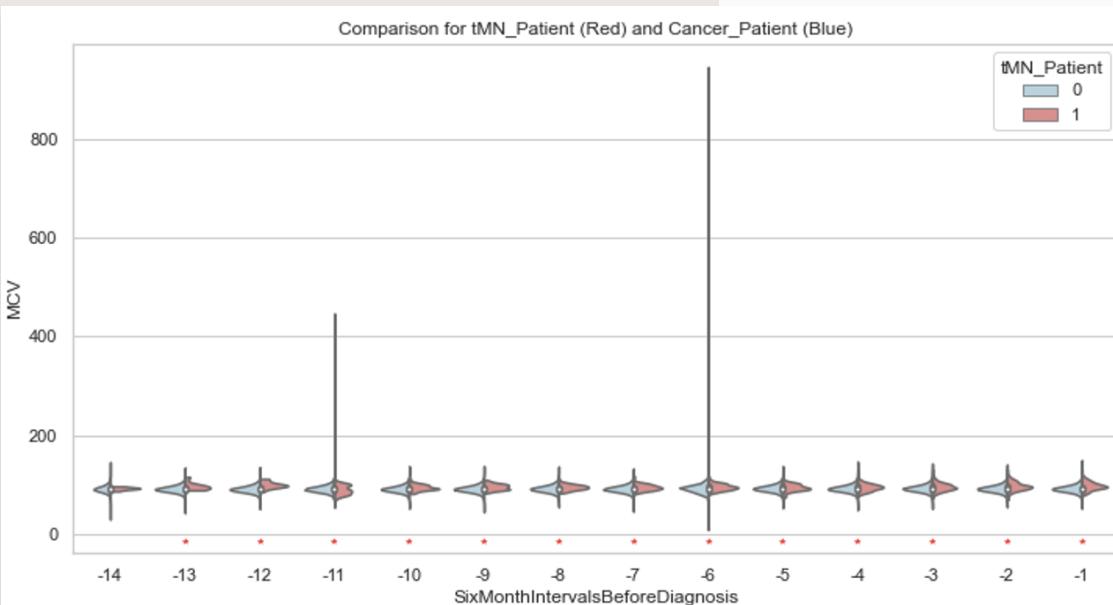
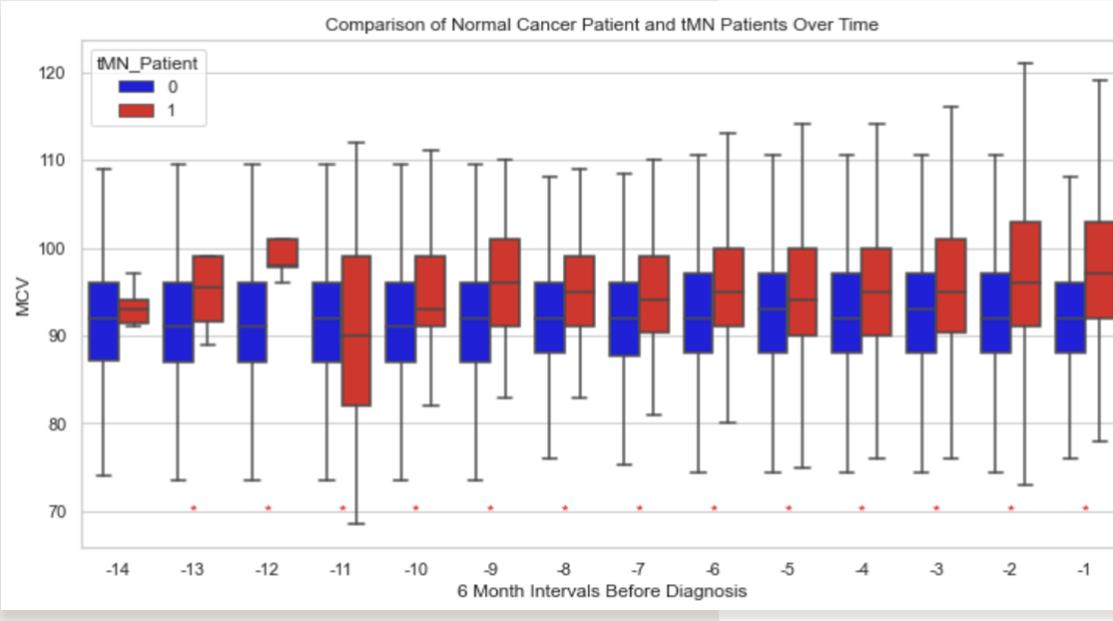
RBC Individual Graphs – All Patients



- Overall trend:
 - Statistically meaningful smaller values and decreasing trend dated back to 5 years before the tMN Transition
- Fluctuations Probable Causes:
 - Hemorrhagic Events, such as physical injury or surgical interventions.
 - Chronic Diseases and Infections
- Conclusion:
 - Relatively stronger indicators compared to WBC, but still need other parameters

MCV

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	2.695275e+09	1.450033e-130	0.00	significant
1	-2	2.290211e+09	2.717519e-88	0.00	significant
2	-3	1.859484e+09	9.723253e-49	0.00	significant
3	-4	1.564212e+09	6.357769e-34	0.00	significant
4	-5	1.259287e+09	1.254044e-21	0.00	significant
5	-6	1.251112e+09	2.835526e-36	0.00	significant
6	-7	7.064555e+08	1.680885e-15	0.00	significant
7	-8	5.398685e+08	2.183690e-15	0.00	significant
8	-9	5.539355e+08	2.148261e-20	0.00	significant
9	-10	3.387634e+08	6.443599e-07	0.00	significant
10	-11	1.599505e+08	2.060035e-02	0.02	significant
11	-12	2.880812e+07	3.476680e-04	0.00	significant
12	-13	3.172098e+07	2.758602e-02	0.03	significant
13	-14	1.136124e+07	7.179813e-01	0.72	not significant

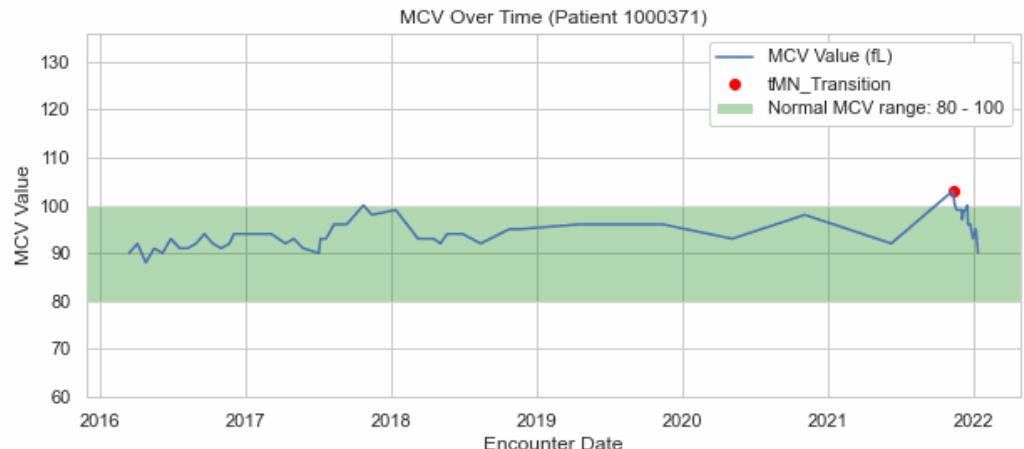
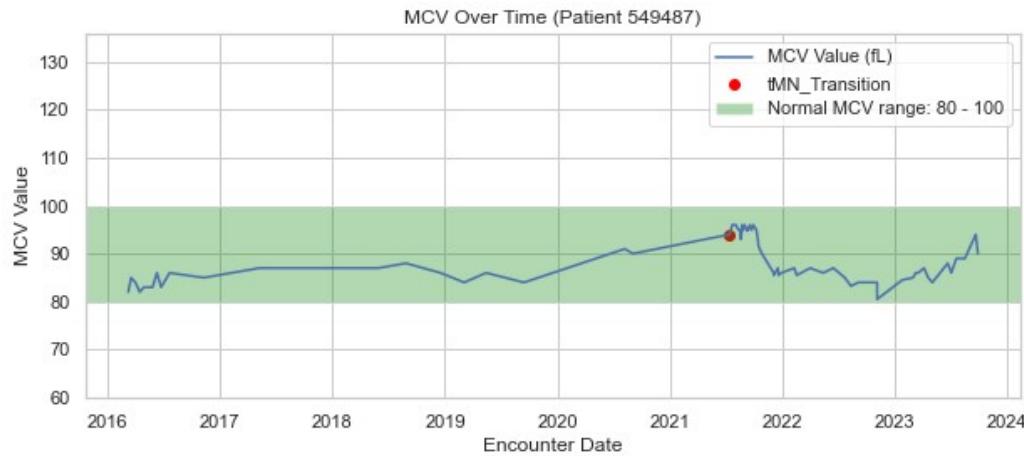
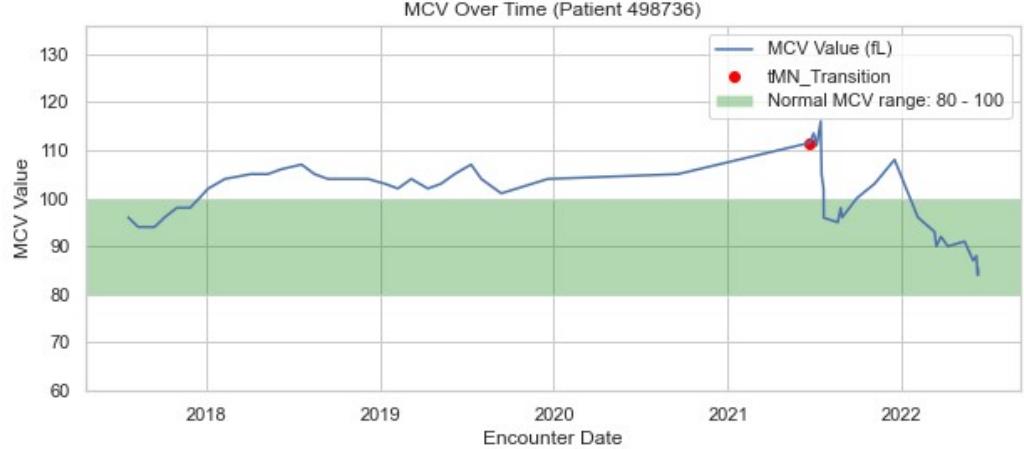


MCV – General Trend*

tMN patients:

- MCV level is higher and consistently increasing dated back to 5 years ago (no medium fluctuation within this set period)
- Fluctuate at interval -11 with outliers and contains highly skewed at time interval -6
- At -6 time interval, longer whisker in the violin plot indicates extreme outliers or a highly skewed distribution that deviate significantly from the majority of the data

MCV Individual Graphs – All Patients

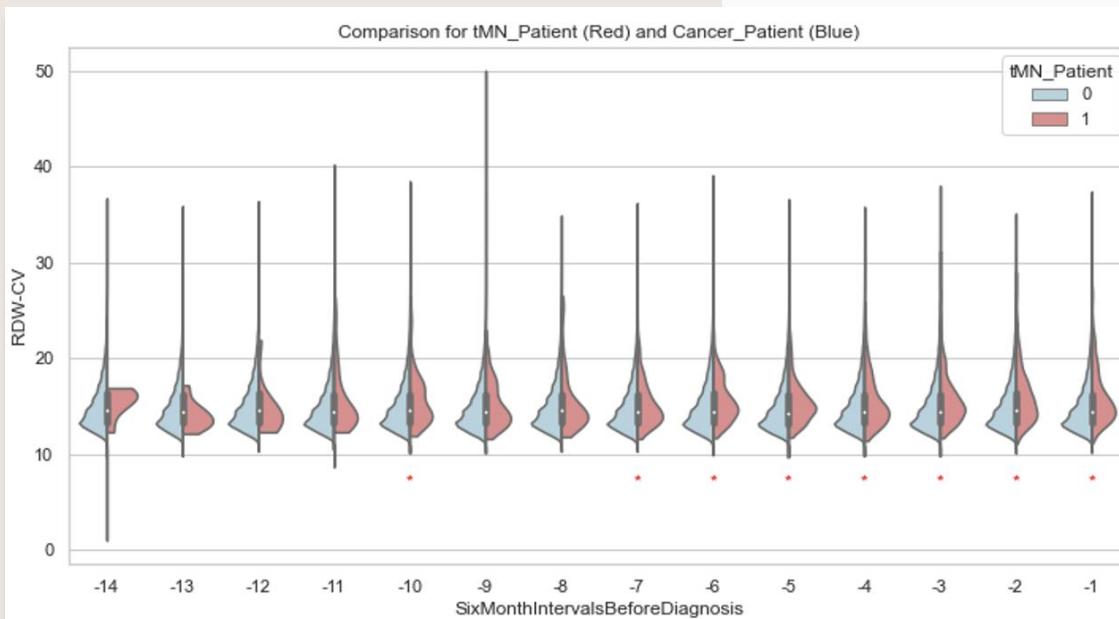
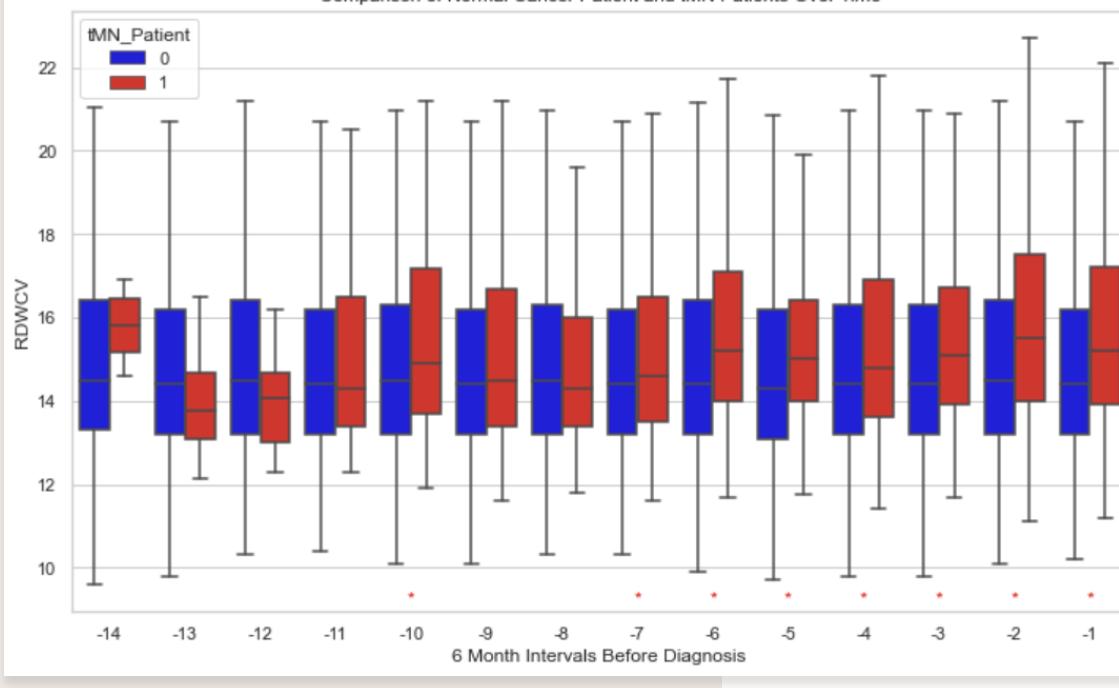


- Overall trend:
 - Statistically meaningful smaller values and decreasing trend dated back to 5 years before the tMN Transition
- Typical Patterns:
 - At the time interval -3, the statistical difference become smaller and fluctuates.
- Conclusion:
 - Very strong and significant indicators with interesting patterns for future analysis

RDW-CV

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	2.324107e+09	1.281265e-37	0.00	significant
1	-2	2.094330e+09	1.284726e-39	0.00	significant
2	-3	1.702185e+09	2.103534e-19	0.00	significant
3	-4	1.392521e+09	2.091837e-08	0.00	significant
4	-5	1.193667e+09	4.023259e-11	0.00	significant
5	-6	1.156488e+09	2.158911e-16	0.00	significant
6	-7	6.150488e+08	1.211486e-02	0.01	significant
7	-8	4.337952e+08	4.740524e-01	0.47	not significant
8	-9	4.371406e+08	2.476238e-01	0.25	not significant
9	-10	3.243508e+08	1.995611e-04	0.00	significant
10	-11	1.906617e+08	3.821091e-01	0.38	not significant
11	-12	1.528149e+07	3.530811e-01	0.35	not significant
12	-13	1.915963e+07	1.551048e-01	0.16	not significant
13	-14	1.298096e+07	2.900307e-01	0.29	not significant

Comparison of Normal Cancer Patient and tMN Patients Over Time



RDW-CV – General Trend

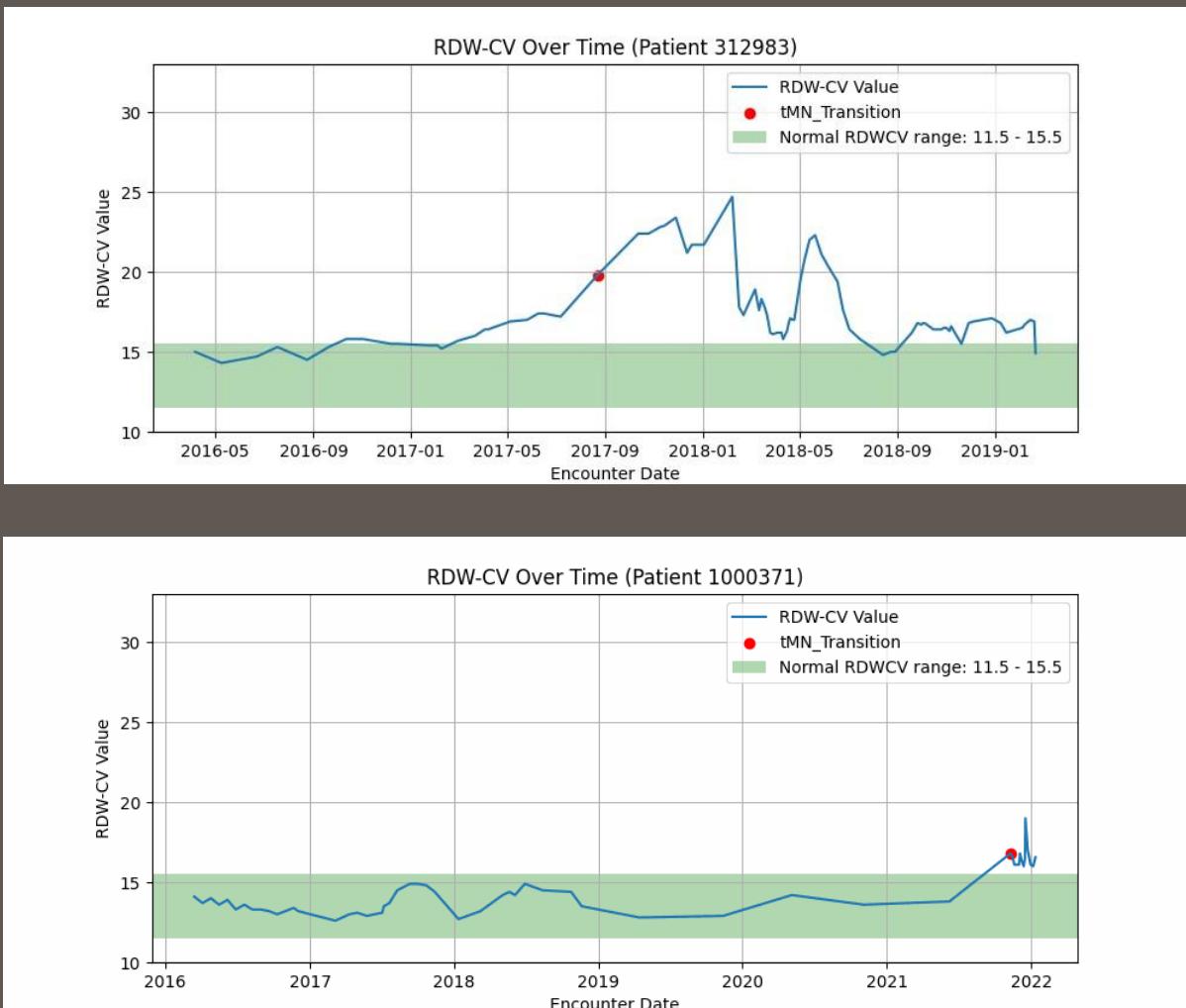
tMN patients:

- RDW-CV level is higher and consistently decreasing dated back to 3.5 years ago
- Comparatively strong indicators with stable trends after the -3.5 time interval

Significance Fluctuations:

- Relatively smaller sample size before time interval -10 and contains outliers noises

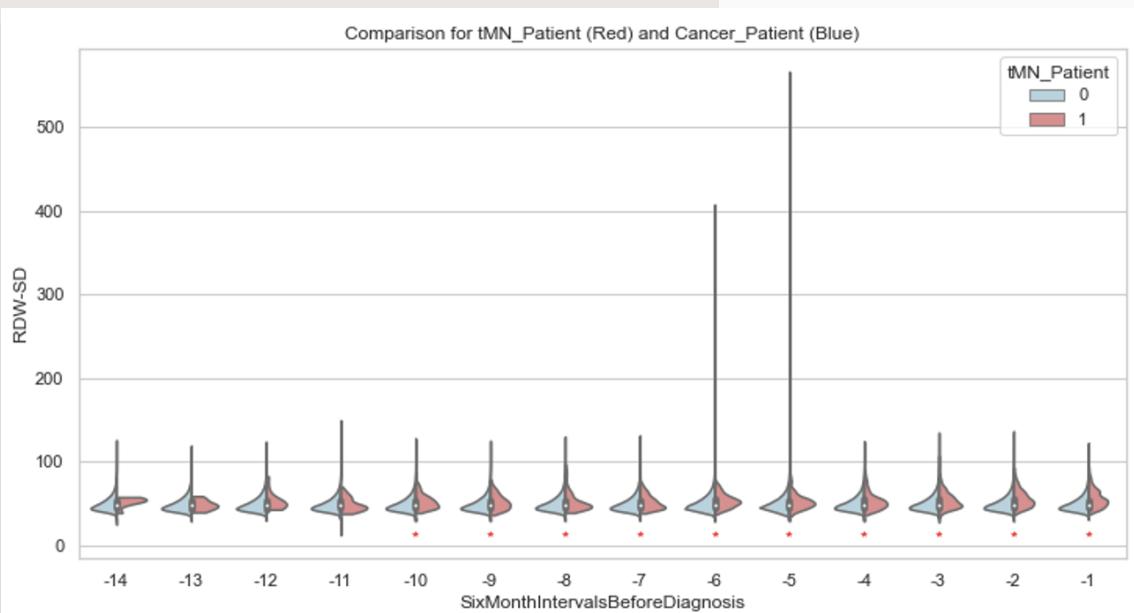
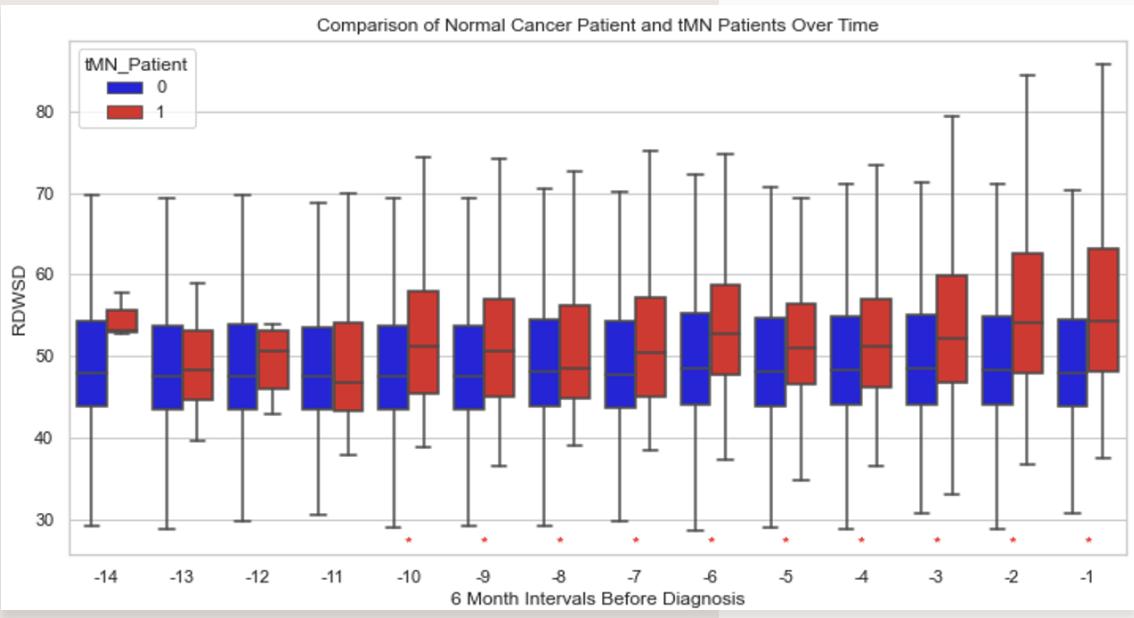
RDW-CV Individual Graphs



- Overall trend:
 - Statistically meaningful larger values dated back to 3.5 years before the tMN Transition
- Fluctuations Probable Causes:
 - RDW-CV tends to be higher in the elderly.
 - Repeated blood transfusions, especially in individuals with chronic anemias or hemoglobinopathies
 - Chronic Diseases, kidney disease or chronic inflammatory disorders
- Conclusion:
 - Relatively strong indicators but varies case by case. Still need helps from other parameters.

RDW-SD

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	2.610024e+09	2.028237e-107	0.00	significant
1	-2	2.313638e+09	7.850505e-95	0.00	significant
2	-3	1.832728e+09	6.616625e-44	0.00	significant
3	-4	1.489506e+09	2.673317e-21	0.00	significant
4	-5	1.248222e+09	2.473511e-19	0.00	significant
5	-6	1.256684e+09	3.376780e-37	0.00	significant
6	-7	6.549187e+08	1.160833e-06	0.00	significant
7	-8	4.636142e+08	5.730066e-03	0.01	significant
8	-9	4.755878e+08	1.422404e-04	0.00	significant
9	-10	3.348585e+08	4.090809e-06	0.00	significant
10	-11	1.706926e+08	2.255435e-01	0.23	not significant
11	-12	2.130349e+07	2.844049e-01	0.28	not significant
12	-13	2.312181e+07	7.776662e-01	0.78	not significant
13	-14	1.370596e+07	1.696425e-01	0.17	not significant



RDW-SD – General Trend*

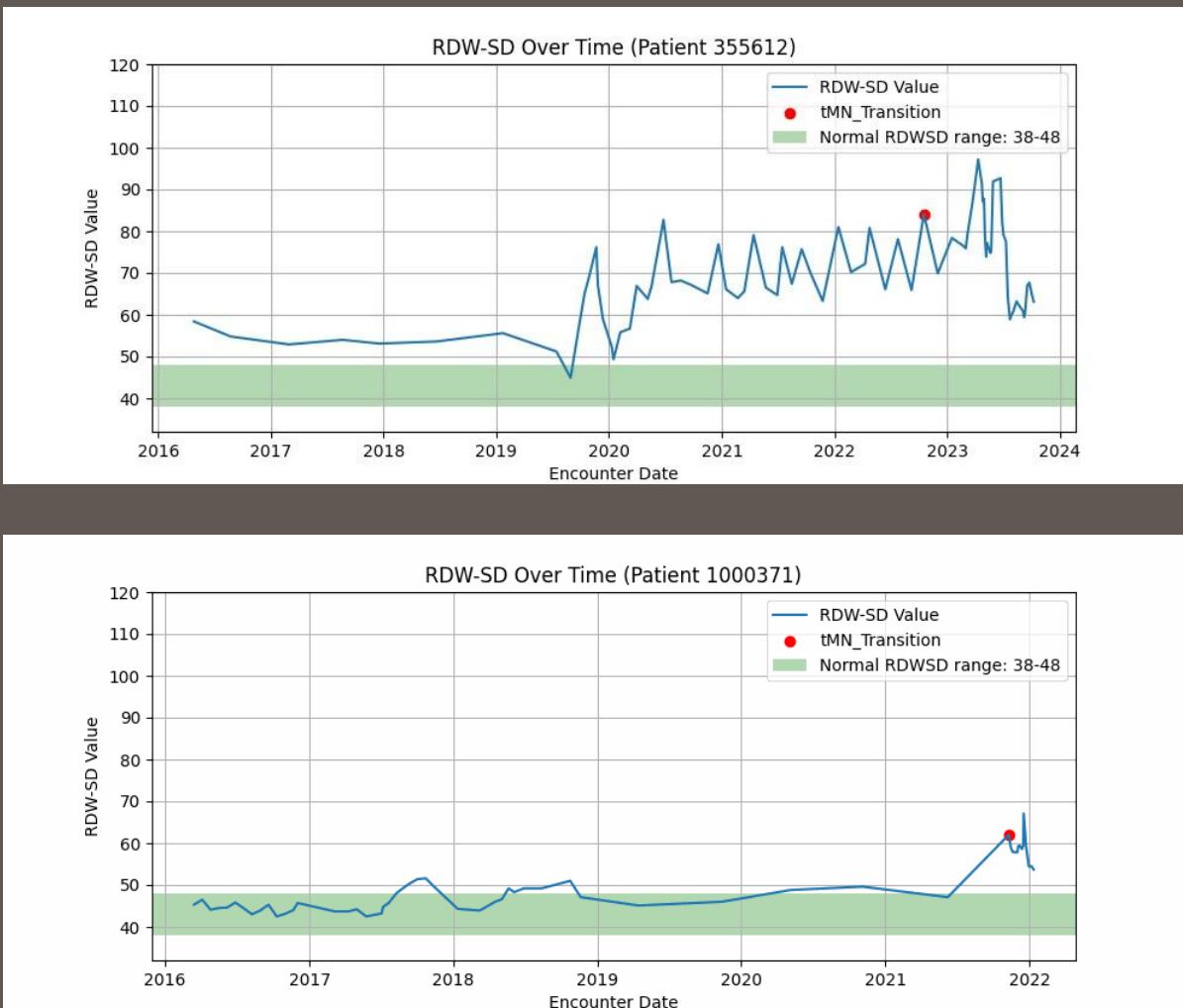
tMN patients:

- RDW-SD level is higher and consistently decreasing dated back to 5 years ago
- Comparatively strong indicators with stable trends after the -10 time interval
- Slightly fluctuates at the -8 time interval

Significance Fluctuations:

- We can easily see the values before time interval -10 obviously fluctuates with outliers and smaller sample size
- -6 and -5 interval on violin plot starts to contain extreme values

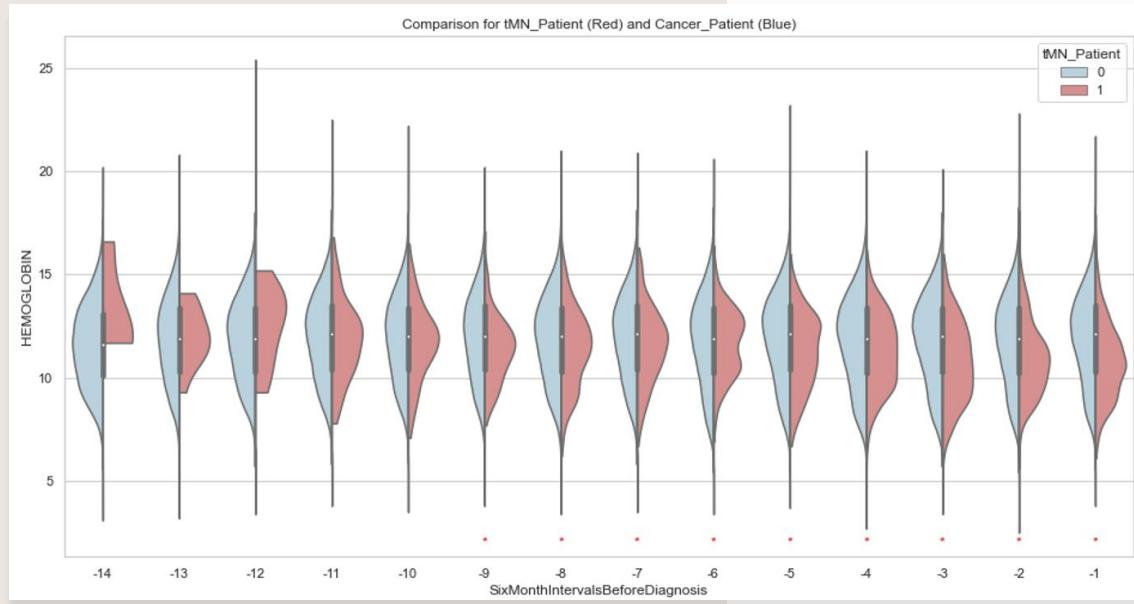
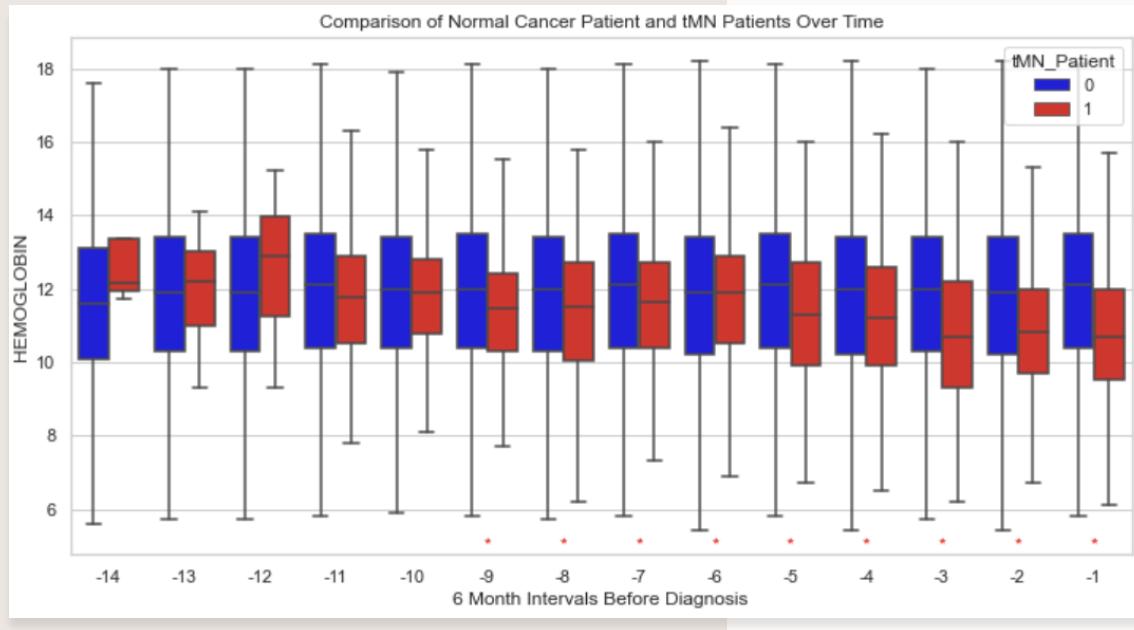
RDW-SD Individual Graphs



- Overall trend:
 - Statistical meaningful larger values dated back 5 years before the tMN Transition
 - At the time interval -8, the statistical difference become smaller and fluctuates.
- Reversing Pattern Causes:
 - Chronic Diseases with Stable Red Blood Cell Size, such as certain chronic inflammatory disorders or chronic kidney disease
 - Certain non-leukemic bone marrow disorders, like myelodysplastic syndromes (MDS) without progression to leukemia, can impact red blood cell production, causing a more uniform size distribution and reducing RDW-SD.
- Conclusion:
 - Relatively strong indicator and is more stable than RDW-CV

HGB

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.108781e+09	3.597333e-72	0.00	significant
1	-2	1.014262e+09	1.933357e-51	0.00	significant
2	-3	8.209793e+08	4.072260e-52	0.00	significant
3	-4	8.249817e+08	2.686023e-16	0.00	significant
4	-5	6.937679e+08	8.183707e-13	0.00	significant
5	-6	7.193744e+08	1.710005e-02	0.02	significant
6	-7	4.010683e+08	6.347885e-03	0.01	significant
7	-8	2.799044e+08	7.740918e-04	0.00	significant
8	-9	2.553841e+08	4.724000e-05	0.00	significant
9	-10	2.082006e+08	2.478135e-01	0.25	not significant
10	-11	1.216439e+08	5.707972e-01	0.57	not significant
11	-12	1.757710e+07	2.672503e-01	0.27	not significant
12	-13	1.807133e+07	8.126951e-01	0.81	not significant
13	-14	6.697567e+06	3.860386e-01	0.39	not significant



HGB – General Trend

tMN patients:

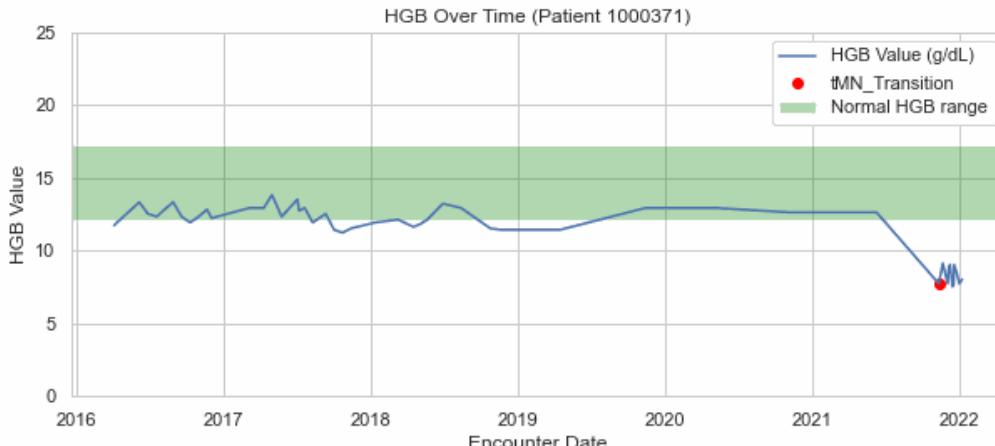
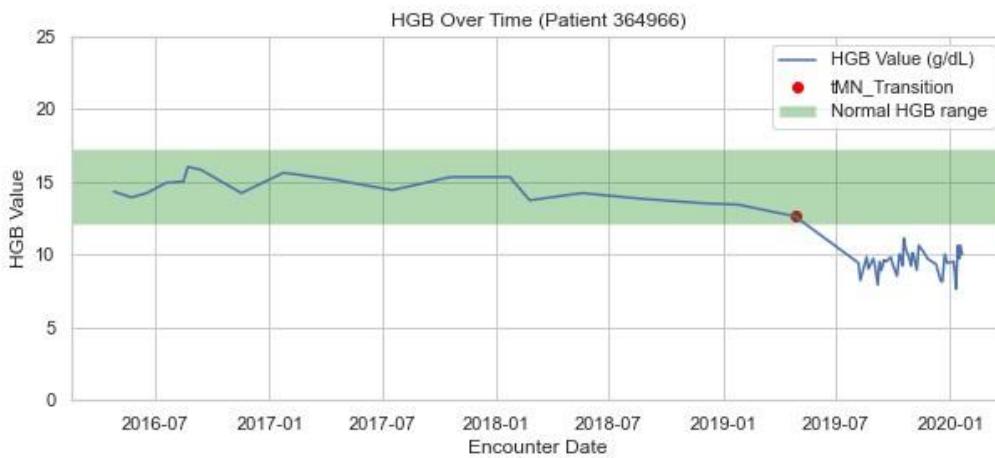
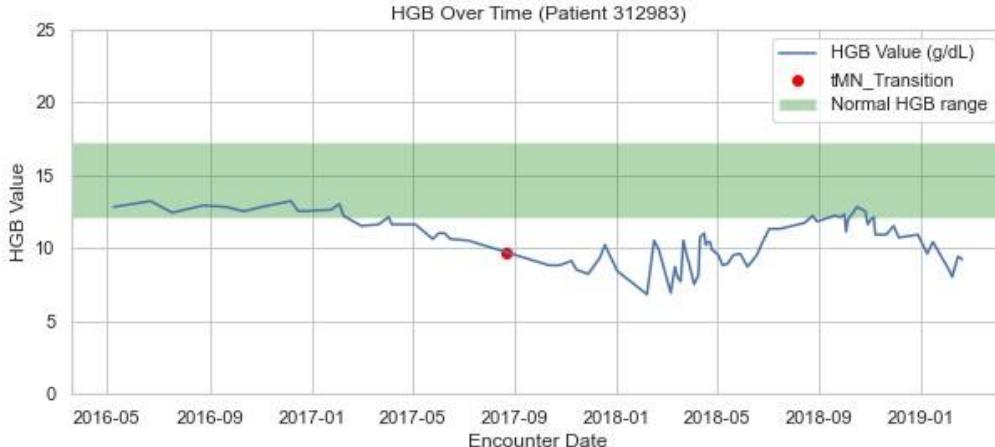
- HGB level is generally lower and consistently decreasing dated back to 5 years ago
- Fluctuates at -6 time interval

Significance Fluctuations:

- General presence of outliers that lead to extended tails
- Multimodal distribution at -6 time interval, suggesting the potential existence of distinct subgroups or patterns within the data
- The presence of two peaks hints at different characteristics or behaviors within the tMN patient group and needs further analysis

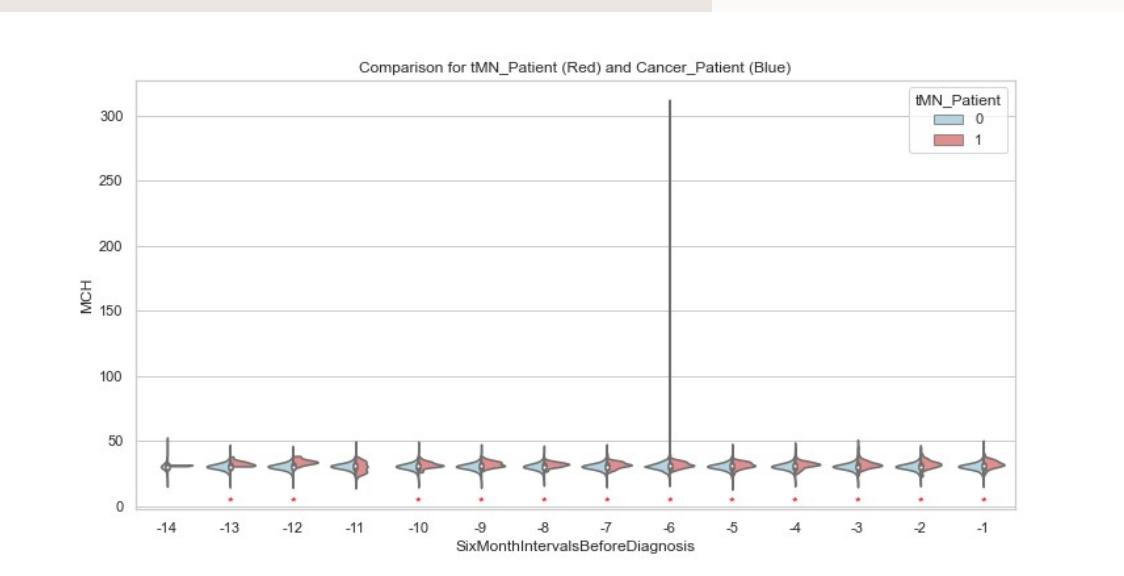
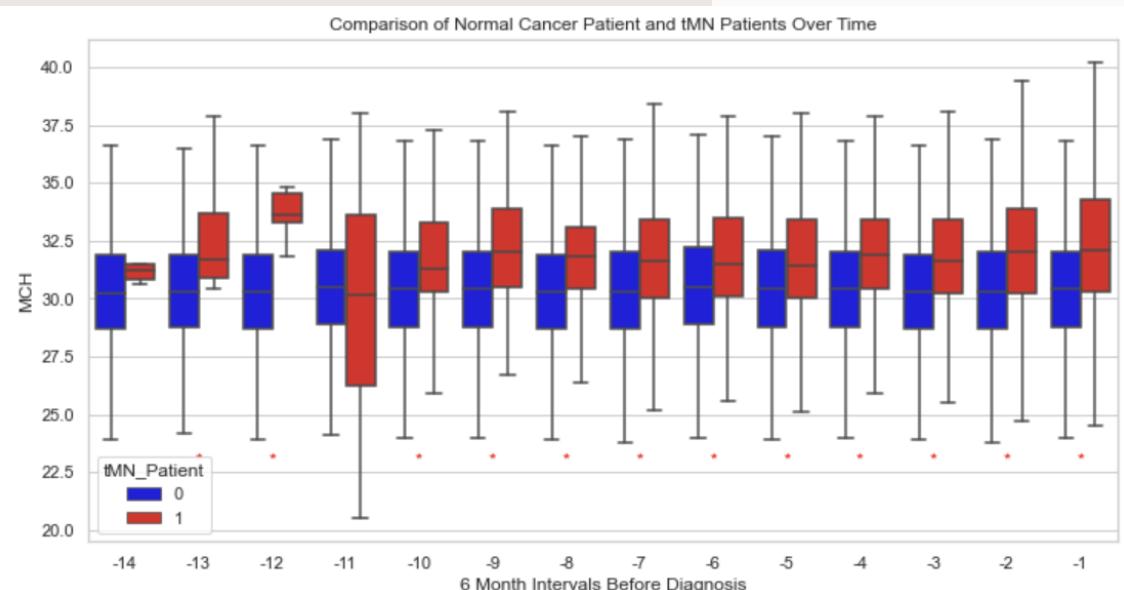
HGB - Individual Graphs

- Overall trend:
 - Statistical meaningful smaller values dated back 2.5 years before the tMN Transition
 - At the time interval -6, the statistical difference become smaller and fluctuates.
- Fluctuation Probable Causes:
 - Clinical Subpopulations, such as subgroups with different medical conditions, age groups, or other demographic characteristics.
 - Disease States, associated with different disease states or medical conditions
 - Treatment Effects, such as medications, therapies, or interventions on hemoglobin levels
- Conclusion:
 - Need more analysis



MCH

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	2.166446e+09	1.419241e-105	0.00	significant
1	-2	1.856448e+09	1.381065e-78	0.00	significant
2	-3	1.513357e+09	4.969138e-52	0.00	significant
3	-4	1.291254e+09	3.191559e-44	0.00	significant
4	-5	1.037359e+09	8.982559e-27	0.00	significant
5	-6	9.832096e+08	6.879927e-33	0.00	significant
6	-7	5.522796e+08	5.622639e-16	0.00	significant
7	-8	4.244530e+08	2.017723e-18	0.00	significant
8	-9	4.195704e+08	6.627789e-24	0.00	significant
9	-10	2.801450e+08	8.756989e-10	0.00	significant
10	-11	1.175913e+08	2.730243e-01	0.27	not significant
11	-12	2.506790e+07	5.400326e-05	0.00	significant
12	-13	2.542838e+07	3.944030e-03	0.00	significant
13	-14	6.677974e+06	3.912997e-01	0.39	not significant



MCH – General Trend

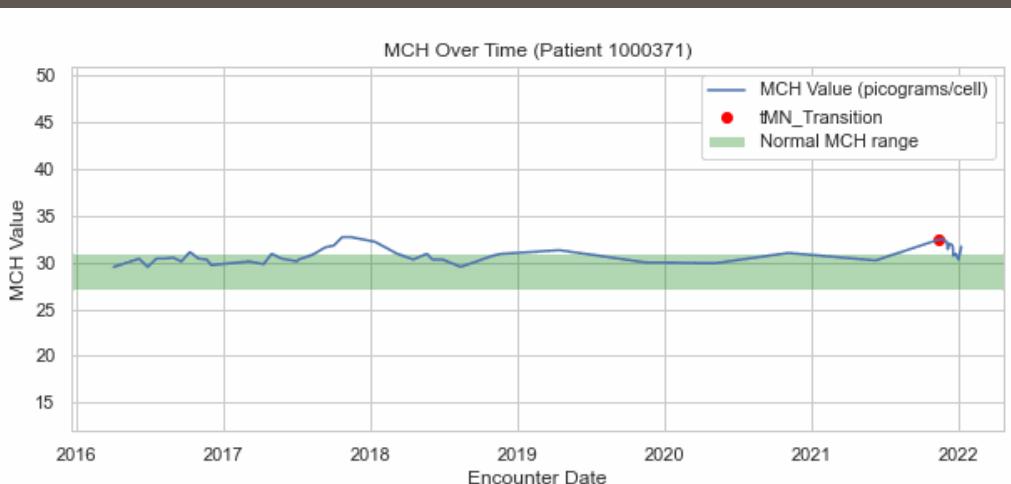
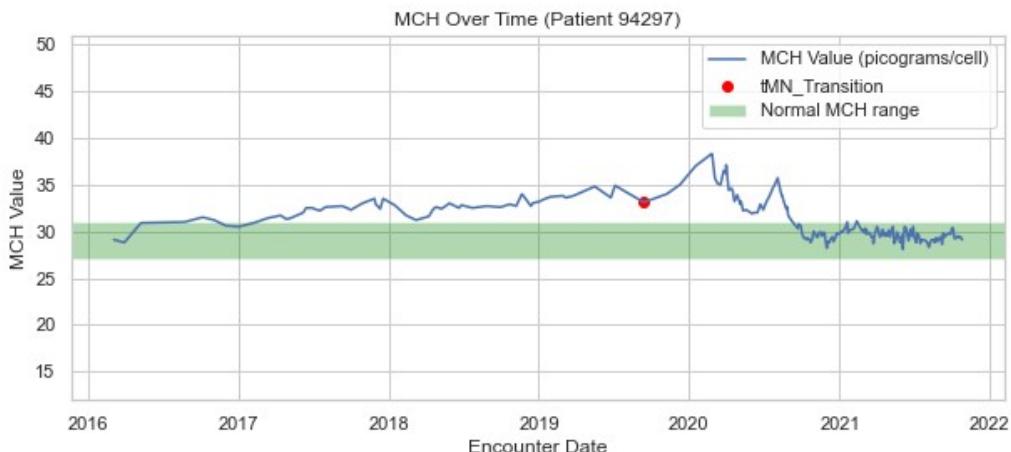
tMN patients:

- MCH level is generally higher and consistently increasing dated back to 5 years ago
- Fluctuates at -11 time interval, needs potential cut-off

Significance Fluctuations:

- Outliers that lead to extended tails in -6 time interval
- The p-value significant result also fits in the graph, for example, the -11 interval has so many value spread out below and loses the significance
- Stable trend from -10 to -1 time intervals

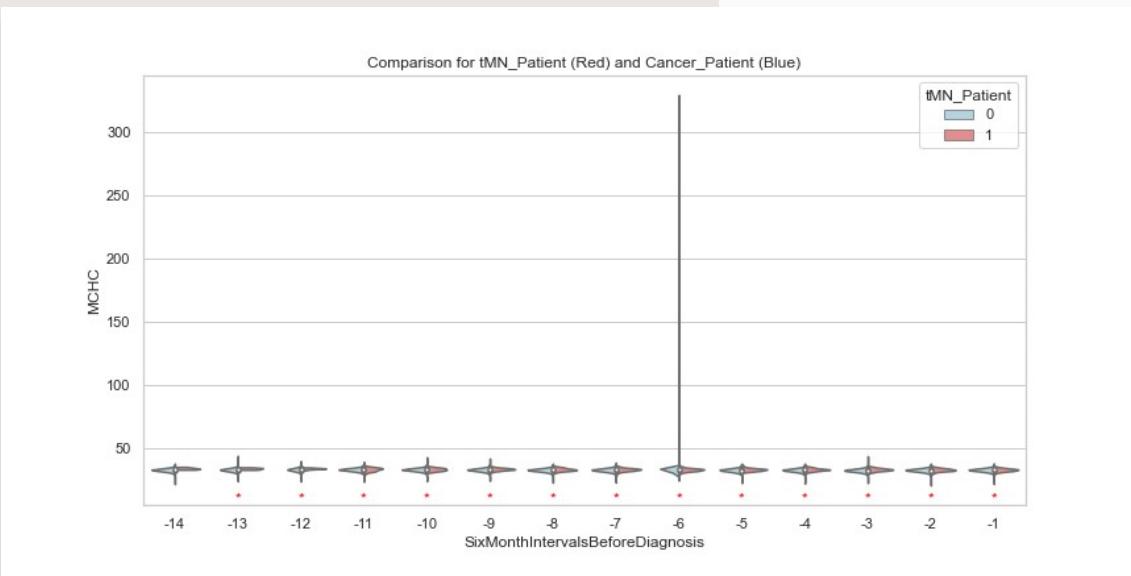
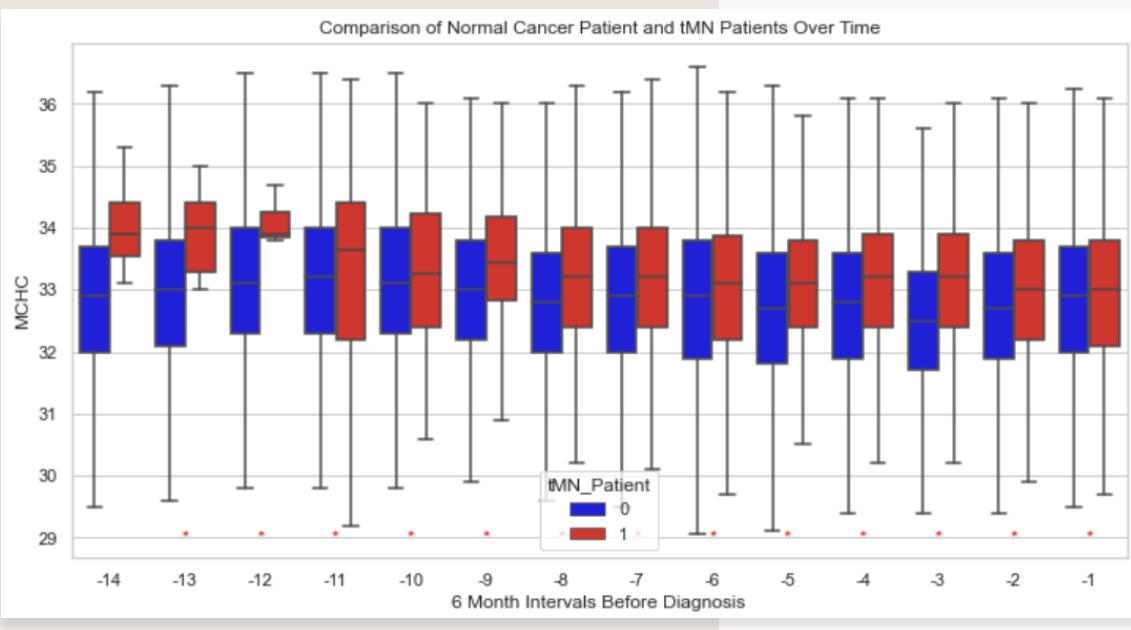
MCH - Individual Graphs



- Overall trend:
 - Statistical meaningful larger values dated back 5 years before the tMN Transition
- Fluctuation Probable Causes:
 - Chronic Inflammation: Cancer and its treatment can induce chronic inflammation, contributing to anemia and affecting MCH levels.
- Conclusion:
 - Need to combine with other parameters for further analysis

MCHC

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.648680e+09	1.827941e-02	0.02	significant
1	-2	1.453362e+09	9.925359e-03	0.01	significant
2	-3	1.327140e+09	2.242279e-12	0.00	significant
3	-4	1.138868e+09	1.751821e-11	0.00	significant
4	-5	9.278905e+08	4.965158e-07	0.00	significant
5	-6	8.249554e+08	7.585050e-04	0.00	significant
6	-7	5.021616e+08	6.281804e-06	0.00	significant
7	-8	3.683161e+08	4.959783e-05	0.00	significant
8	-9	3.905529e+08	3.310929e-14	0.00	significant
9	-10	2.633226e+08	9.365063e-06	0.00	significant
10	-11	1.513355e+08	6.514886e-04	0.00	significant
11	-12	2.291762e+07	1.378434e-03	0.00	significant
12	-13	2.695063e+07	6.040976e-04	0.00	significant
13	-14	8.340593e+06	5.315934e-02	0.05	not significant



MCHC – General Trend

tMN patients:

- MCHC level is consistently higher and in all time intervals
- No significant variation but slightly fluctuates at -10 time interval

Significance Fluctuations:

- No significant fluctuations, very strong indicators and shows a more stable trend compared to other parameters
- Outliers that lead to extended tails in -6 time interval

MCHC - Individual Graphs

➤ Overall trend:

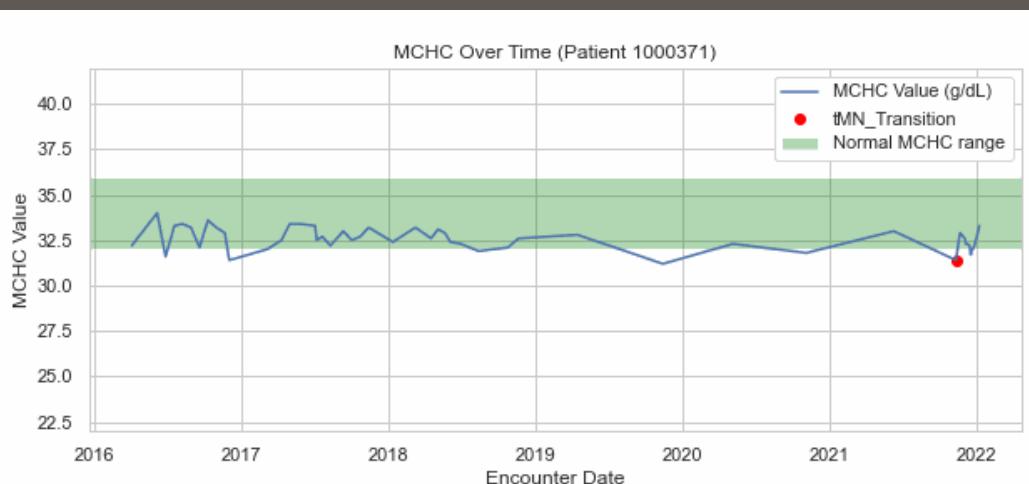
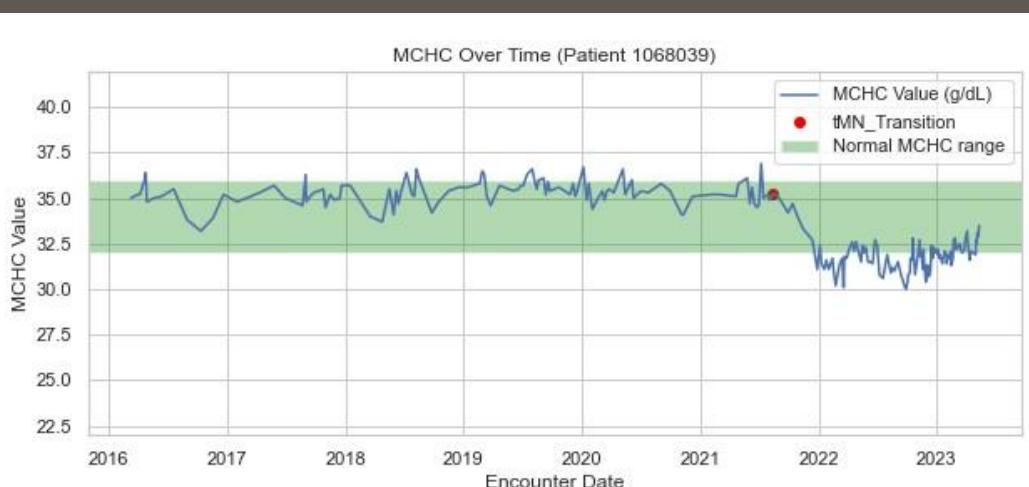
- Statistical meaningful larger values dated back 5 years before the tMN transition
- Mean in tMN patients is slightly higher than the control group average. But could be hard to capture the pattern by itself and need the help from other parameters

➤ Fluctuation Probable Causes:

- Chemotherapy-Induced Anemia: Decreased production of red blood cells and subsequently lower MCHC levels.
- Chronic Inflammation: Cancer and its treatment could induce chronic inflammation, affecting the balance of red blood cell components and leading to changes in MCHC.

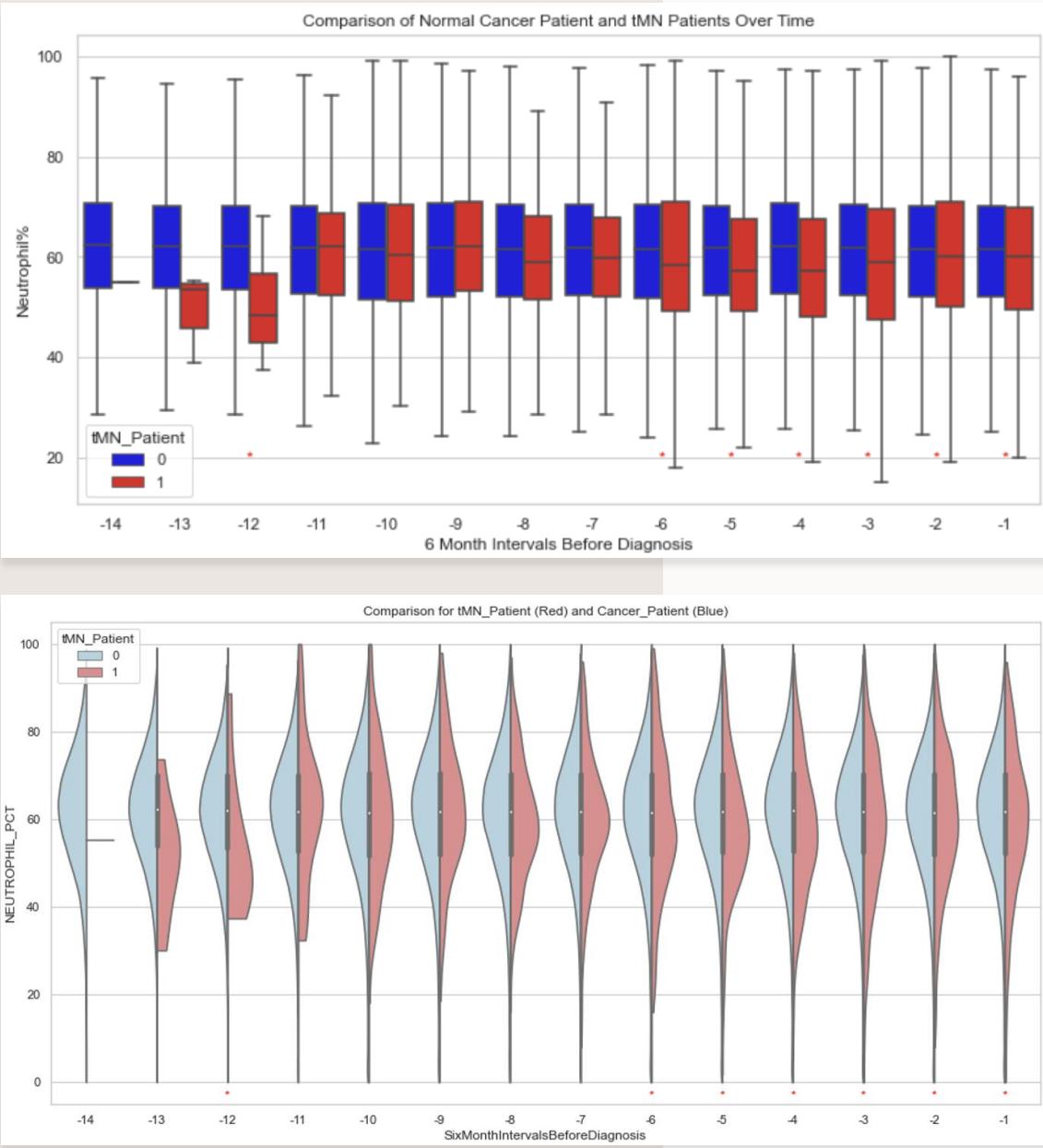
➤ Conclusion:

- Strong performance indicator but need to combine with other parameters for further analysis



NEUTROPHIL %

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.165672e+09	2.639058e-03	0.00	significant
1	-2	1.044556e+09	1.463180e-02	0.01	significant
2	-3	8.445476e+08	2.692314e-06	0.00	significant
3	-4	6.673539e+08	2.819599e-09	0.00	significant
4	-5	5.061795e+08	4.785350e-06	0.00	significant
5	-6	5.649542e+08	1.577526e-02	0.02	significant
6	-7	3.251550e+08	9.411713e-02	0.09	not significant
7	-8	2.082848e+08	5.630662e-02	0.06	not significant
8	-9	1.983807e+08	4.691894e-01	0.47	not significant
9	-10	1.511297e+08	6.711630e-01	0.67	not significant
10	-11	7.976356e+07	8.265141e-01	0.83	not significant
11	-12	8.315218e+06	2.381855e-02	0.02	significant
12	-13	5.188392e+06	6.330296e-02	0.06	not significant
13	-14	7.797150e+05	5.167454e-01	0.52	not significant



NEUT% – General Trend

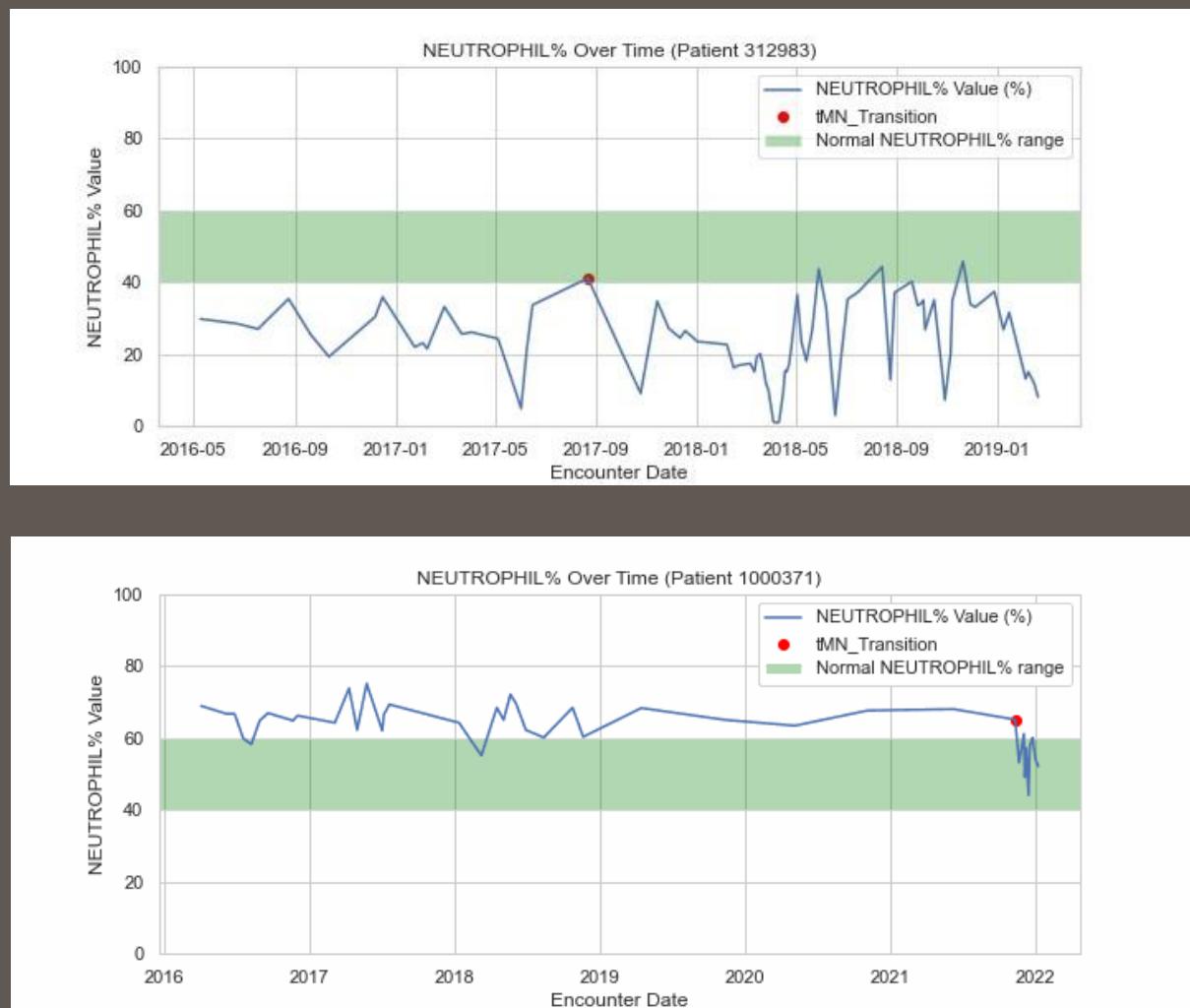
tMN patients:

- NEUT% level is generally lower dated back to 3 years ago
- Incorporate noticeable trend change in the -2 and -1 time interval

Significance Fluctuations:

- General presence of outliers that lead to extended tails in all time intervals
- Trend are constantly fluctuating and not in a stable state

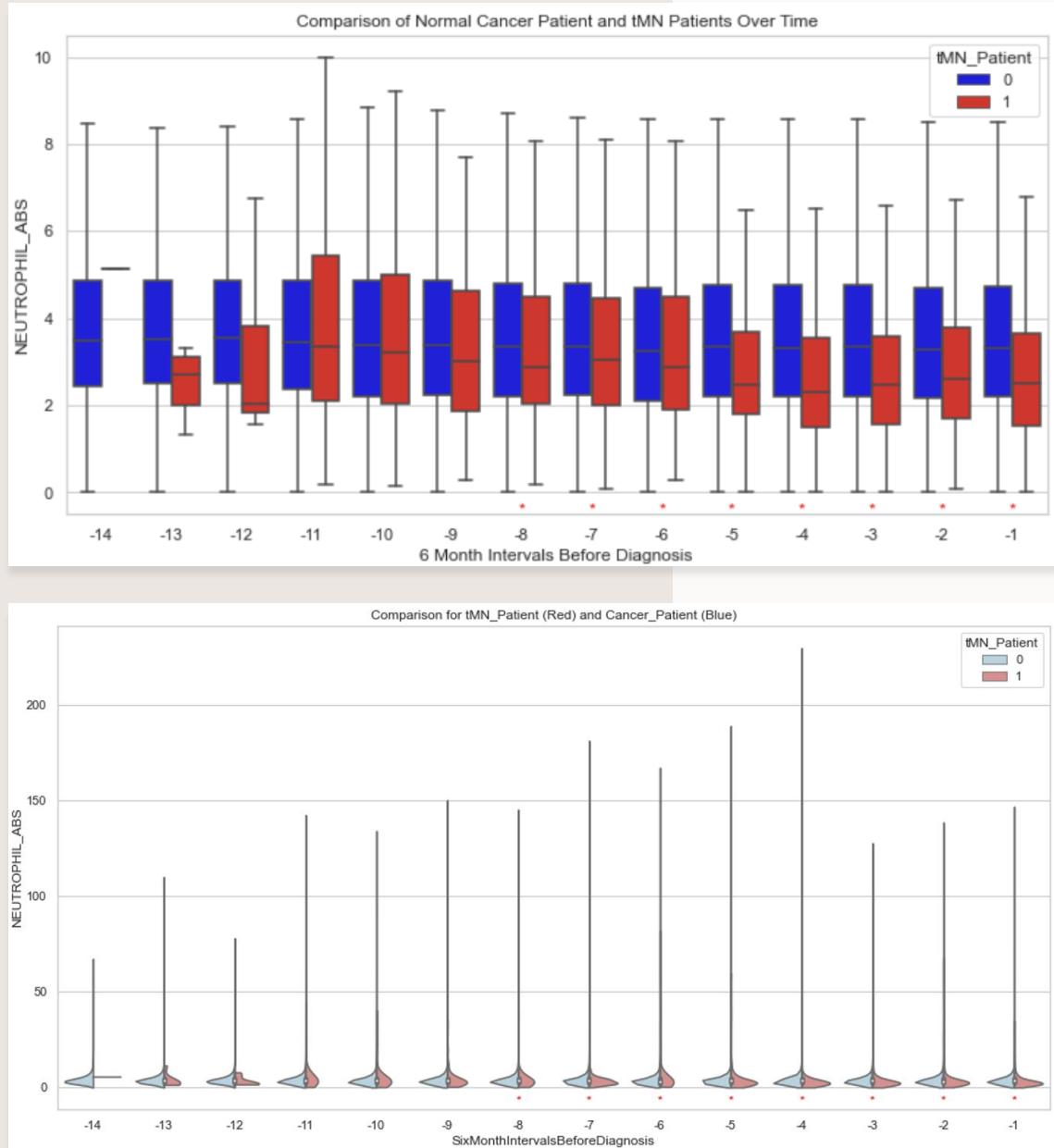
NEUT% - Individual Graphs



- Overall trend:
 - In general, statistically meaningful lower values dated back 3 years before the tMN transition
 - When it comes to -2 and -1 time interval, trend fluctuates and reflects a slope of increasing
- Fluctuation Probable Causes:
 - Medication Effects: Certain medications can alter NEUT% levels by directly or indirectly affecting neutrophil production or survival, leading to fluctuations.
 - Infections and Sepsis: Cancer patients are at an increased risk of infections, and episodes of infection or sepsis can lead to rapid changes in neutrophil percentages as the body responds to the infection.
- Conclusion:
 - Would not be an optimal indicator

NEUTROPHIL ABS

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.165672e+09	2.639058e-03	0.00	significant
1	-2	1.044556e+09	1.463180e-02	0.01	significant
2	-3	8.445476e+08	2.692314e-06	0.00	significant
3	-4	6.673539e+08	2.819599e-09	0.00	significant
4	-5	5.061795e+08	4.785350e-06	0.00	significant
5	-6	5.649542e+08	1.577526e-02	0.02	significant
6	-7	3.251550e+08	9.411713e-02	0.09	not significant
7	-8	2.082848e+08	5.630662e-02	0.06	not significant
8	-9	1.983807e+08	4.691894e-01	0.47	not significant
9	-10	1.511297e+08	6.711630e-01	0.67	not significant
10	-11	7.976356e+07	8.265141e-01	0.83	not significant
11	-12	8.315218e+06	2.381855e-02	0.02	significant
12	-13	5.188392e+06	6.330296e-02	0.06	not significant
13	-14	7.797150e+05	5.167454e-01	0.52	not significant



NEUT ABS – General Trend

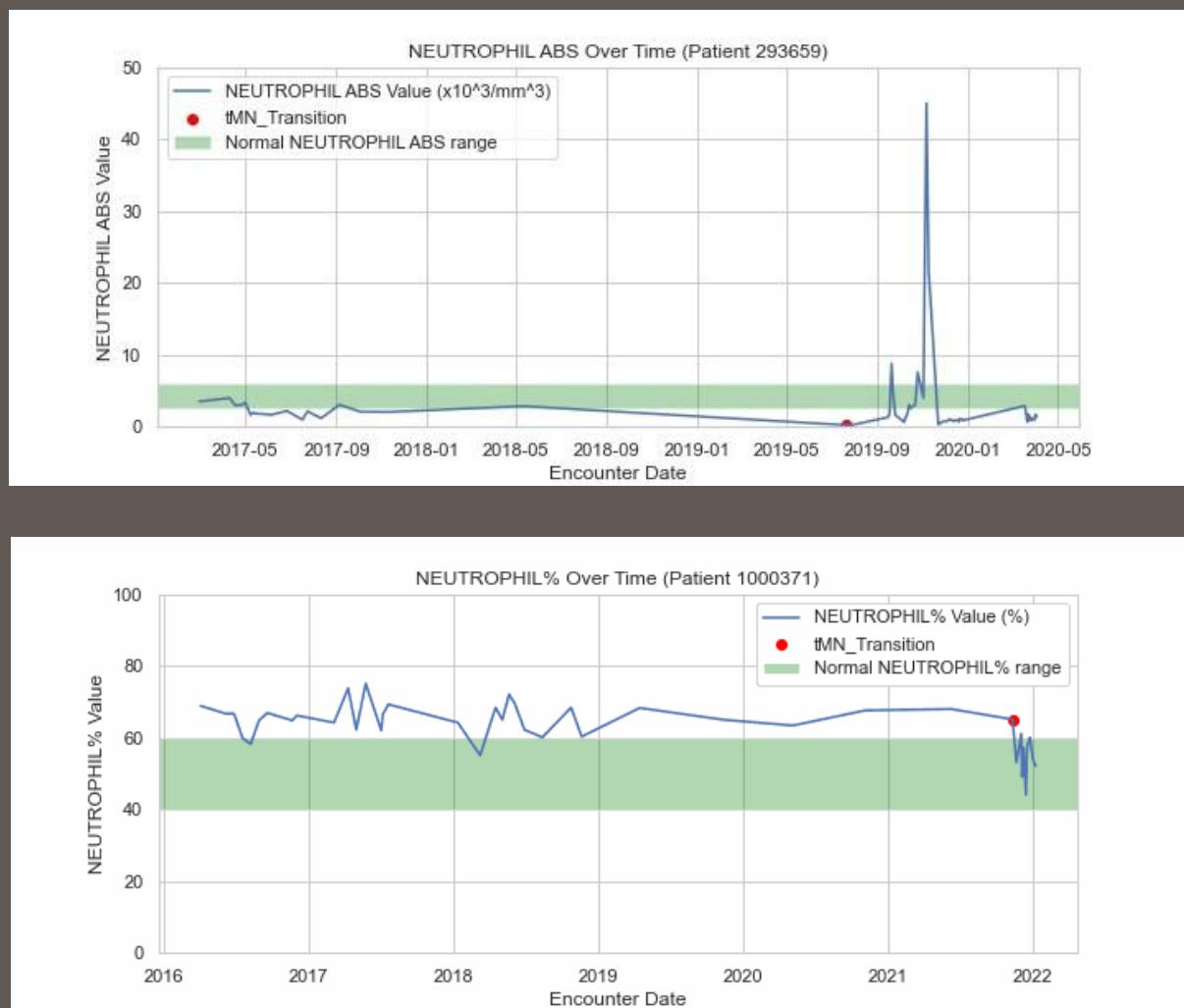
tMN patients:

- NEUT ABS level is statistically significant lower and consistently decreasing dated back to 4 years ago
- Better and more statistically stable indicator than NEUT%

Significance Fluctuations:

- General presence of outliers that lead to extended tails in all time intervals, especially in the -4 time interval
- Time-interval data distribution are generally left skewed with large value outliers

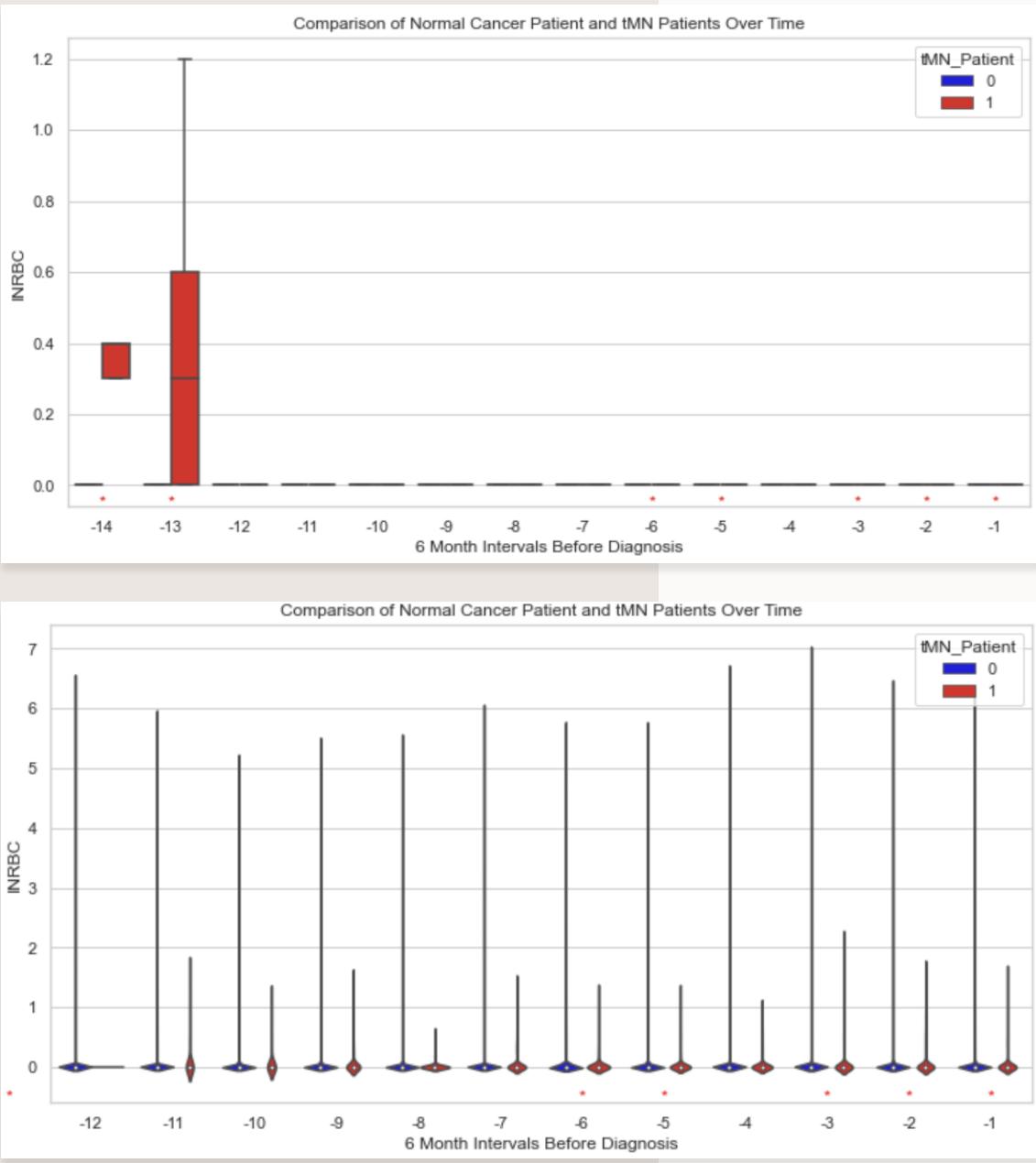
NEUT ABS - Individual Graphs



- Overall trend:
 - Statistical lower values and decreasing trend dated back 4 years before the tMN transition
- Fluctuation Probable Causes:
 - Medication Effects
 - Infections and Sepsis
 - Treatment Effects
 - ❖ Positive response to treatment that influence body's overall immune function and consequently reflected in neutrophil percentages.
- Conclusion:
 - Better performance than NEUT% but still highly fluctuates and thus need helps from other parameters

INRBC

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	974206687.5	6.388418e-07	0.00	significant
1	-2	810063223.5	1.207227e-08	0.00	significant
2	-3	713925195.5	1.102017e-03	0.00	significant
3	-4	579716439.5	5.502792e-02	0.06	not significant
4	-5	468672769.0	4.539125e-02	0.05	significant
5	-6	411326131.0	1.055792e-02	0.01	significant
6	-7	269455213.0	5.662785e-01	0.57	not significant
7	-8	173828573.0	7.475327e-01	0.75	not significant
8	-9	177225193.5	7.711553e-01	0.77	not significant
9	-10	109143557.0	8.435793e-02	0.08	not significant
10	-11	68259181.5	1.499791e-01	0.15	not significant
11	-12	10303474.5	2.963389e-01	0.30	not significant
12	-13	19345193.5	2.330700e-08	0.00	significant
13	-14	6723331.5	1.075591e-05	0.00	significant



INRBC – General Trend

tMN patients:

- INRBC level is statistically significant lower and consistently decreasing dated back to 4 years ago
- Better and more statistically stable indicator than NEUT%

Significance Fluctuations:

- General presence of outliers that lead to extended tails in all time intervals, especially in the -4 time interval
- Time-interval data distribution are generally left skewed with large value outliers

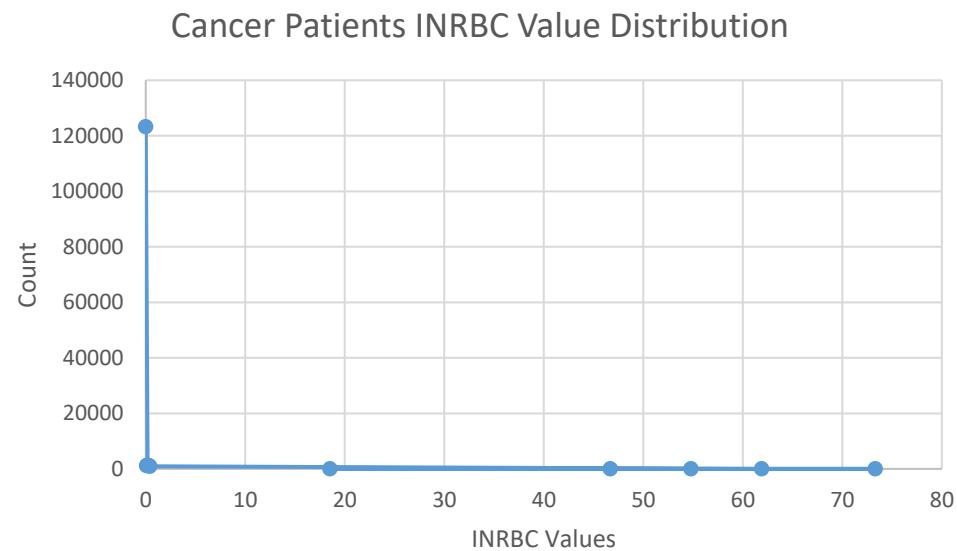
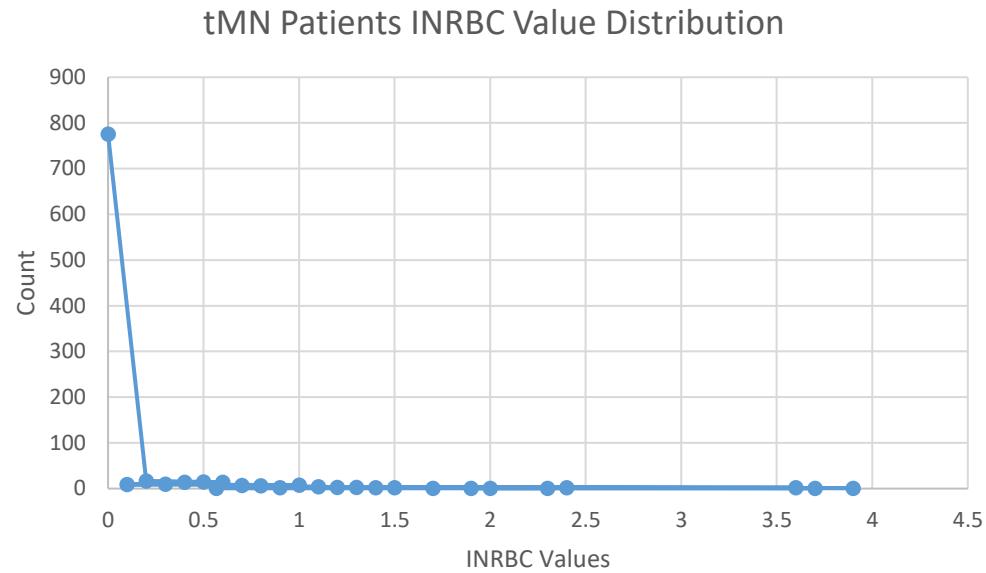
INRBC – Data Distribution

tMN patients:

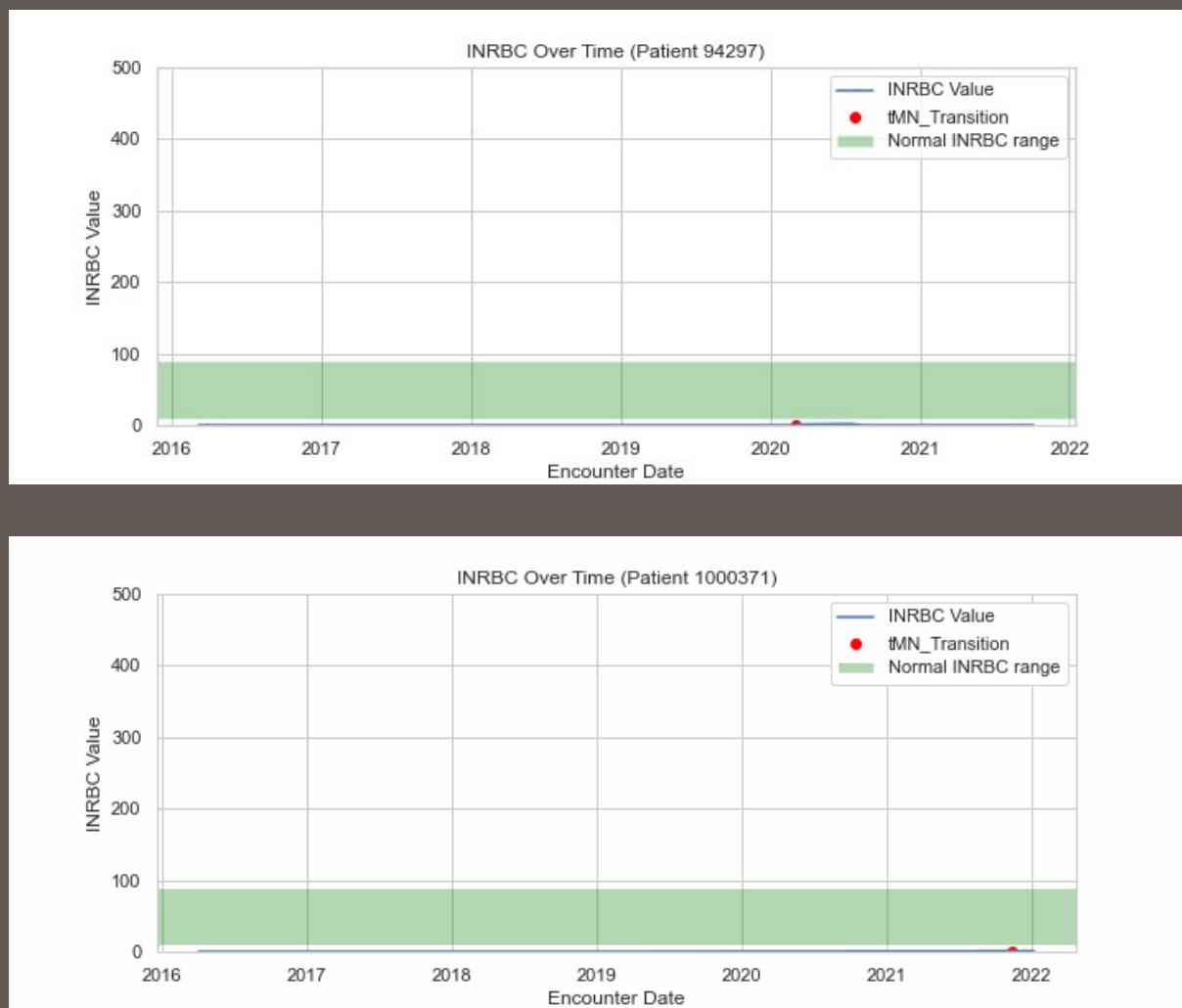
- Value are generally smaller and more focus

Cancer Patients:

- Contain spread out outliers and thus influence the general graph results when plotting these two groups side by side



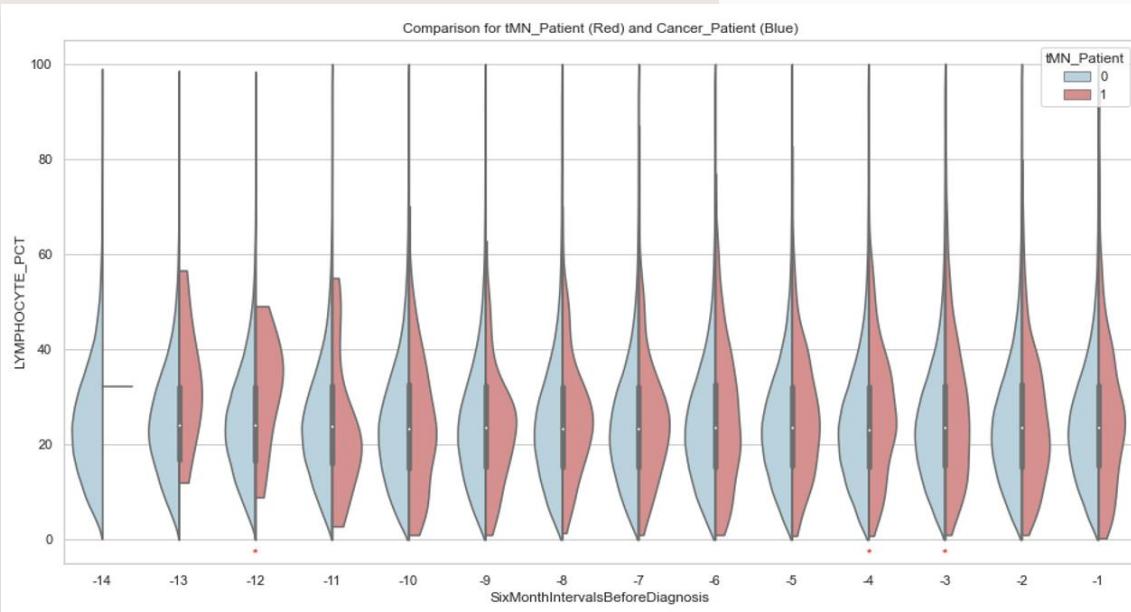
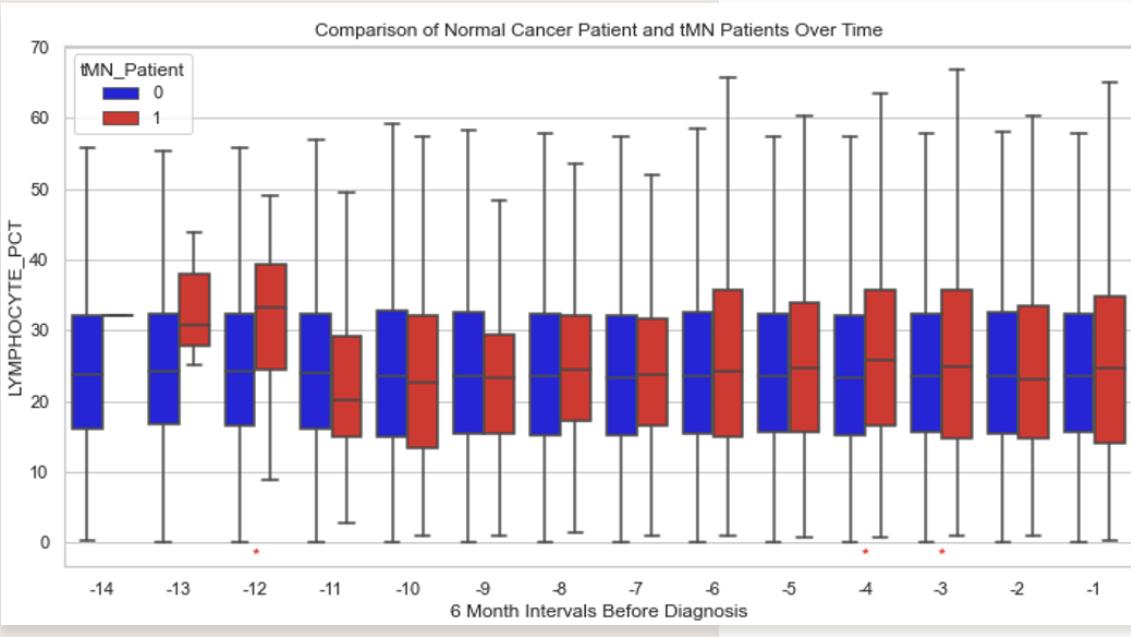
INRBC - Individual Graphs



- Overall trend:
 - Statistical lower values close to zeros trend dated back 1.5 years before the tMN transition
- Data Pattern:
 - Zero values for INRBC may indicate results within the normal range, potentially rounded down due to the laboratory's practices or limitations in the testing method's detection sensitivity.
- Fluctuation Probable Causes:
 - Treatment Side Effects: Certain chemotherapy and radiation therapy, may directly impact the bone marrow, leading to an increased production of immature blood cells.
 - Bone Marrow Infiltration: Cancer cells infiltrating the bone marrow can disrupt normal blood cell production, resulting in an elevated count of immature red blood cells.
- Conclusion:
 - Need to work with other parameters

LYMPHOCYTE %

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.250002e+09	0.403095	0.40	not significant
1	-2	1.070585e+09	0.294153	0.29	not significant
2	-3	9.733175e+08	0.037294	0.04	significant
3	-4	8.371502e+08	0.000343	0.00	significant
4	-5	5.969630e+08	0.156916	0.16	not significant
5	-6	6.160483e+08	0.390862	0.39	not significant
6	-7	3.465377e+08	0.885024	0.89	not significant
7	-8	2.368393e+08	0.288713	0.29	not significant
8	-9	1.798232e+08	0.183040	0.18	not significant
9	-10	1.439820e+08	0.320697	0.32	not significant
10	-11	6.859013e+07	0.190853	0.19	not significant
11	-12	1.836644e+07	0.049904	0.05	significant
12	-13	1.179675e+07	0.104912	0.10	not significant
13	-14	1.853772e+06	0.396796	0.40	not significant



LYMPH% – General Trend

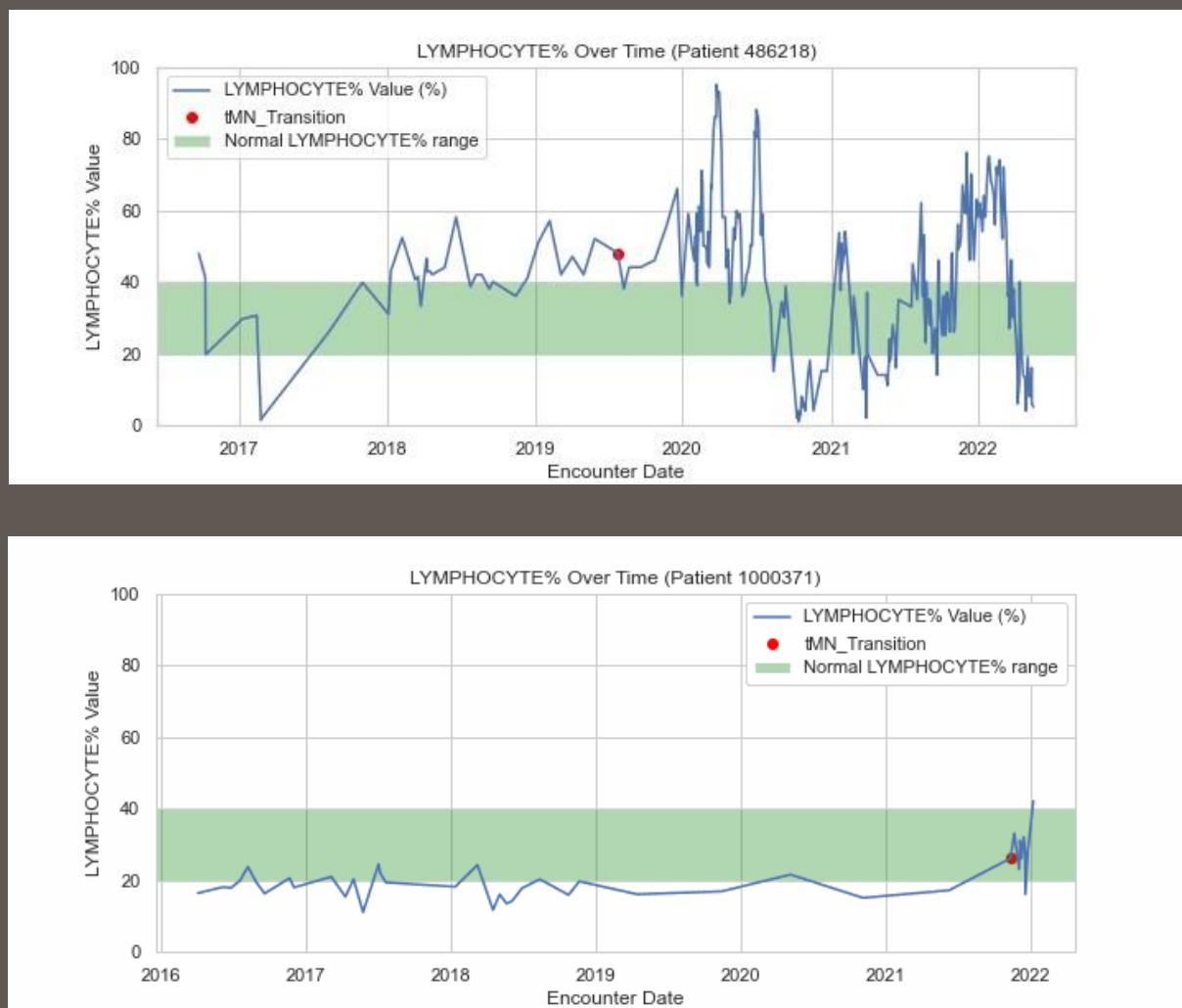
tMN patients:

- LYMPH% level is statistically significant larger at time interval -3 and -4, but not a consistently trend

Significance Fluctuations:

- General presence of outliers that lead to extended tails in all time intervals
- Time-interval data distribution are generally left skewed with large value outliers.

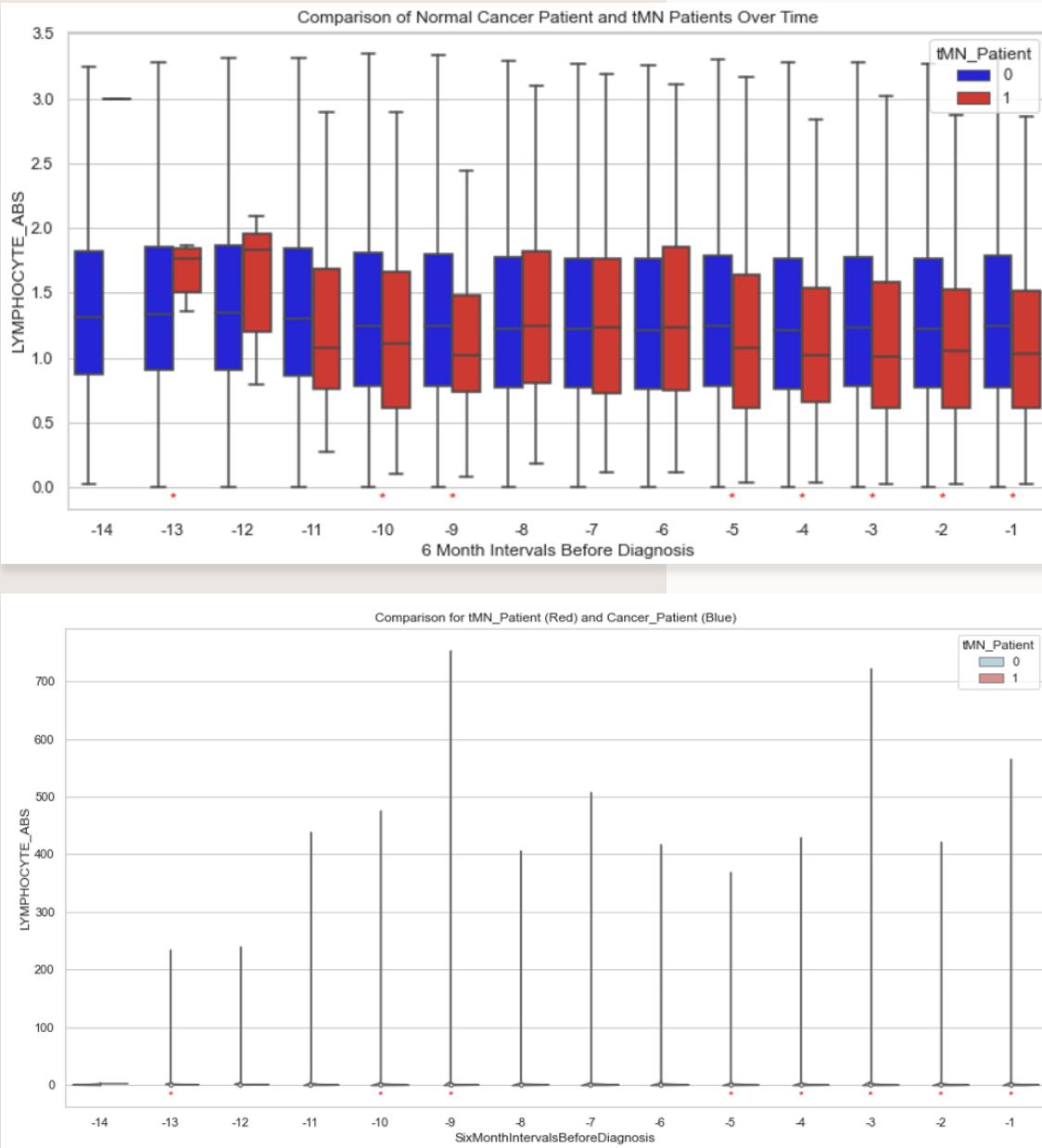
LYMPH% - Individual Graphs



- Overall trend:
 - Statistical higher values 1 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Treatment Effects: Cancer treatments, such as chemotherapy or radiation therapy, can have direct effects on the immune system, potentially influencing the LYMPH% in blood test results.
 - Inflammatory Response: Cancer is often associated with inflammation, and an inflammatory response can affect the distribution of white blood cells, including lymphocytes.
- Conclusion:
 - Since both groups perform similarly, hard to separate by the parameter itself

LYMPHOCYTE ABS

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	996396322.5	5.184495e-25	0.00	significant
1	-2	917150952.0	1.587296e-16	0.00	significant
2	-3	790528688.5	5.740056e-13	0.00	significant
3	-4	649523598.5	5.503206e-12	0.00	significant
4	-5	496562671.0	3.754284e-07	0.00	significant
5	-6	597694015.5	7.642552e-01	0.76	not significant
6	-7	331711244.0	2.744704e-01	0.27	not significant
7	-8	229731017.0	7.417220e-01	0.74	not significant
8	-9	161576847.0	7.356008e-04	0.00	significant
9	-10	135672578.5	4.149762e-02	0.04	significant
10	-11	70515000.5	3.349399e-01	0.33	not significant
11	-12	17599151.5	1.011237e-01	0.10	not significant
12	-13	12442726.0	4.979542e-02	0.05	significant
13	-14	2372877.5	1.164354e-01	0.12	not significant



LYMPH ABS – General Trend

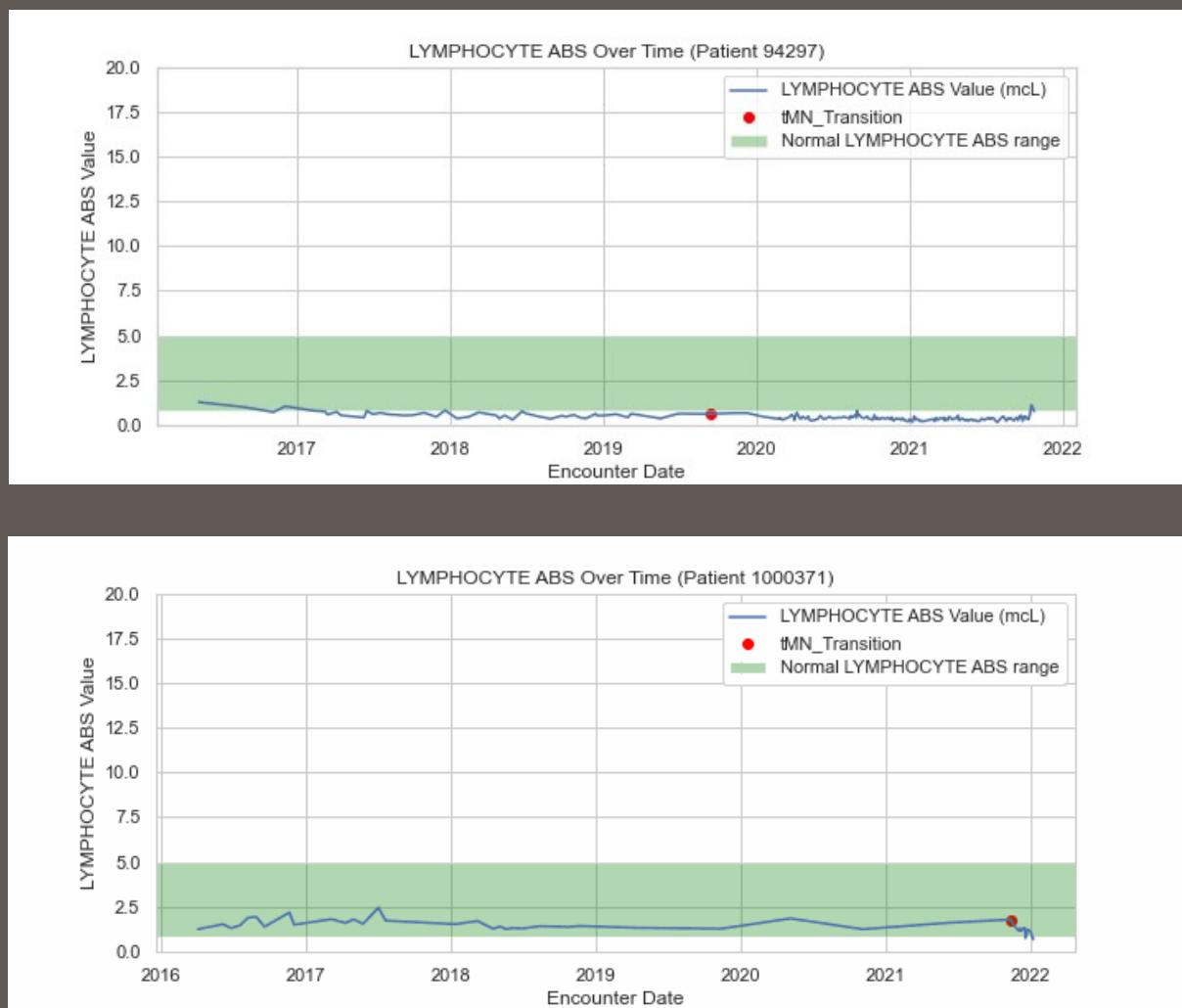
tMN patients:

- LYMPH ABS level is statistically significant lower dated back to 2.5 years

Data Patterns:

- Outliers concentrate in time interval -9 and -3 for non-tMNs group
- Time-interval data distribution are generally left skewed with large value outliers
- Too many large outliers compressed the violin graph

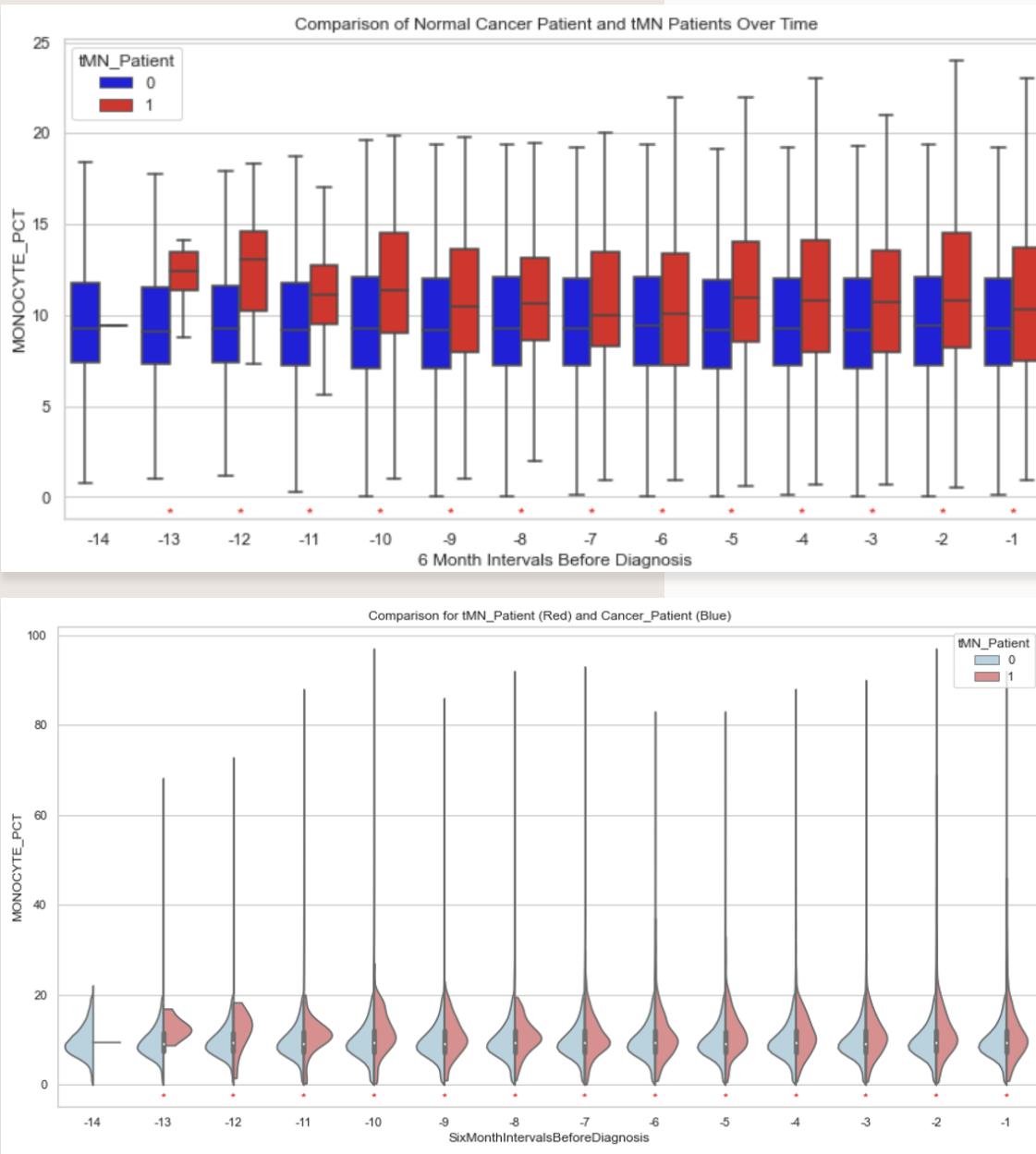
LYMPH ABS - Individual Graphs



- Overall trend:
 - Statistical lower values 2.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Treatment Effects: Cancer treatments, such as chemotherapy or radiation therapy, can have direct effects on the immune system, potentially influencing the LYMPH% in blood test results.
 - Inflammatory Response: Cancer is often associated with inflammation, and an inflammatory response can affect the distribution of white blood cells, including lymphocytes.
- Conclusion:
 - Perform better than LYMPH% but not the statistically strong factors compared to other parameters.

MONOCYTE %

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.357085e+09	2.817918e-11	0.00	significant
1	-2	1.294035e+09	5.021241e-25	0.00	significant
2	-3	1.072713e+09	1.119558e-14	0.00	significant
3	-4	9.051319e+08	5.512878e-16	0.00	significant
4	-5	6.886121e+08	3.207076e-16	0.00	significant
5	-6	6.487931e+08	8.429347e-05	0.00	significant
6	-7	3.851465e+08	4.399683e-05	0.00	significant
7	-8	2.621533e+08	2.966494e-05	0.00	significant
8	-9	2.155933e+08	2.451621e-03	0.00	significant
9	-10	1.899139e+08	2.342690e-07	0.00	significant
10	-11	9.124633e+07	1.867134e-03	0.00	significant
11	-12	1.853129e+07	3.621247e-02	0.04	significant
12	-13	1.289810e+07	2.400250e-02	0.02	significant
13	-14	1.266362e+06	9.644026e-01	0.96	not significant



MONO% – General Trend

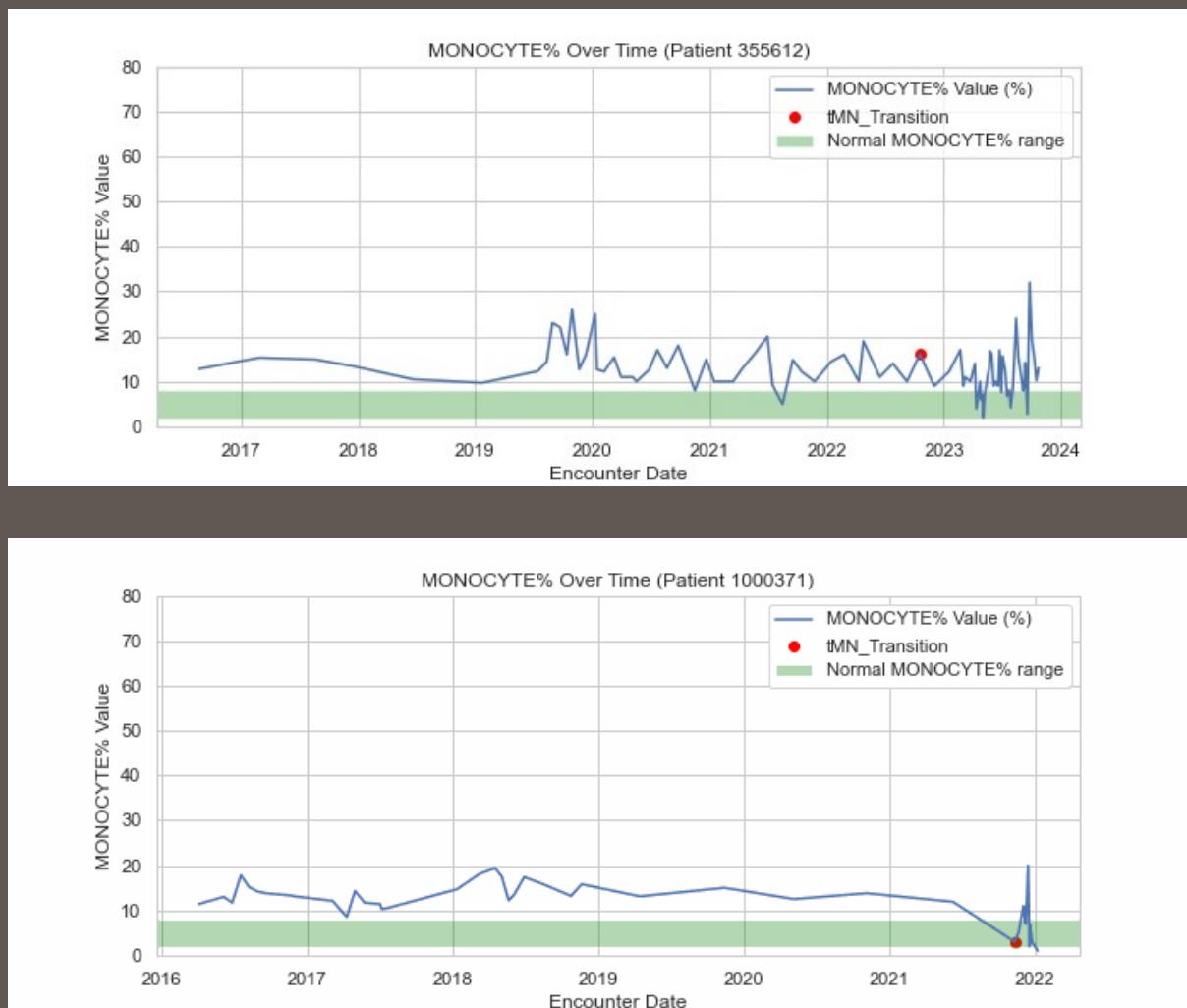
tMN patients:

- MONO% level is statistically significant larger dated back to 5-6 years

Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers

MONO% - Individual Graphs



➤ Overall trend:

- Statistical higher values 5.5 year before the tMN transition but stays

➤ Fluctuation Probable Causes:

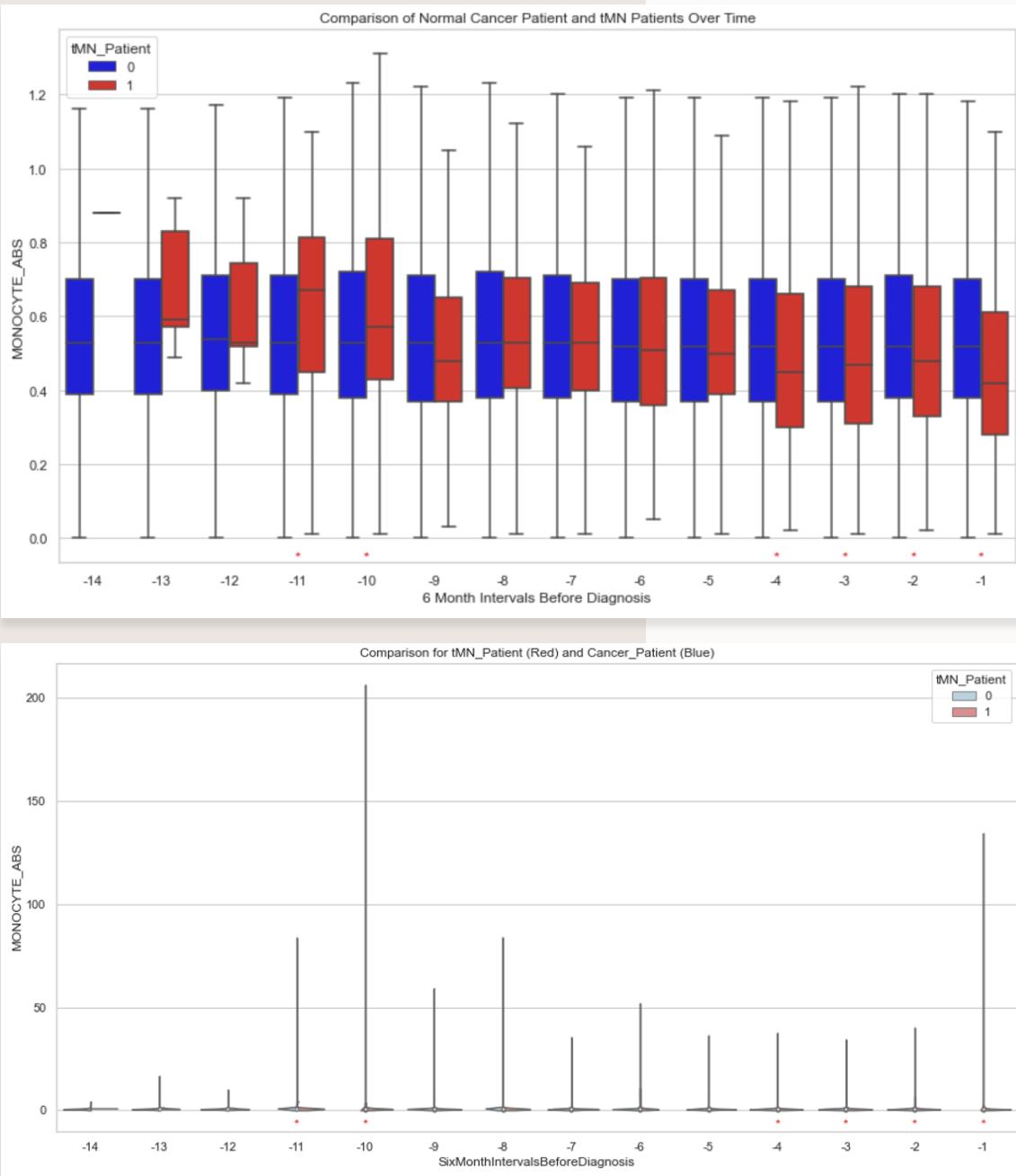
- Chemotherapy Medications: Some chemotherapy drugs may suppress monocyte production, while others may lead to an increase in monocyte levels.
- Inflammation and Infection: cancer patients may be more susceptible to infections, which can also affect the MONO% in the blood.
- Cancer Type and Stage: Different types and stages of cancer can influence the immune response, leading to fluctuations in monocyte levels.

➤ Conclusion:

- Perform better than LYMPH% but not the statistically strong factors compared to other parameters.

MONOCYTE ABS

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	962100438.0	3.910765e-28	0.00	significant
1	-2	987397590.5	2.349831e-05	0.00	significant
2	-3	820224914.5	1.697086e-07	0.00	significant
3	-4	653300367.5	9.416255e-10	0.00	significant
4	-5	538780613.5	9.623388e-02	0.10	not significant
5	-6	581102594.0	6.771192e-01	0.68	not significant
6	-7	337527951.5	9.637534e-01	0.96	not significant
7	-8	229950214.5	4.171891e-01	0.42	not significant
8	-9	174458970.0	1.022251e-01	0.10	not significant
9	-10	168501195.0	1.463224e-02	0.01	significant
10	-11	90202887.5	3.459994e-03	0.00	significant
11	-12	15390179.0	4.433620e-01	0.44	not significant
12	-13	11936855.0	8.044030e-02	0.08	not significant
13	-14	2157258.5	1.953218e-01	0.20	not significant



MONO ABS – General Trend

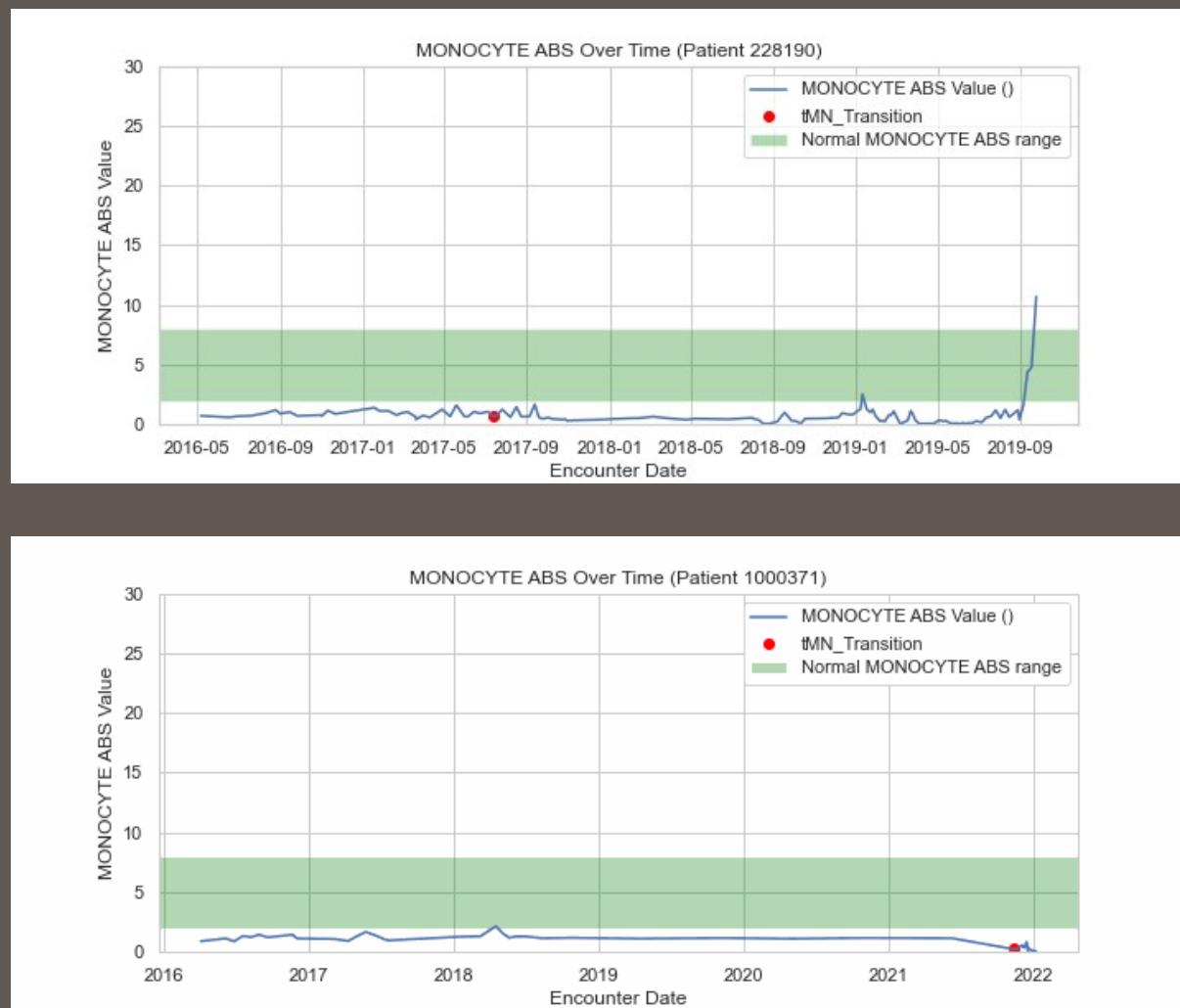
tMN patients:

- LYMPH ABS level is statistically significant lower dated back to 2 years, need a clean cut-off for the time frame
- Outliers at and before -10 time interval

Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers, especially in the -10 time interval
- Too many large outliers compressed the violin graph

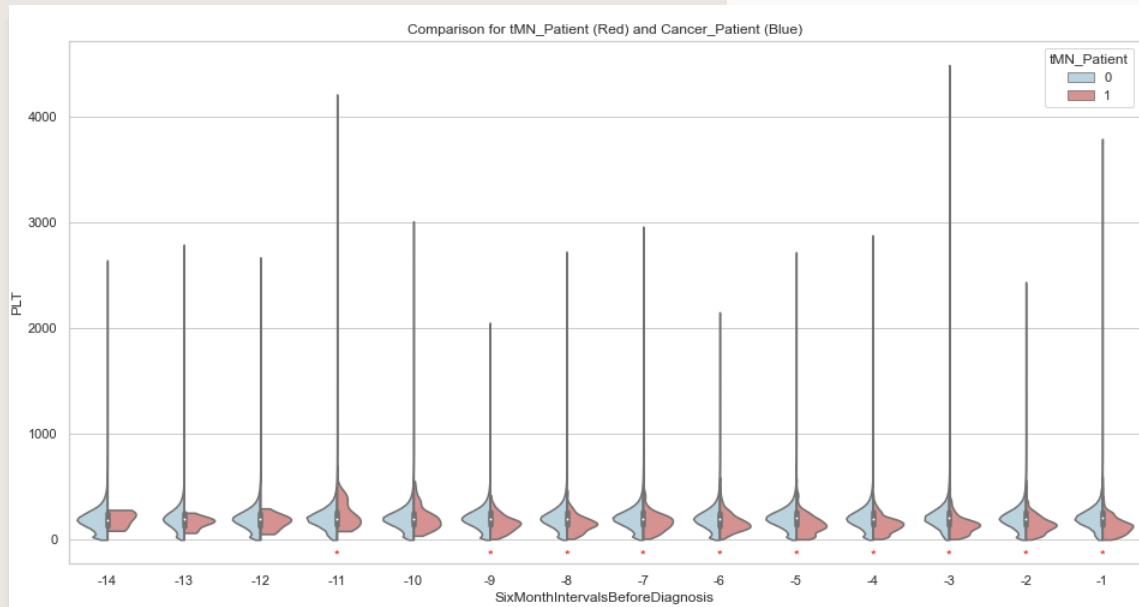
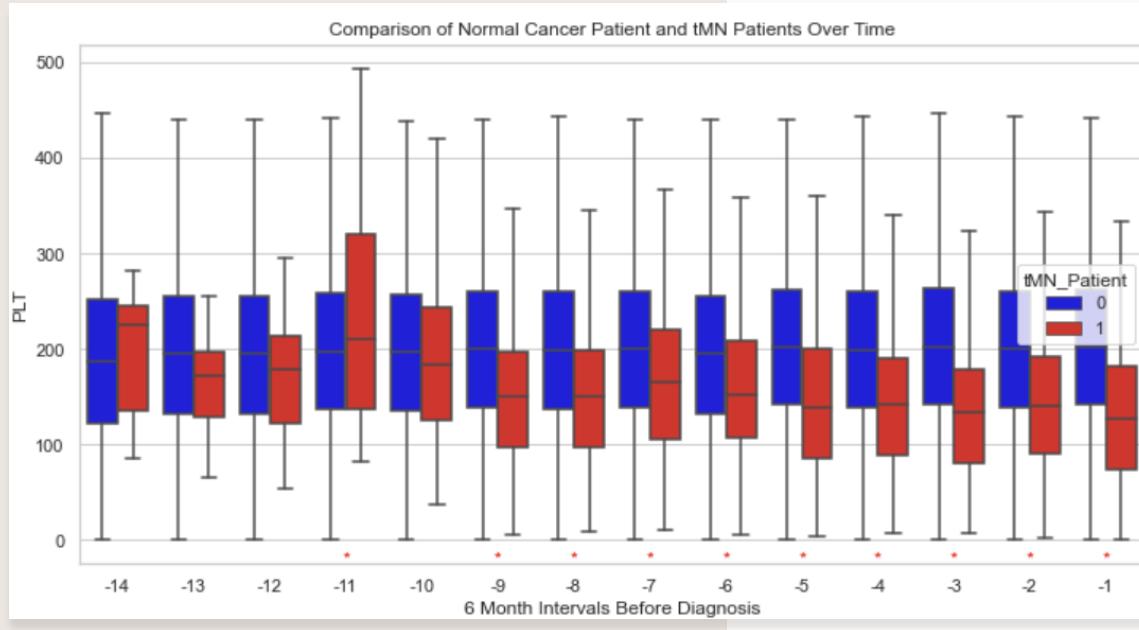
MONO ABS - Individual Graphs



- Overall trend:
 - Statistical lower values 2.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Treatment Effects: Cancer treatments, such as chemotherapy or radiation therapy, can have direct effects on the immune system, potentially influencing the LYMPH% in blood test results.
 - Inflammatory Response: Cancer is often associated with inflammation, and an inflammatory response can affect the distribution of white blood cells, including lymphocytes.
- Conclusion:
 - Perform better than LYMPH% but not the statistically strong factors compared to other parameters.

PLT

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	943287402.0	7.084383e-132	0.00	significant
1	-2	917702249.5	9.035808e-82	0.00	significant
2	-3	721178868.0	3.082969e-87	0.00	significant
3	-4	652245071.5	2.527043e-61	0.00	significant
4	-5	562746200.0	2.929608e-45	0.00	significant
5	-6	586643015.0	5.141659e-23	0.00	significant
6	-7	350073575.5	8.293898e-11	0.00	significant
7	-8	226656256.5	2.005612e-15	0.00	significant
8	-9	216207568.5	1.797498e-14	0.00	significant
9	-10	209261459.5	1.789842e-01	0.18	not significant
10	-11	142787949.0	2.616990e-02	0.03	significant
11	-12	12015900.0	2.861883e-01	0.29	not significant
12	-13	14063180.5	1.036115e-01	0.10	not significant
13	-14	6735387.0	9.869067e-01	0.99	not significant



PLT – General Trend

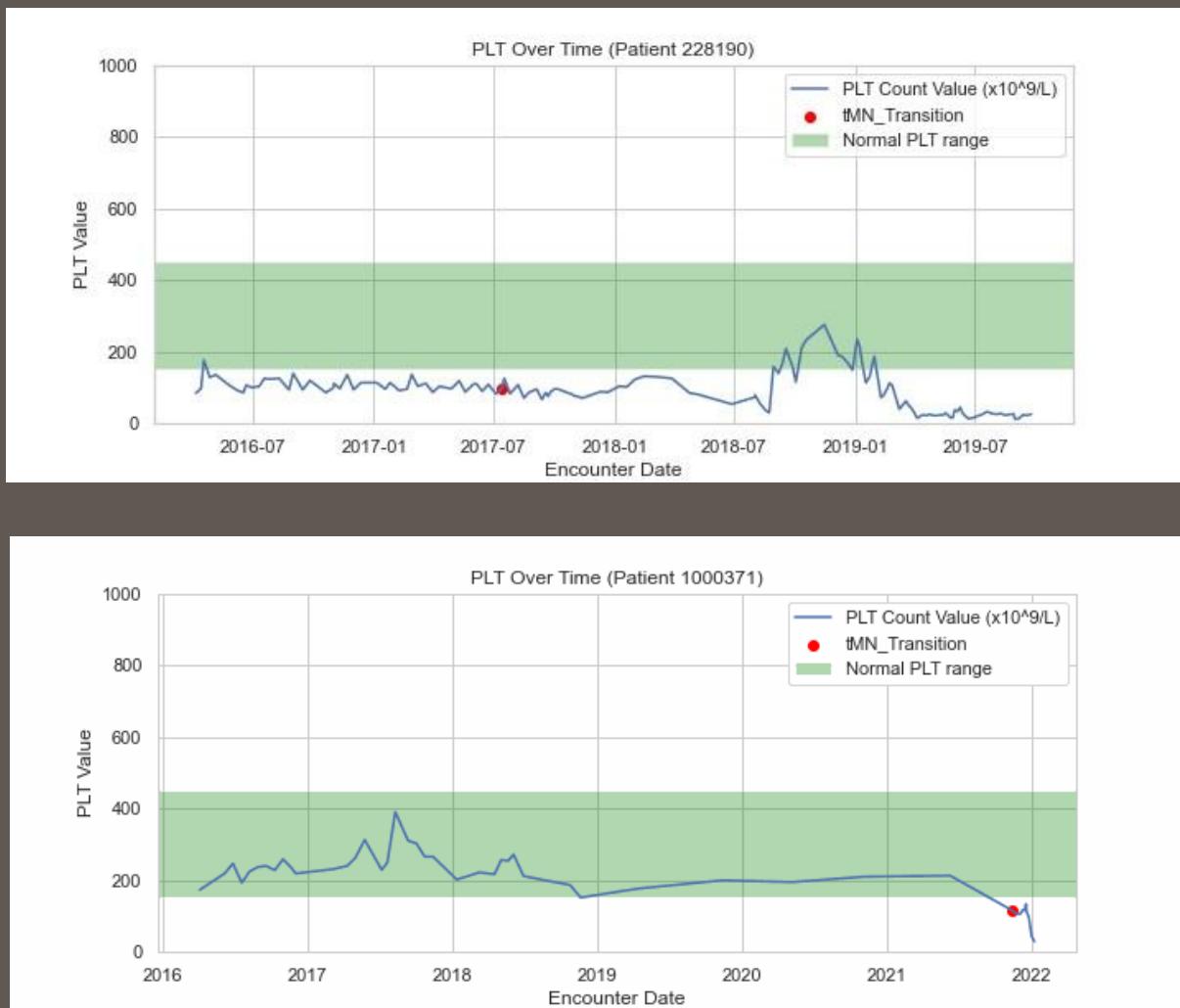
tMN patients:

- PLT level is statistically significant lower and decreasing dated back to 4.5 years

Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers, especially in the -3 time interval
- Multi-model patterns in -4 and -5 time interval for tMN Patients
- Multi-model pattern generally exist in the cancer patients for PLT values.

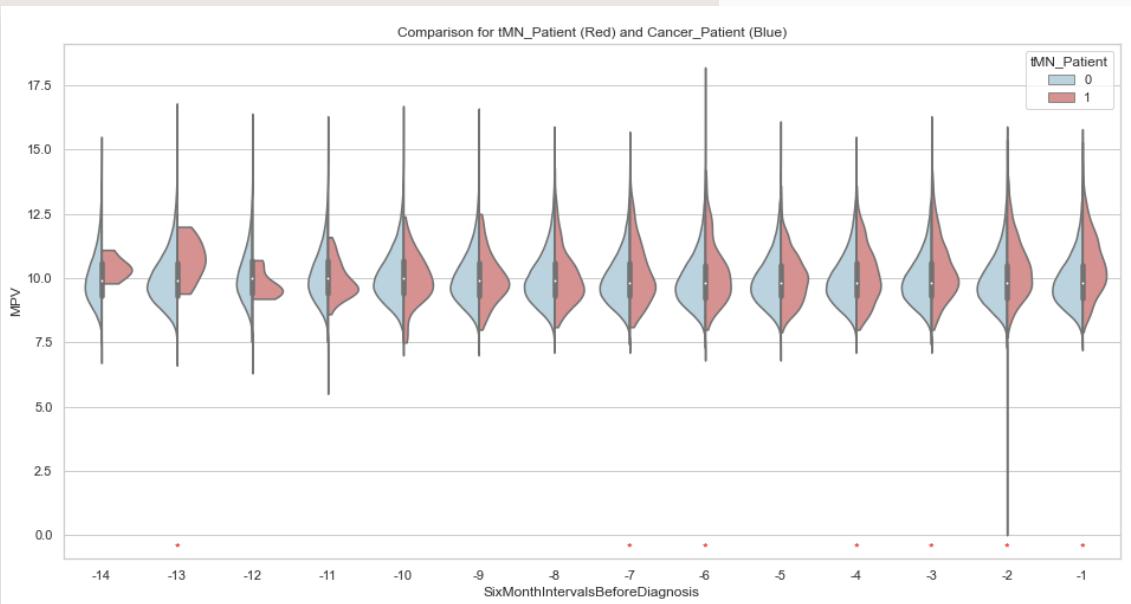
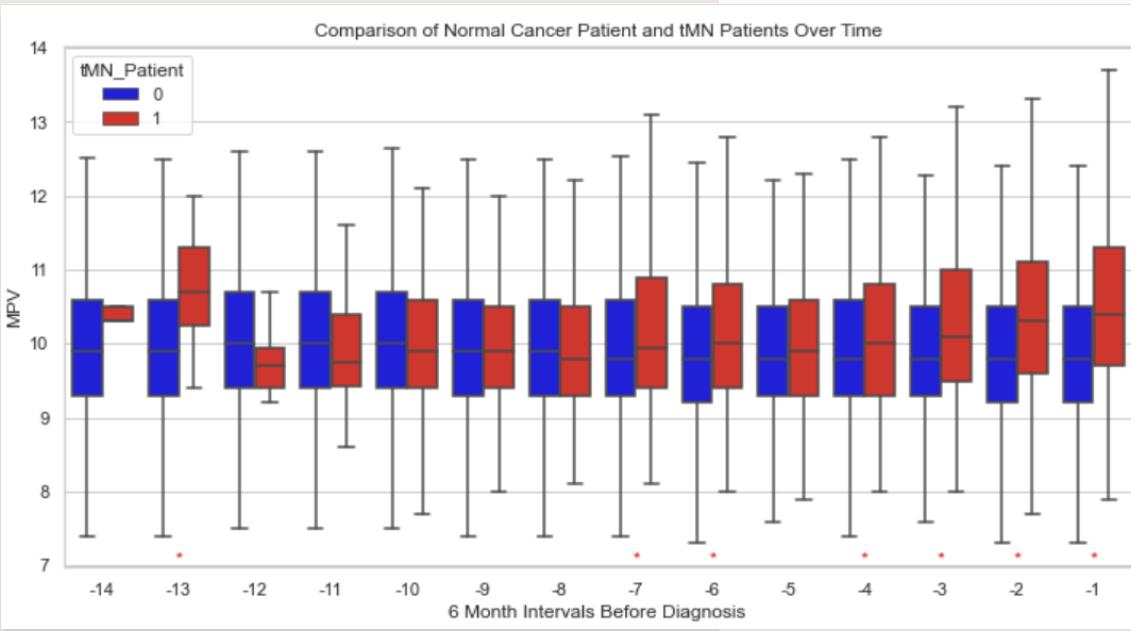
PLT - Individual Graphs



- Overall trend:
 - Statistical lower values 4.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Bone Marrow Infiltration: In some cases, cancer cells may infiltrate the bone marrow, disrupting normal blood cell production, including platelets.
 - Inflammatory Response: Cancer and its treatments can trigger inflammatory responses in the body, affecting the balance of blood components, including platelets.
- Conclusion:
 - Statistically strong predictor, but need other parameters to help with the multi-model problem

MPV

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	943287402.0	7.084383e-132	0.00	significant
1	-2	917702249.5	9.035808e-82	0.00	significant
2	-3	721178868.0	3.082969e-87	0.00	significant
3	-4	652245071.5	2.527043e-61	0.00	significant
4	-5	562746200.0	2.929608e-45	0.00	significant
5	-6	586643015.0	5.141659e-23	0.00	significant
6	-7	350073575.5	8.293898e-11	0.00	significant
7	-8	226656256.5	2.005612e-15	0.00	significant
8	-9	216207568.5	1.797498e-14	0.00	significant
9	-10	209261459.5	1.789842e-01	0.18	not significant
10	-11	142787949.0	2.616990e-02	0.03	significant
11	-12	12015900.0	2.861883e-01	0.29	not significant
12	-13	14063180.5	1.036115e-01	0.10	not significant
13	-14	6735387.0	9.869067e-01	0.99	not significant



MPV – General Trend

tMN patients:

- MPV level is statistically significant higher dated back to 2 years ago

Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers
- But -2 time interval are right skewed with extreme small values

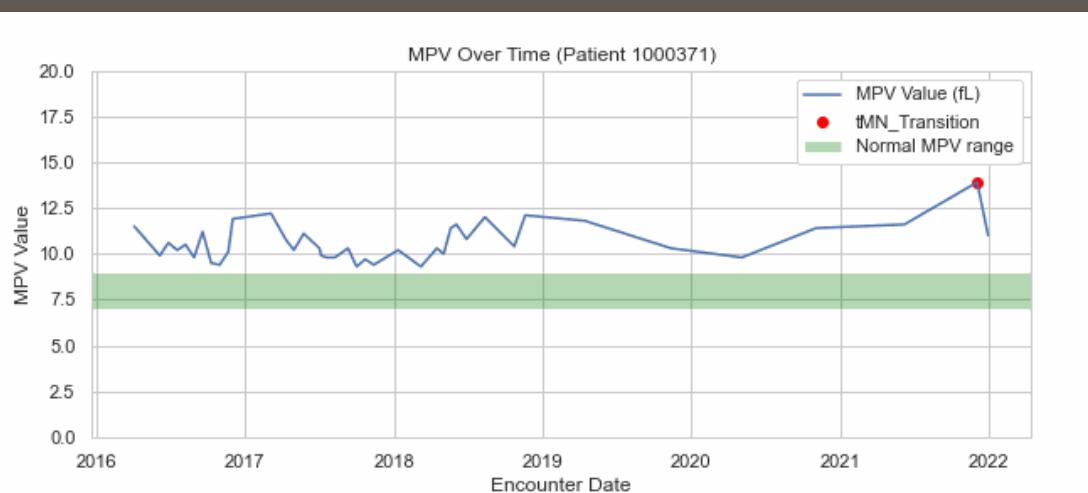
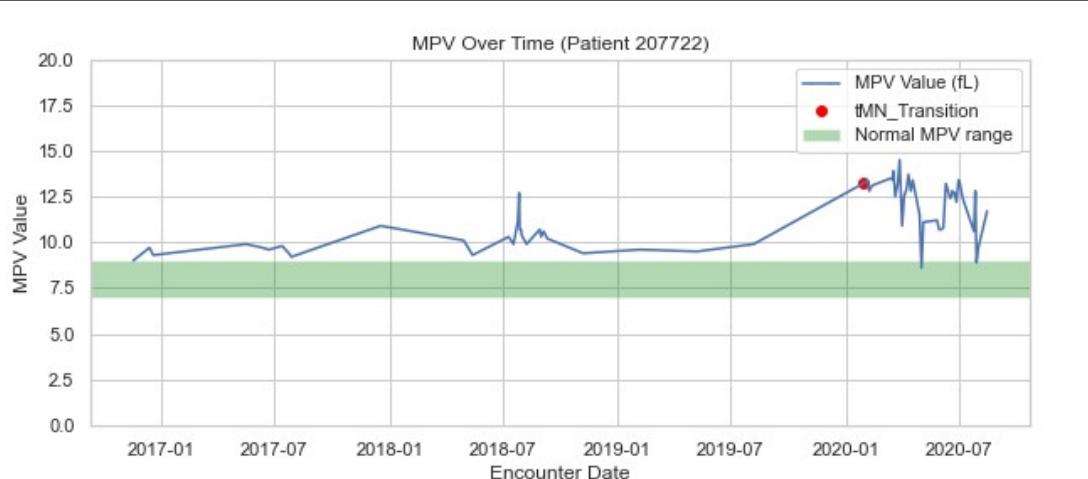
MPV - Individual Graphs

➤ Overall trend:

- Statistically significant higher values 2 year before the tMN transition
- Stable and for staying the above of control group most of the time

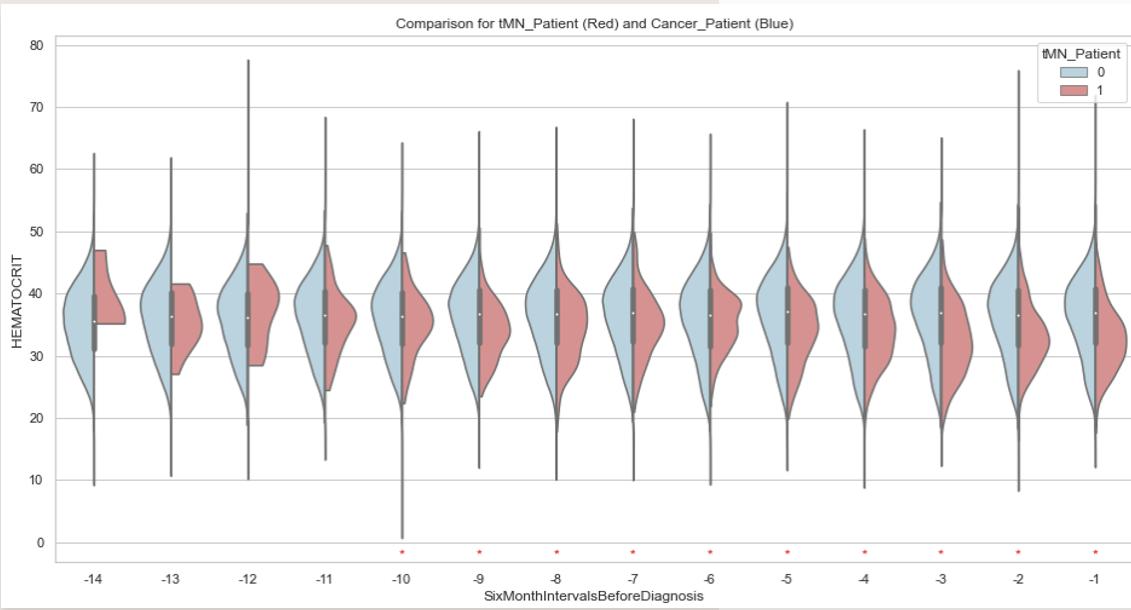
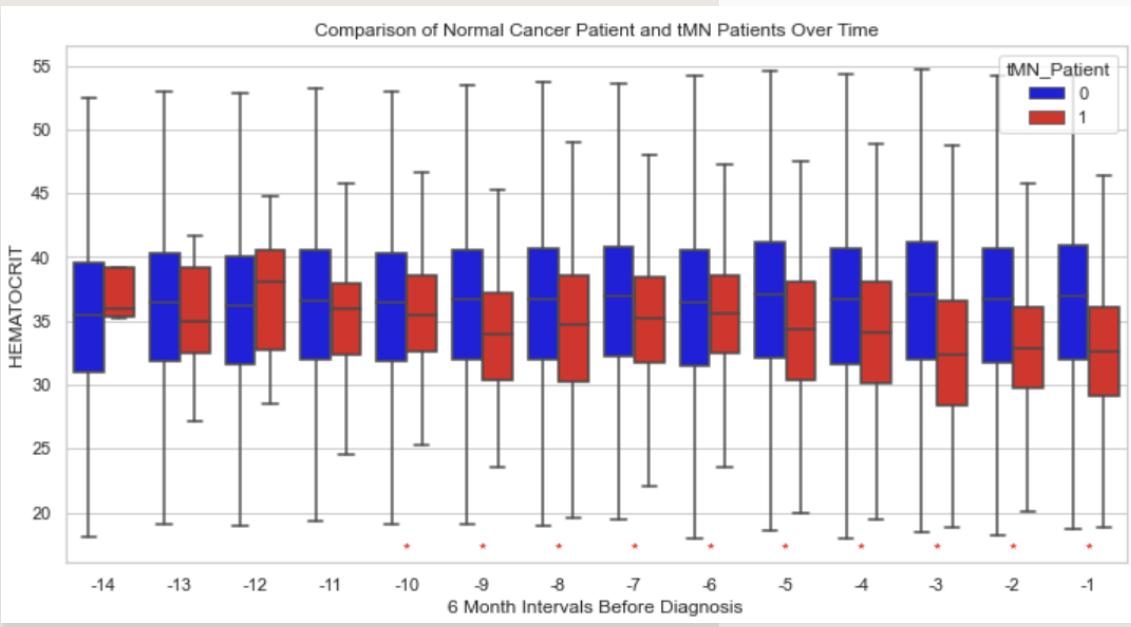
➤ Conclusion:

- Statistically strong predictor



HEMATOCRIT (HCT)

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.061194e+09	9.057116e-87	0.00	significant
1	-2	9.702961e+08	1.166947e-63	0.00	significant
2	-3	7.660627e+08	3.074351e-69	0.00	significant
3	-4	7.793637e+08	4.661585e-25	0.00	significant
4	-5	6.574009e+08	1.600426e-19	0.00	significant
5	-6	6.881454e+08	4.736281e-05	0.00	significant
6	-7	3.795775e+08	2.061915e-05	0.00	significant
7	-8	2.643728e+08	3.218270e-06	0.00	significant
8	-9	2.327515e+08	1.838927e-09	0.00	significant
9	-10	1.936406e+08	8.742779e-03	0.01	significant
10	-11	1.140230e+08	1.140396e-01	0.11	not significant
11	-12	1.583058e+07	6.675243e-01	0.67	not significant
12	-13	1.593123e+07	5.964301e-01	0.60	not significant
13	-14	6.163138e+06	6.016844e-01	0.60	not significant



HCT – General Trend

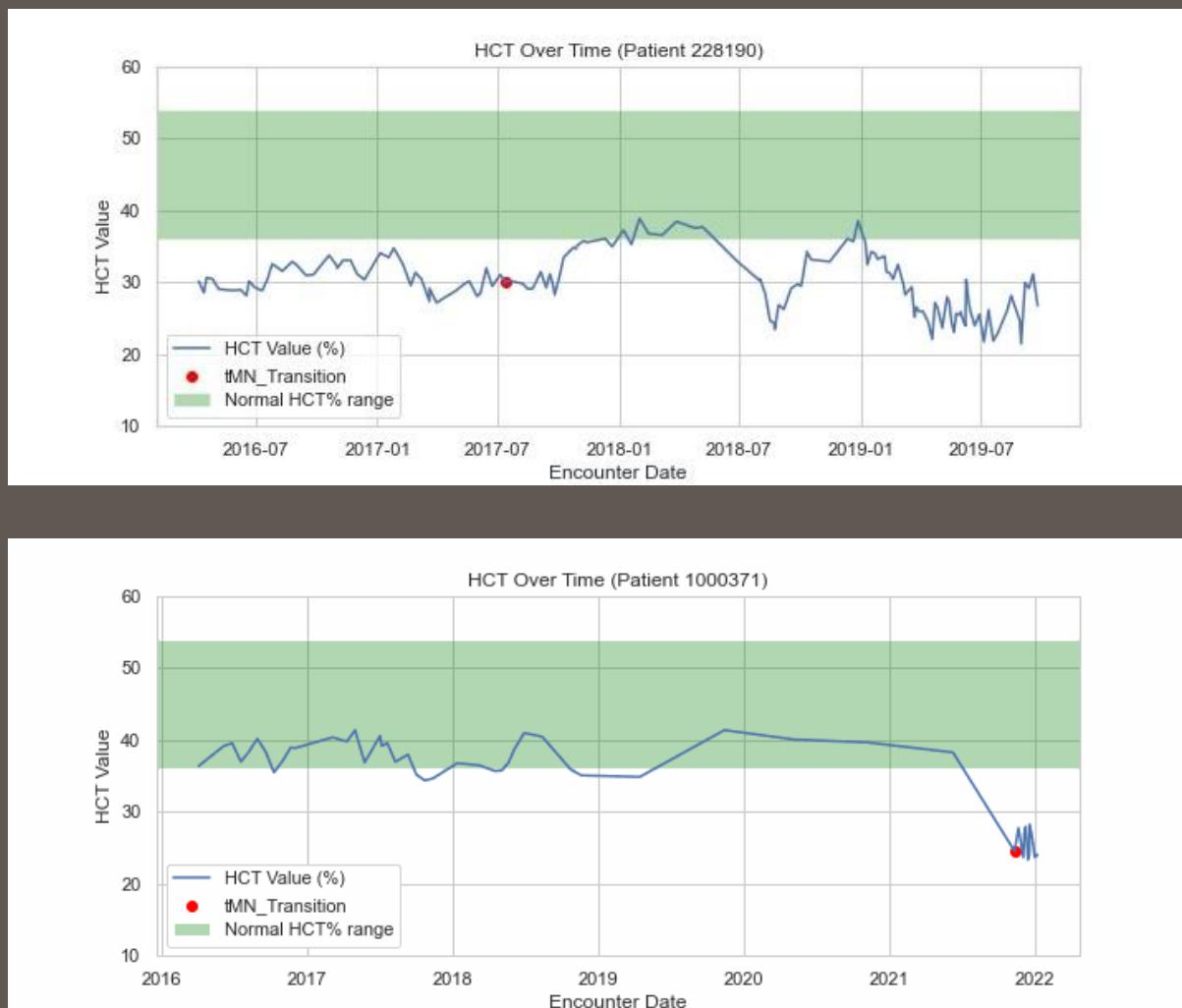
tMN patients:

- HCT level is statistically significant lower dated back to 5 years
- Decreasing trend is consistently dated back to 2.5 years ago

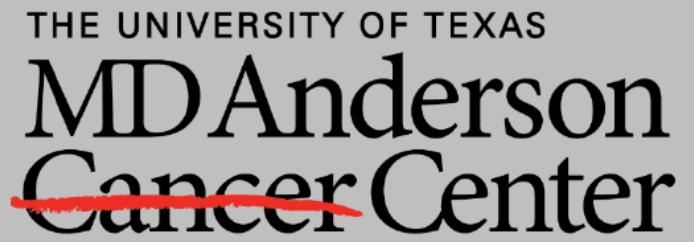
Data Patterns:

- General presence of outliers that lead to extended tails
- Multimodal distribution at -6 time interval, suggesting the potential existence of distinct subgroups or patterns within the data
- The presence of two peaks hints at different characteristics or behaviors within the tMN patient group and needs further analysis

HCT - Individual Graphs



- Overall trend:
 - Statistical lower values 2.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Medication Side Effects: Certain medications used in cancer treatment can have direct effects on red blood cell production and survival, influencing hematocrit levels.
 - Inflammatory Response: Cancer and its treatments can trigger inflammatory responses in the body, affecting red blood cell production and survival, which may influence hematocrit levels.
- Conclusion:
 - Relatively strong performing parameter but needs the combined analysis with other parameters to solve the multi-model pattern problem



Making Cancer History®

11.21

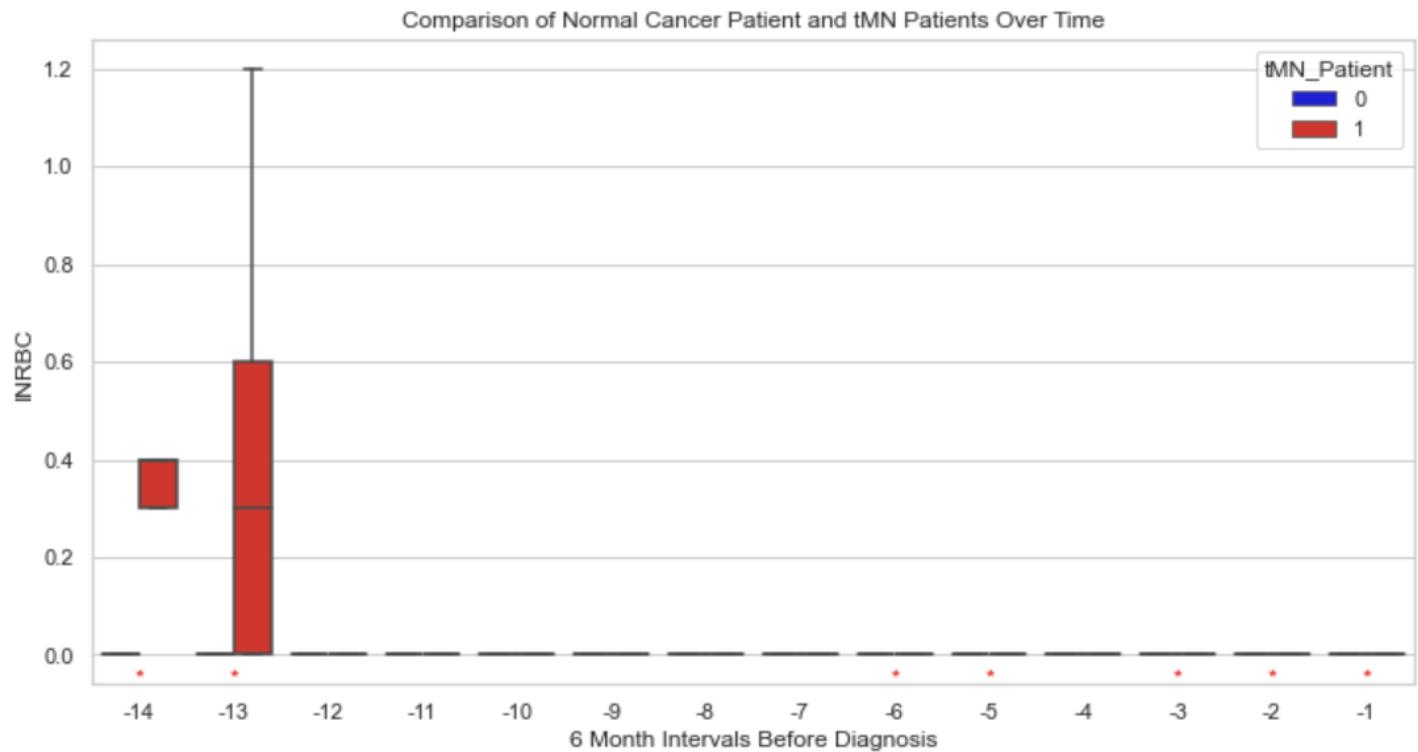
INRBC

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	974206687.5	6.388418e-07	0.00	significant
1	-2	810063223.5	1.207227e-08	0.00	significant
2	-3	713925195.5	1.102017e-03	0.00	significant
3	-4	579716439.5	5.502792e-02	0.06	not significant
4	-5	468672769.0	4.539125e-02	0.05	significant
5	-6	411326131.0	1.055792e-02	0.01	significant
6	-7	269455213.0	5.662785e-01	0.57	not significant
7	-8	173828573.0	7.475327e-01	0.75	not significant
8	-9	177225193.5	7.711553e-01	0.77	not significant
9	-10	109143557.0	8.435793e-02	0.08	not significant
10	-11	68259181.5	1.499791e-01	0.15	not significant
11	-12	10303474.5	2.963389e-01	0.30	not significant
12	-13	19345193.5	2.330700e-08	0.00	significant
13	-14	6723331.5	1.075591e-05	0.00	significant

INRBC– General Trend

tMN patients:

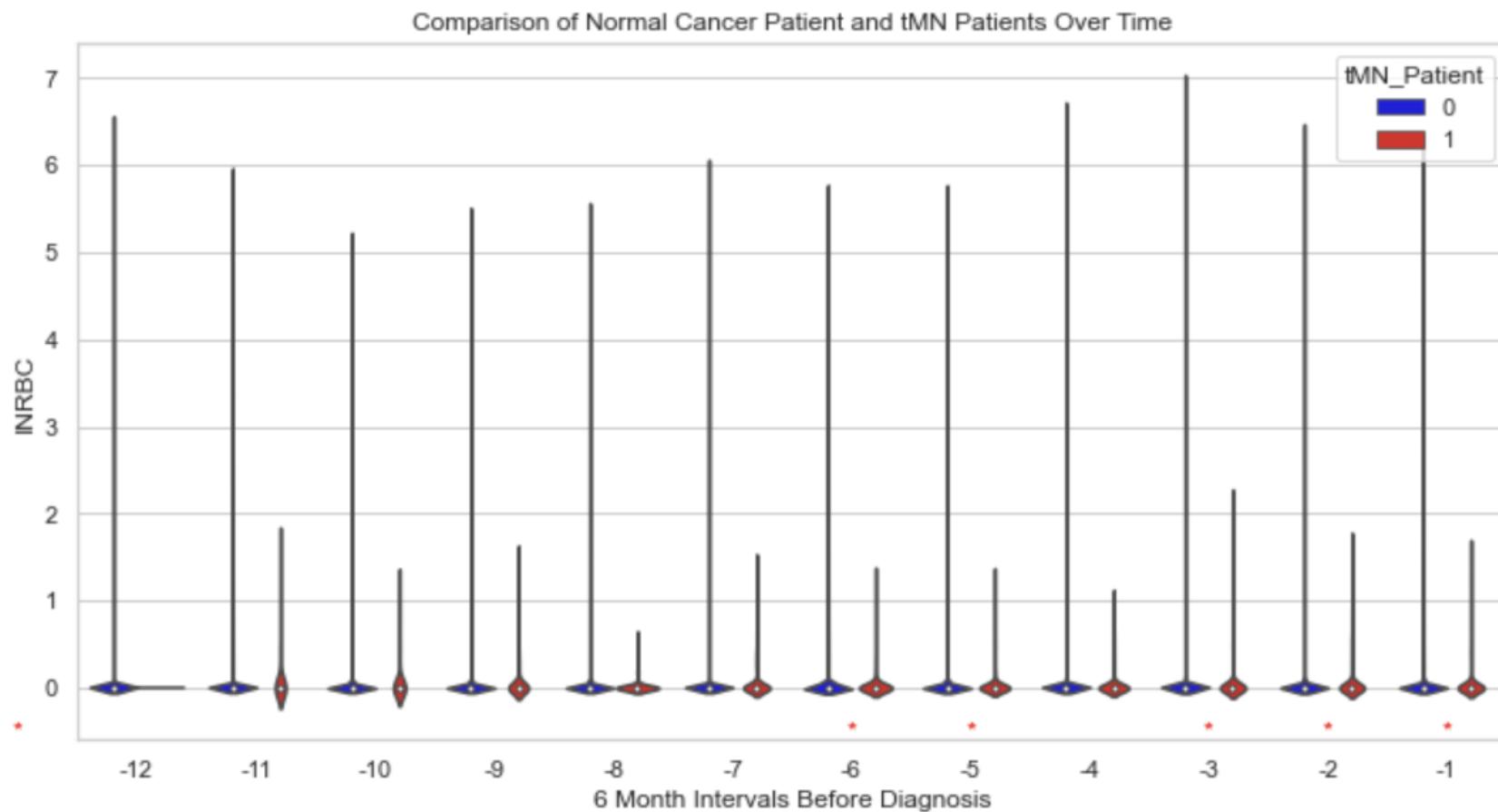
- INRBC level is statistically significant lower and consistently decreasing dated back to 4 years ago
- Better and more statistically stable indicator than NEUT%



INRBC – General Trend

Significance Fluctuations:

- General presence of outliers that lead to extended tails in all time intervals, especially in the -4 time interval
- Time-interval data distribution are generally left skewed with large value outliers



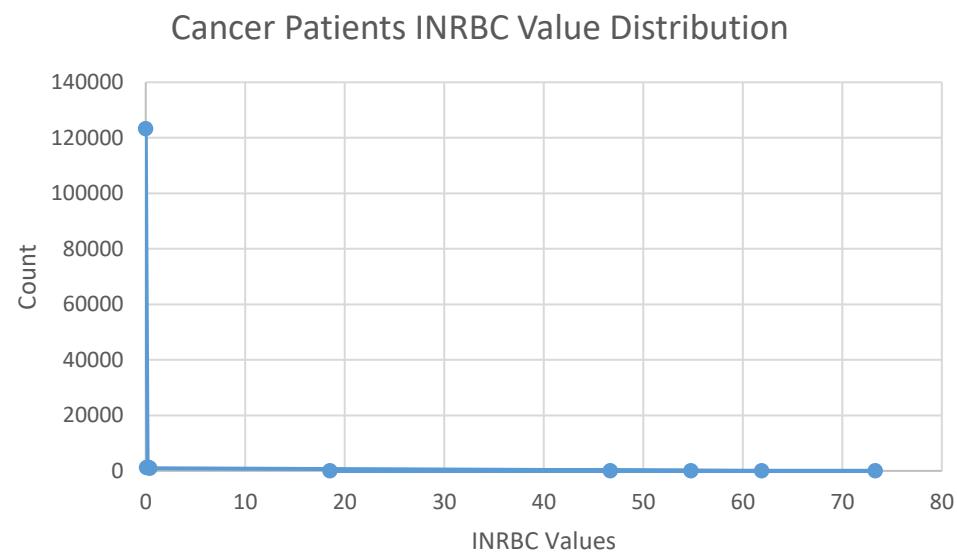
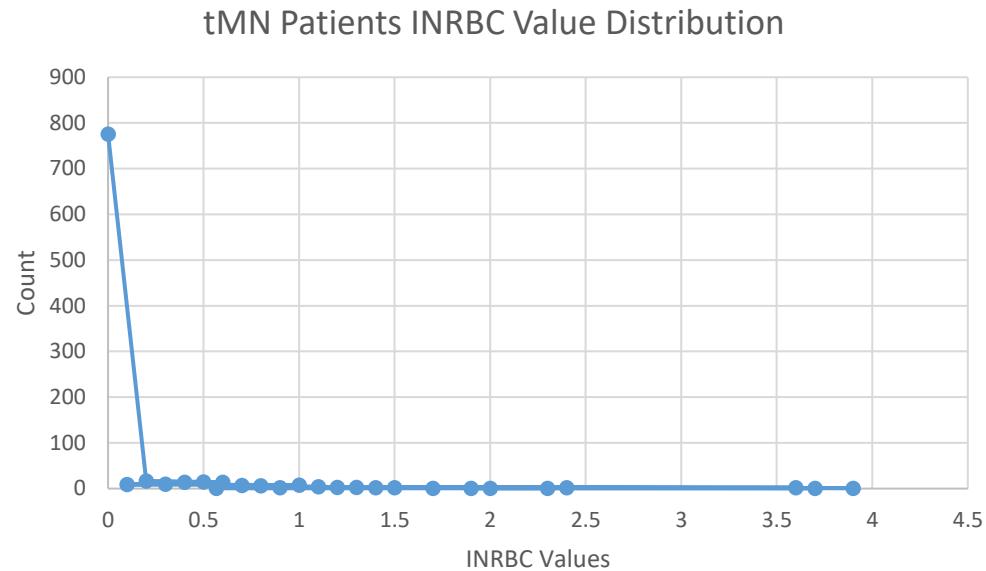
INRBC – Data Distribution

tMN patients:

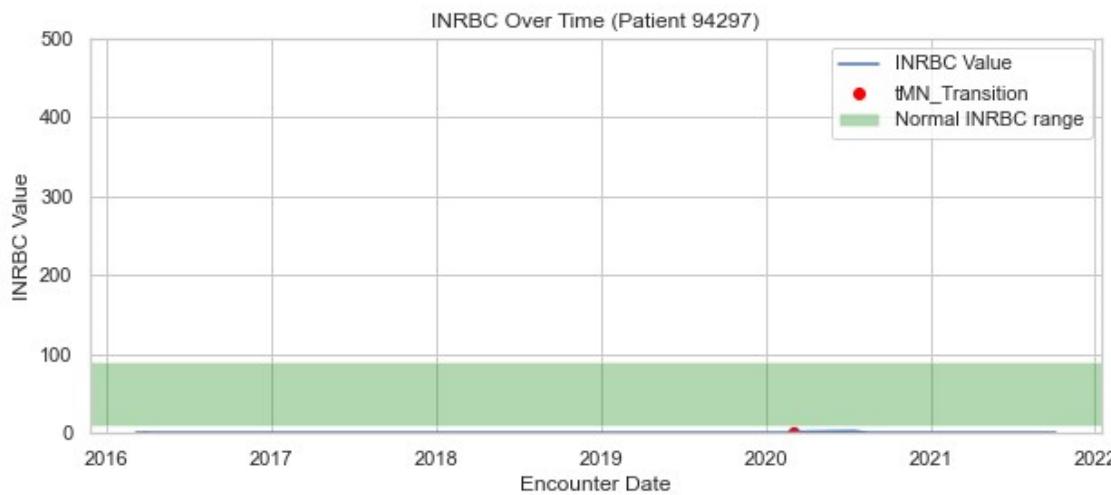
- Value are generally smaller and more focus

Cancer Patients:

- Contain spread out outliers and thus influence the general graph results when plotting these two groups side by side



INRBC - Individual Graphs



- Overall trend:
 - Statistical lower values close to zeros trend dated back 1.5 years before the tMN transition
- Data Pattern:
 - Zero values for INRBC may indicate results within the normal range, potentially rounded down due to the laboratory's practices or limitations in the testing method's detection sensitivity.
- Fluctuation Probable Causes:
 - Treatment Side Effects: Certain chemotherapy and radiation therapy, may directly impact the bone marrow, leading to an increased production of immature blood cells.
 - Bone Marrow Infiltration: Cancer cells infiltrating the bone marrow can disrupt normal blood cell production, resulting in an elevated count of immature red blood cells.
- Conclusion:
 - Need to work with other parameters

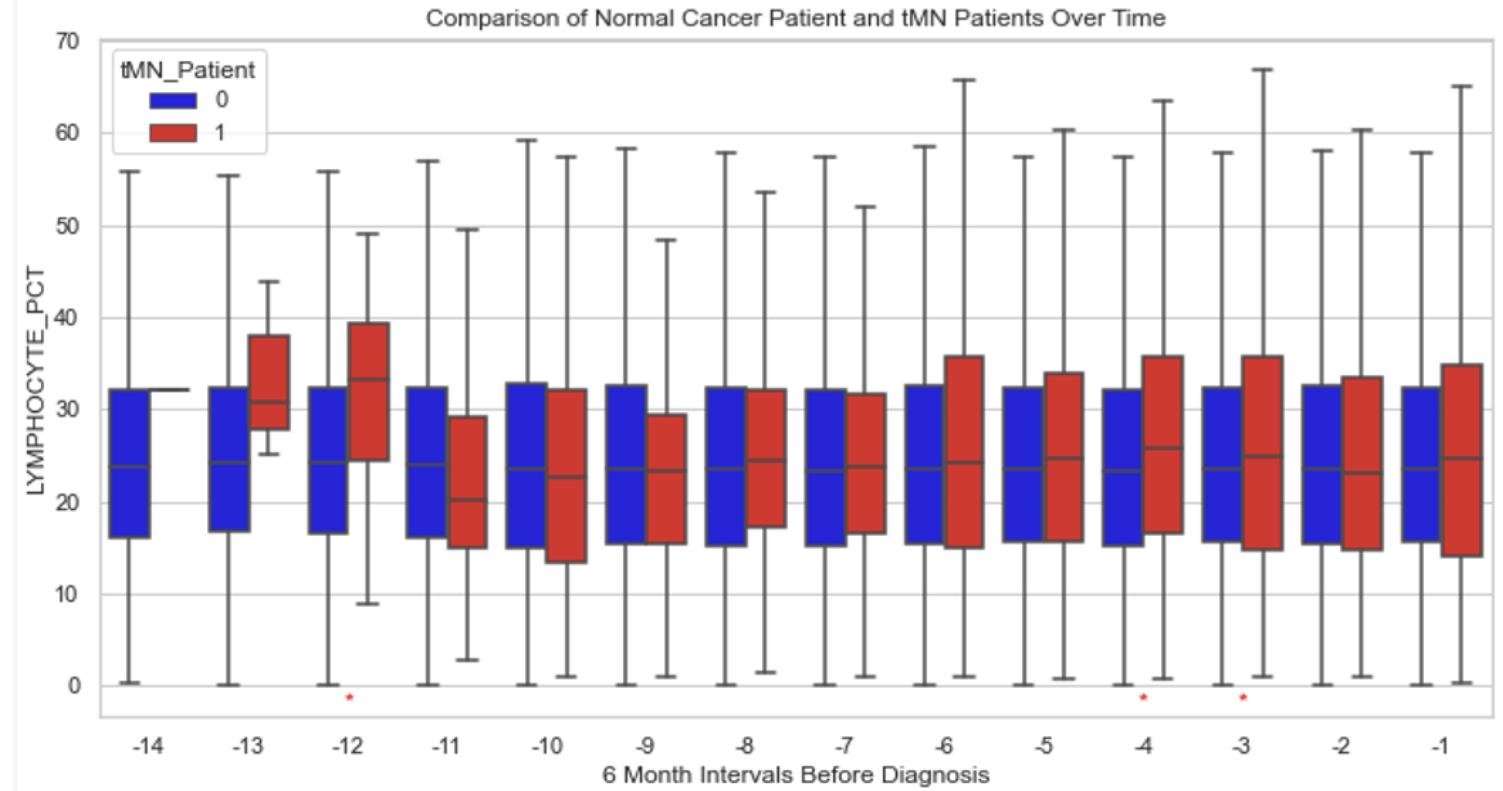
LYMPHOCYTE %

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.250002e+09	0.403095	0.40	not significant
1	-2	1.070585e+09	0.294153	0.29	not significant
2	-3	9.733175e+08	0.037294	0.04	significant
3	-4	8.371502e+08	0.000343	0.00	significant
4	-5	5.969630e+08	0.156916	0.16	not significant
5	-6	6.160483e+08	0.390862	0.39	not significant
6	-7	3.465377e+08	0.885024	0.89	not significant
7	-8	2.368393e+08	0.288713	0.29	not significant
8	-9	1.798232e+08	0.183040	0.18	not significant
9	-10	1.439820e+08	0.320697	0.32	not significant
10	-11	6.859013e+07	0.190853	0.19	not significant
11	-12	1.836644e+07	0.049904	0.05	significant
12	-13	1.179675e+07	0.104912	0.10	not significant
13	-14	1.853772e+06	0.396796	0.40	not significant

LYMPH% – General Trend

tMN patients:

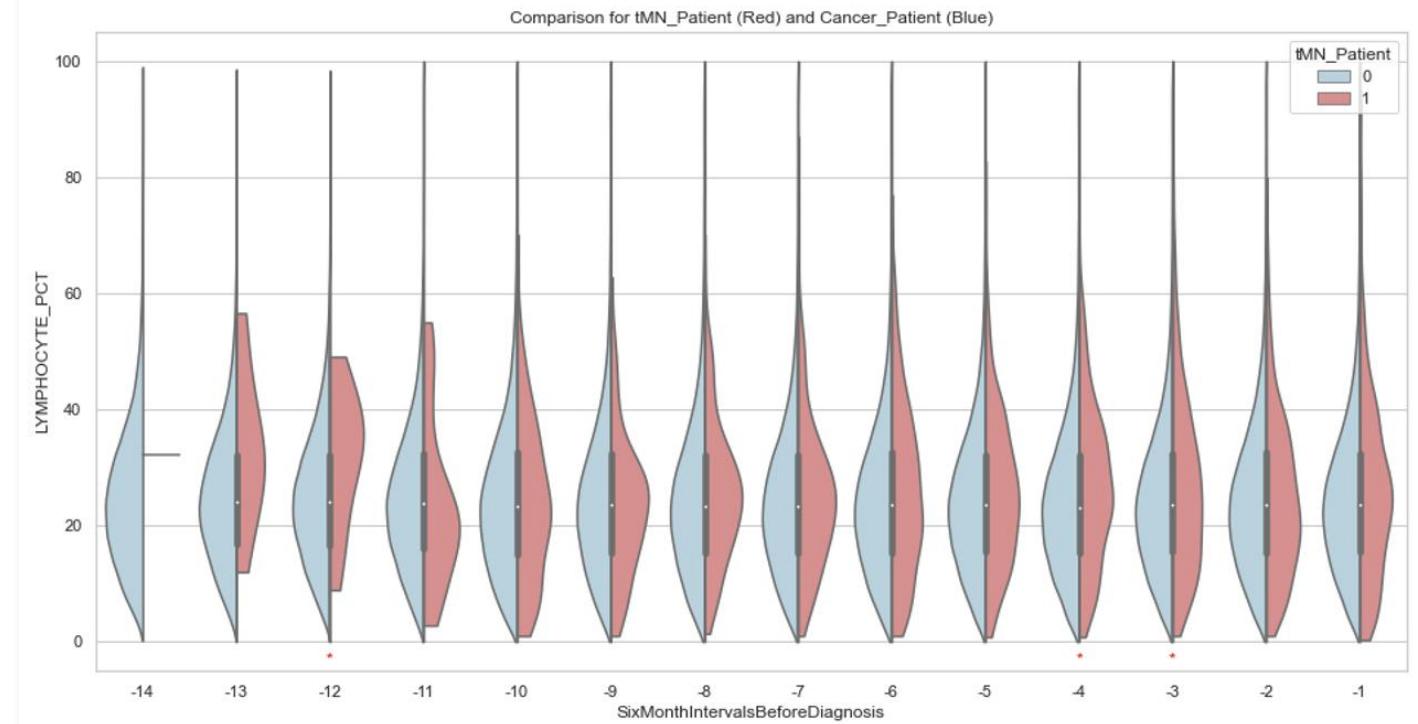
- LYMPH% level is statistically significant larger at time interval -3 and -4, but not a consistently trend



LYMPH% – General Trend

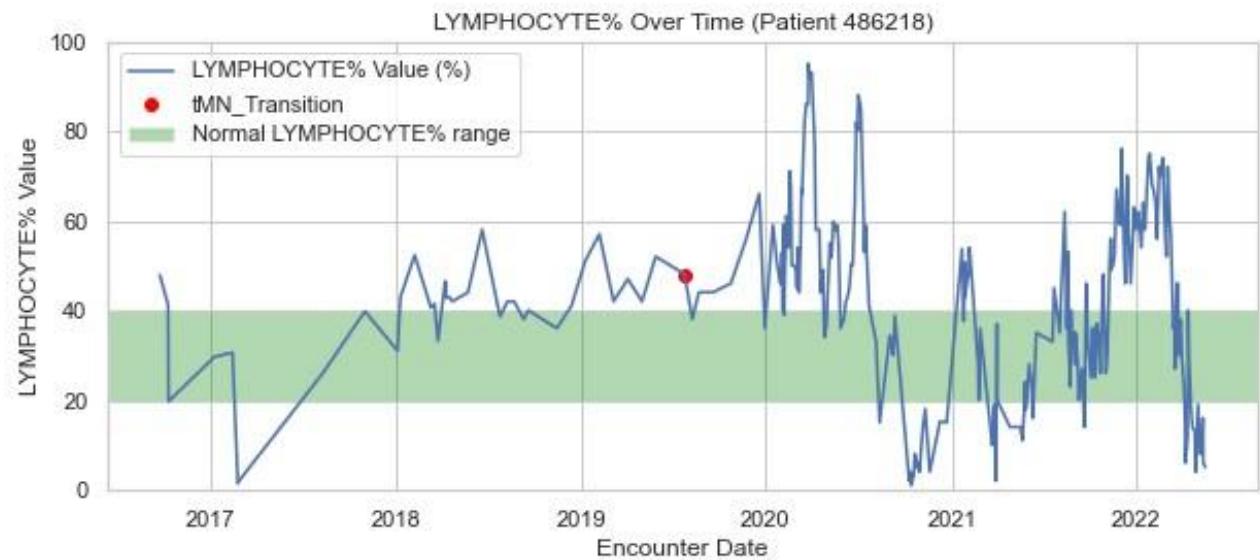
Significance Fluctuations:

- General presence of outliers that lead to extended tails in all time intervals
- Time-interval data distribution are generally left skewed with large value outliers.



LYMPH% - Individual Graphs

- Overall trend:
 - Statistical higher values 1 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Treatment Effects: Cancer treatments, such as chemotherapy or radiation therapy, can have direct effects on the immune system, potentially influencing the LYMPH% in blood test results.
 - Inflammatory Response: Cancer is often associated with inflammation, and an inflammatory response can affect the distribution of white blood cells, including lymphocytes.
- Conclusion:
 - Since both groups perform similarly, hard to separate by the parameter itself.



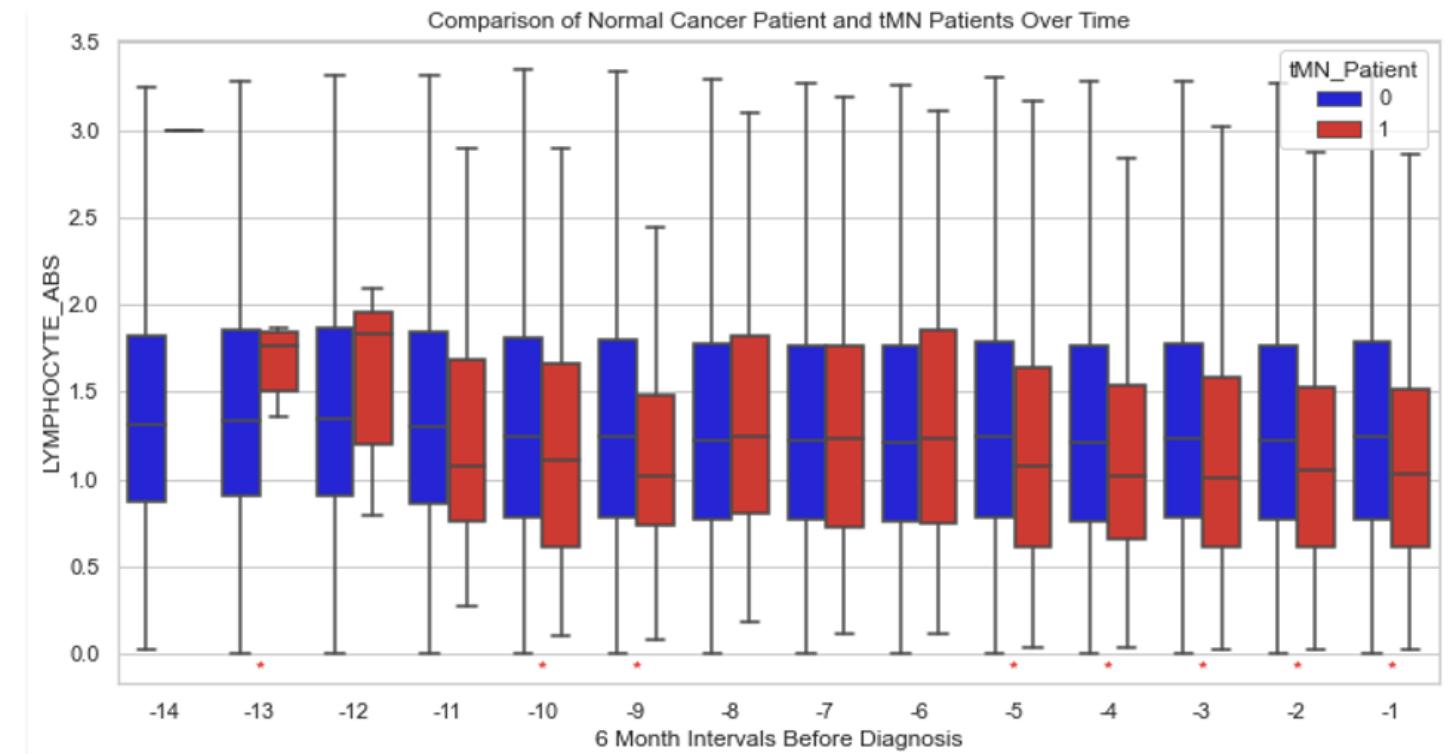
LYMPHOCYTE ABS

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	996396322.5	5.184495e-25	0.00	significant
1	-2	917150952.0	1.587296e-16	0.00	significant
2	-3	790528688.5	5.740056e-13	0.00	significant
3	-4	649523598.5	5.503206e-12	0.00	significant
4	-5	496562671.0	3.754284e-07	0.00	significant
5	-6	597694015.5	7.642552e-01	0.76	not significant
6	-7	331711244.0	2.744704e-01	0.27	not significant
7	-8	229731017.0	7.417220e-01	0.74	not significant
8	-9	161576847.0	7.356008e-04	0.00	significant
9	-10	135672578.5	4.149762e-02	0.04	significant
10	-11	70515000.5	3.349399e-01	0.33	not significant
11	-12	17599151.5	1.011237e-01	0.10	not significant
12	-13	12442726.0	4.979542e-02	0.05	significant
13	-14	2372877.5	1.164354e-01	0.12	not significant

LYMPH ABS – General Trend

tMN patients:

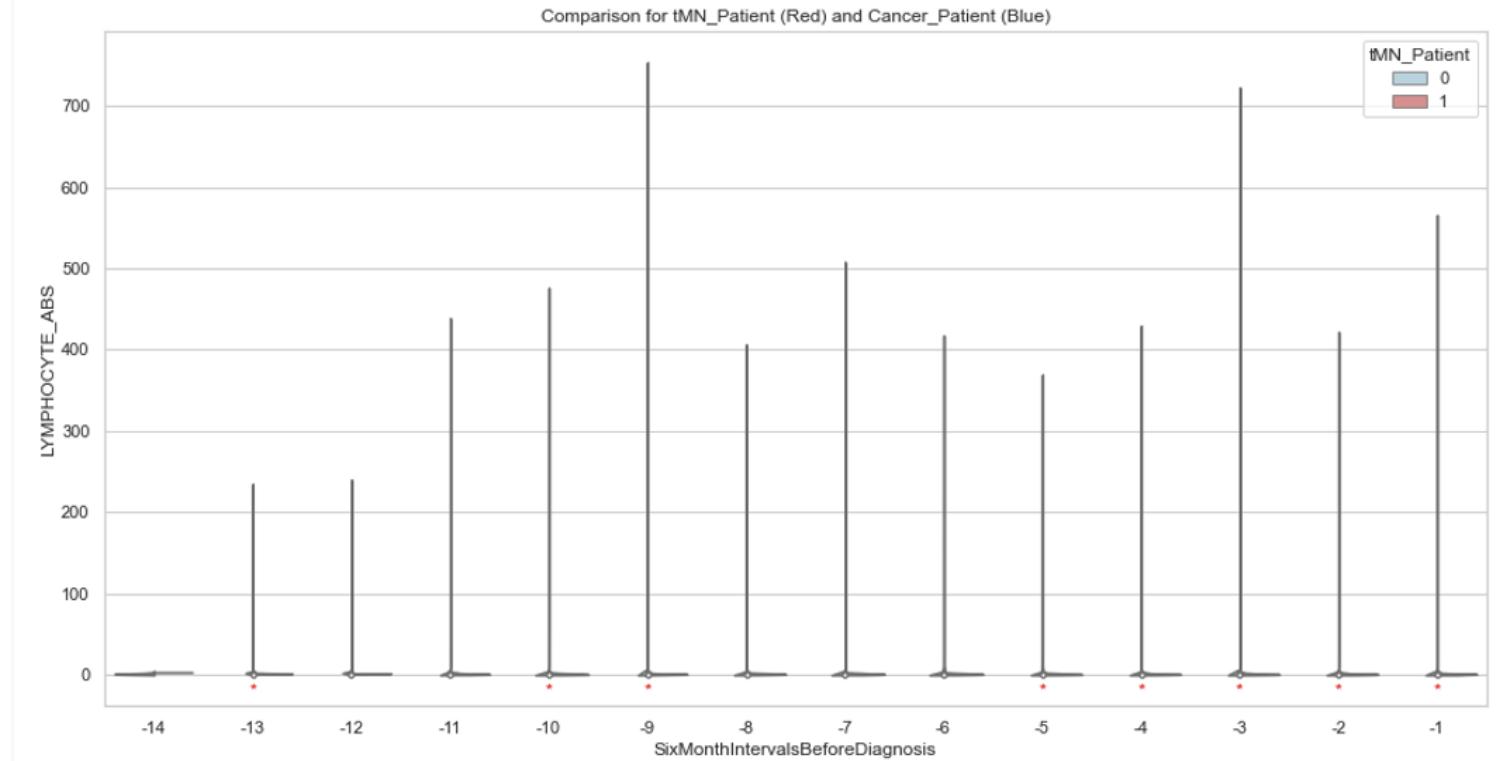
- LYMPH ABS level is statistically significant lower dated back to 2.5 years



LYMPH ABS – General Trend

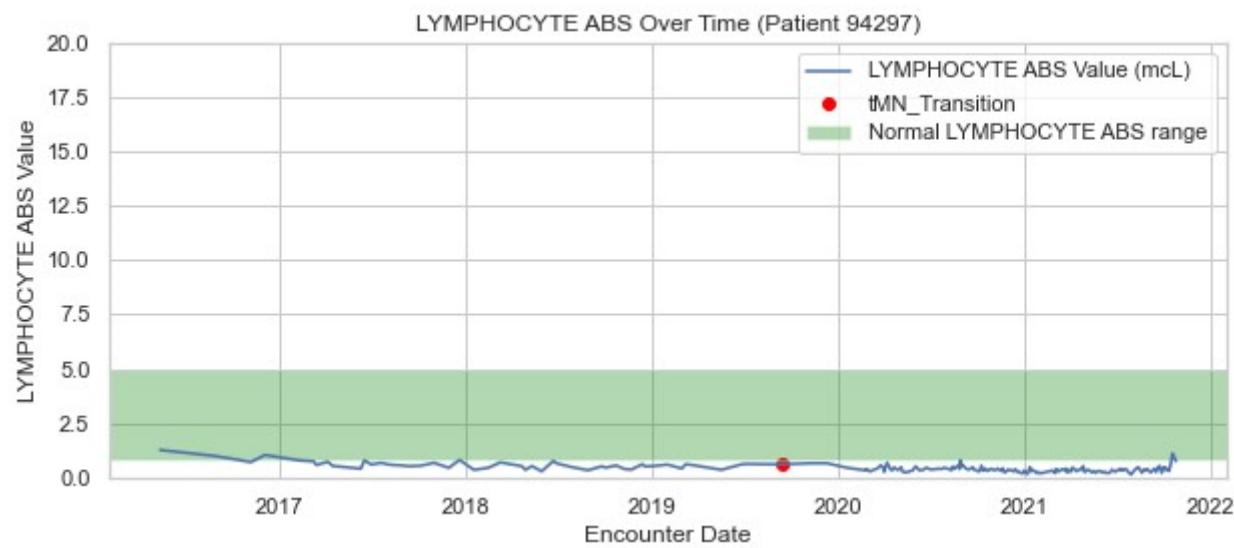
Data Patterns:

- Outliers concentrate in time interval -9 and -3 for non-tMNs group
- Time-interval data distribution are generally left skewed with large value outliers
- Too many large outliers compressed the violin graph



LYMPH ABS - Individual Graphs

- Overall trend:
 - Statistical lower values 2.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Treatment Effects: Cancer treatments, such as chemotherapy or radiation therapy, can have direct effects on the immune system, potentially influencing the LYMPH% in blood test results.
 - Inflammatory Response: Cancer is often associated with inflammation, and an inflammatory response can affect the distribution of white blood cells, including lymphocytes.
- Conclusion:
 - Perform better than LYMPH% but not the statistically strong factors compared to other parameters.



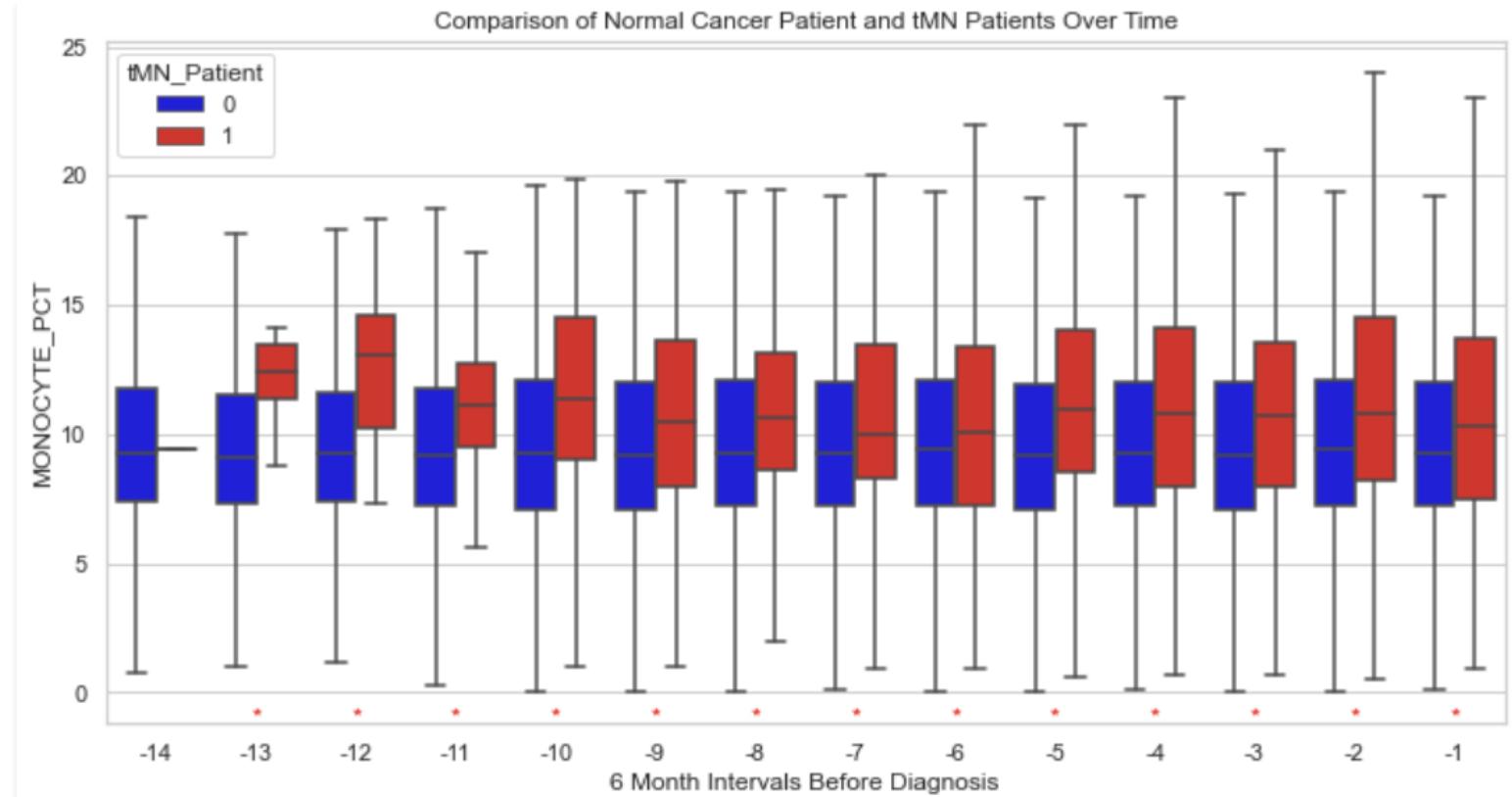
MONOCYTE %

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.357085e+09	2.817918e-11	0.00	significant
1	-2	1.294035e+09	5.021241e-25	0.00	significant
2	-3	1.072713e+09	1.119558e-14	0.00	significant
3	-4	9.051319e+08	5.512878e-16	0.00	significant
4	-5	6.886121e+08	3.207076e-16	0.00	significant
5	-6	6.487931e+08	8.429347e-05	0.00	significant
6	-7	3.851465e+08	4.399683e-05	0.00	significant
7	-8	2.621533e+08	2.966494e-05	0.00	significant
8	-9	2.155933e+08	2.451621e-03	0.00	significant
9	-10	1.899139e+08	2.342690e-07	0.00	significant
10	-11	9.124633e+07	1.867134e-03	0.00	significant
11	-12	1.853129e+07	3.621247e-02	0.04	significant
12	-13	1.289810e+07	2.400250e-02	0.02	significant
13	-14	1.266362e+06	9.644026e-01	0.96	not significant

MONO% – General Trend

tMN patients:

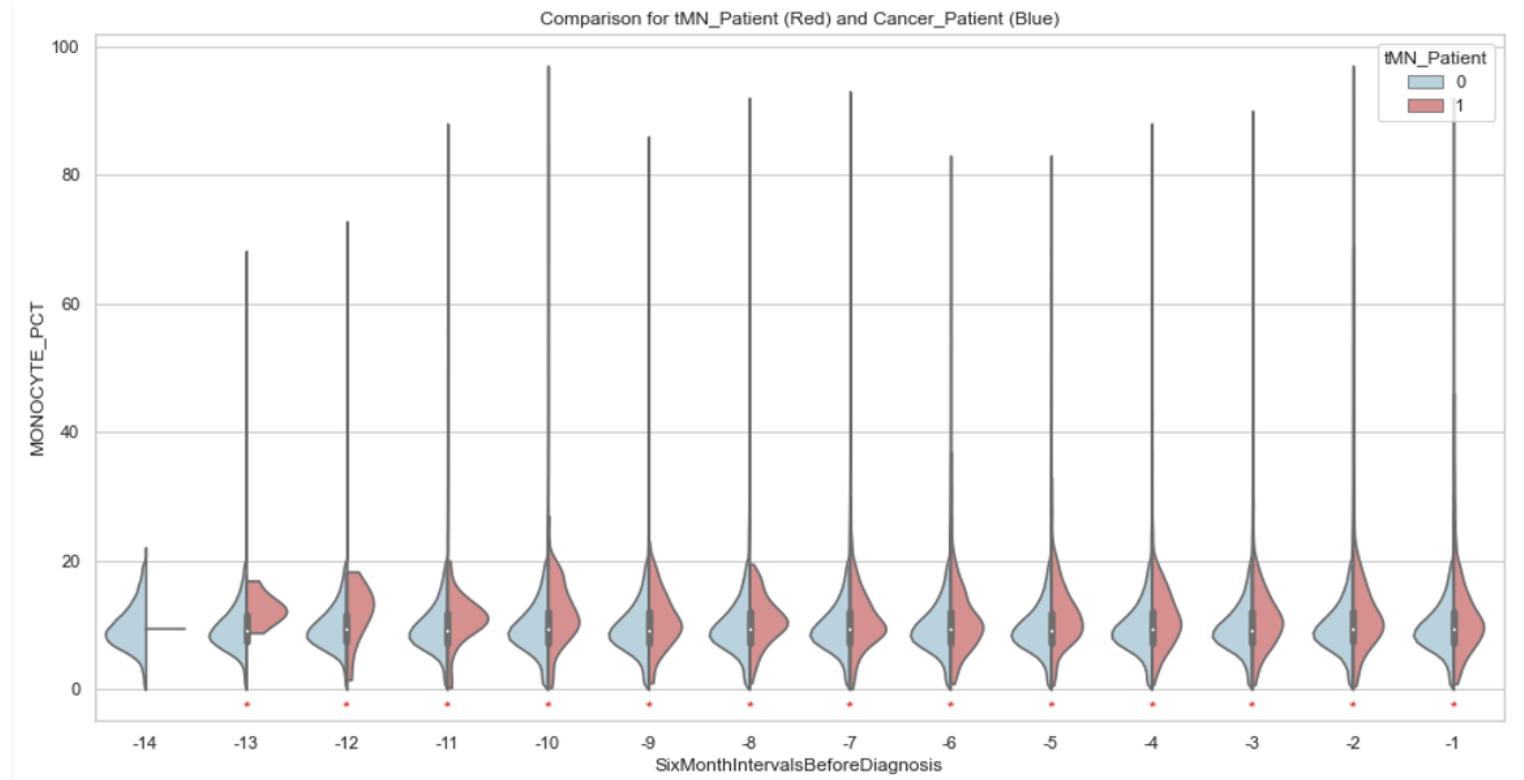
- MONO% level is statistically significant larger dated back to 5-6 years



MONO% – General Trend

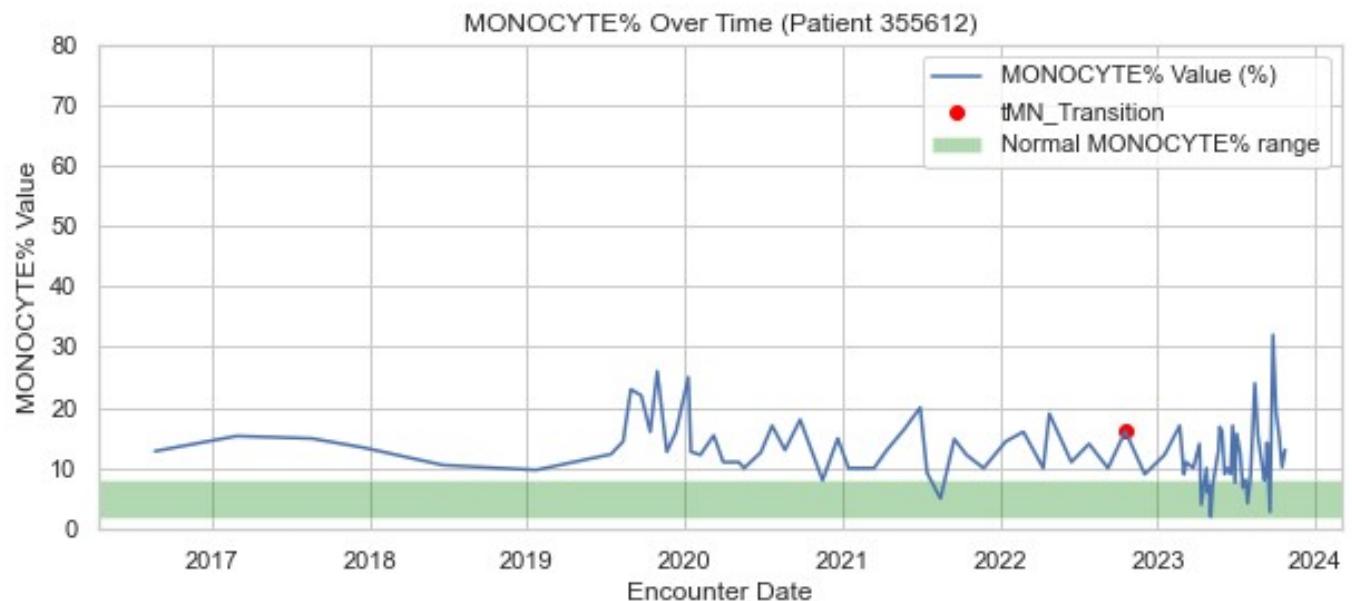
Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers



MONO% - Individual Graphs

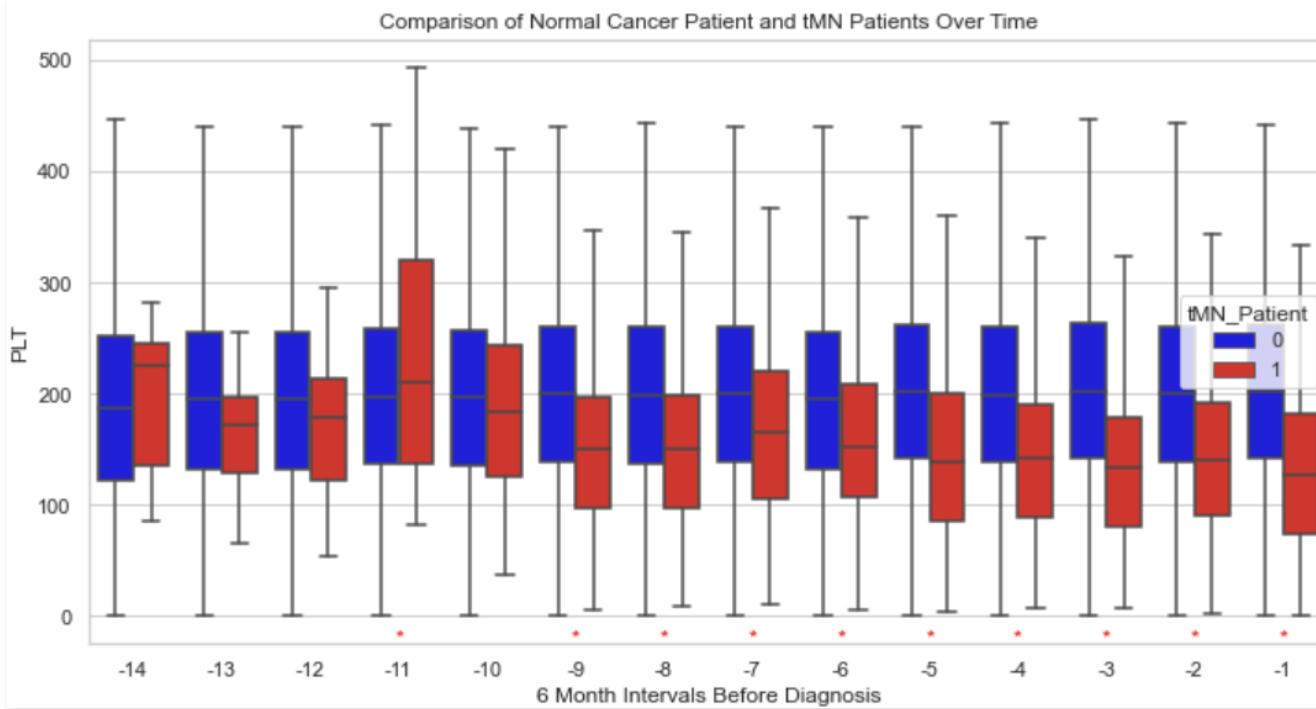
- Overall trend:
 - Statistical higher values 5.5 year before the tMN transition but stays
- Fluctuation Probable Causes:
 - Chemotherapy Medications: Some chemotherapy drugs may suppress monocyte production, while others may lead to an increase in monocyte levels.
 - Inflammation and Infection: cancer patients may be more susceptible to infections, which can also affect the MONO% in the blood.
 - Cancer Type and Stage: Different types and stages of cancer can influence the immune response, leading to fluctuations in monocyte levels.
- Conclusion:
 - Perform better than LYMPH% but not the statistically strong factors compared to other parameters.



PLT

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	943287402.0	7.084383e-132	0.00	significant
1	-2	917702249.5	9.035808e-82	0.00	significant
2	-3	721178868.0	3.082969e-87	0.00	significant
3	-4	652245071.5	2.527043e-61	0.00	significant
4	-5	562746200.0	2.929608e-45	0.00	significant
5	-6	586643015.0	5.141659e-23	0.00	significant
6	-7	350073575.5	8.293898e-11	0.00	significant
7	-8	226656256.5	2.005612e-15	0.00	significant
8	-9	216207568.5	1.797498e-14	0.00	significant
9	-10	209261459.5	1.789842e-01	0.18	not significant
10	-11	142787949.0	2.616990e-02	0.03	significant
11	-12	12015900.0	2.861883e-01	0.29	not significant
12	-13	14063180.5	1.036115e-01	0.10	not significant
13	-14	6735387.0	9.869067e-01	0.99	not significant

PLT – General Trend



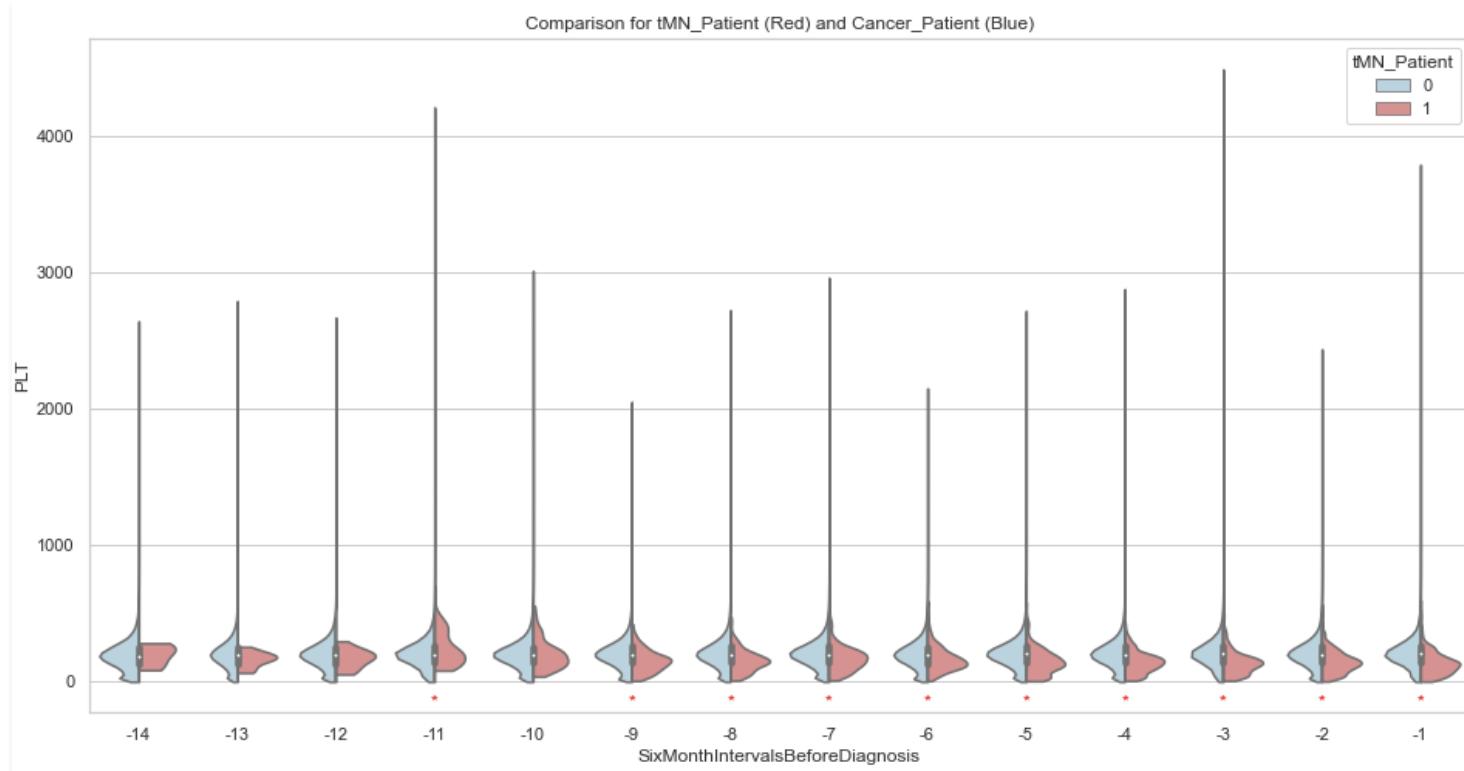
tMN patients:

- PLT level is statistically significant lower and decreasing dated back to 4.5 years

Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers, especially in the -3 time interval
- Multi-model patterns in -4 and -5 time interval for tMN Patients
- Multi-model pattern generally exist in the cancer patients for PLT values.

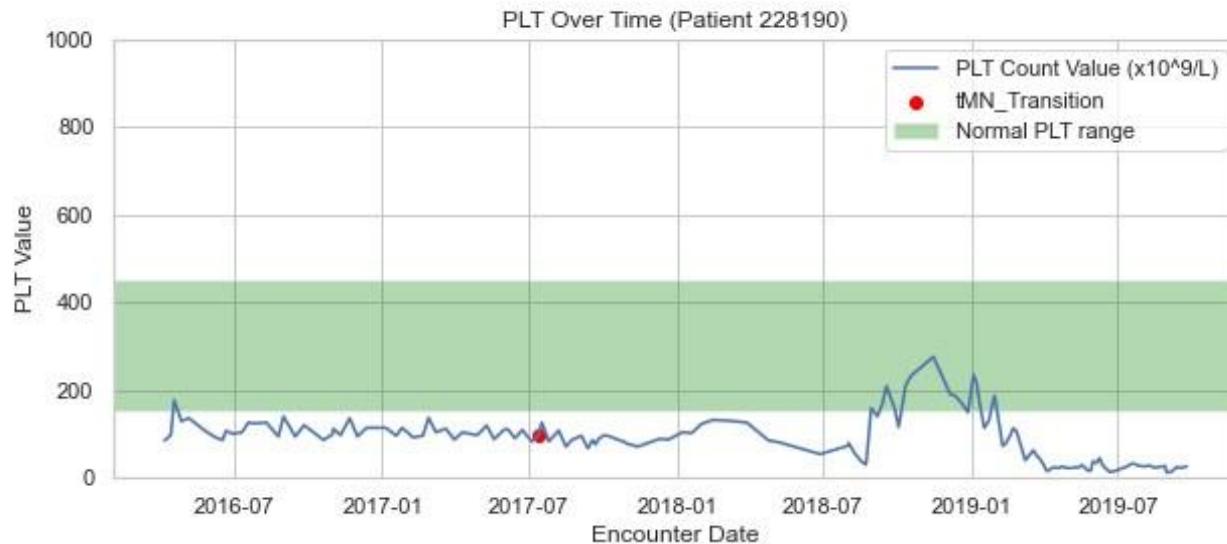
PLT – General Trend



Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers, especially in the -3 time interval
- Multi-model patterns in -4 and -5 time interval for tMN Patients
- Multi-model pattern generally exist in the cancer patients for PLT values.

PLT - Individual Graphs

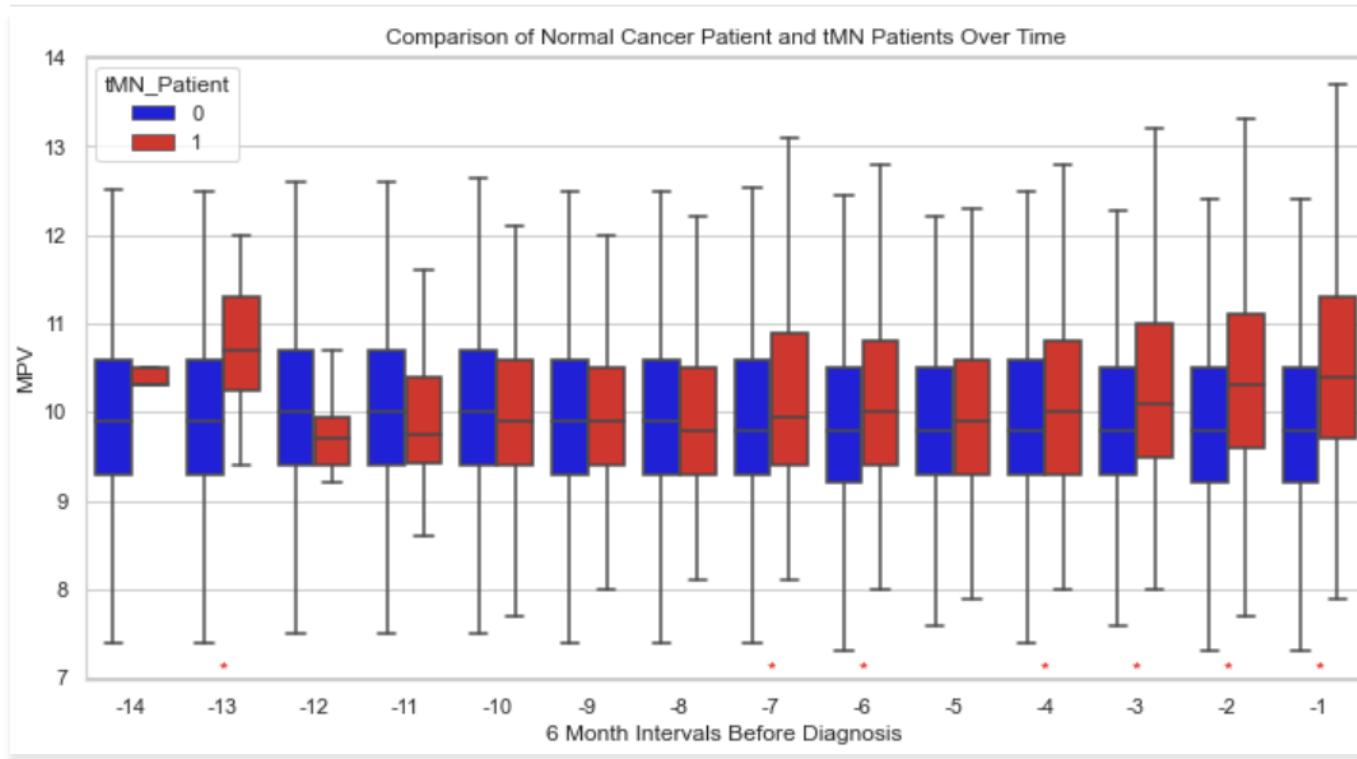


- Overall trend:
 - Statistical lower values 4.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Bone Marrow Infiltration: In some cases, cancer cells may infiltrate the bone marrow, disrupting normal blood cell production, including platelets.
 - Inflammatory Response: Cancer and its treatments can trigger inflammatory responses in the body, affecting the balance of blood components, including platelets.
- Conclusion:
 - Statistically strong predictor, but need other parameters to help with the multi-model problem

MPV

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	943287402.0	7.084383e-132	0.00	significant
1	-2	917702249.5	9.035808e-82	0.00	significant
2	-3	721178868.0	3.082969e-87	0.00	significant
3	-4	652245071.5	2.527043e-61	0.00	significant
4	-5	562746200.0	2.929608e-45	0.00	significant
5	-6	586643015.0	5.141659e-23	0.00	significant
6	-7	350073575.5	8.293898e-11	0.00	significant
7	-8	226656256.5	2.005612e-15	0.00	significant
8	-9	216207568.5	1.797498e-14	0.00	significant
9	-10	209261459.5	1.789842e-01	0.18	not significant
10	-11	142787949.0	2.616990e-02	0.03	significant
11	-12	12015900.0	2.861883e-01	0.29	not significant
12	-13	14063180.5	1.036115e-01	0.10	not significant
13	-14	6735387.0	9.869067e-01	0.99	not significant

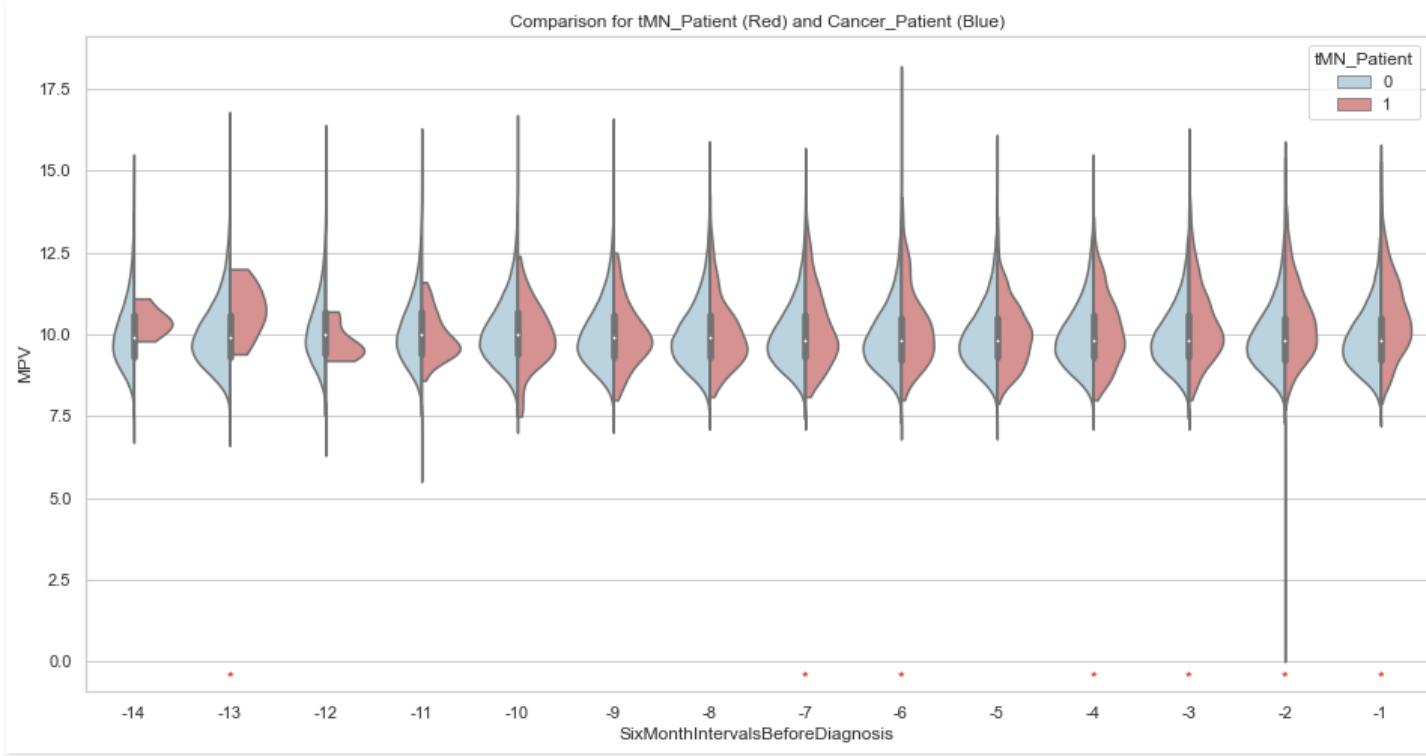
MPV – General Trend



tMN patients:

- MPV level is statistically significant higher dated back to 2 years ago

MPV – General Trend

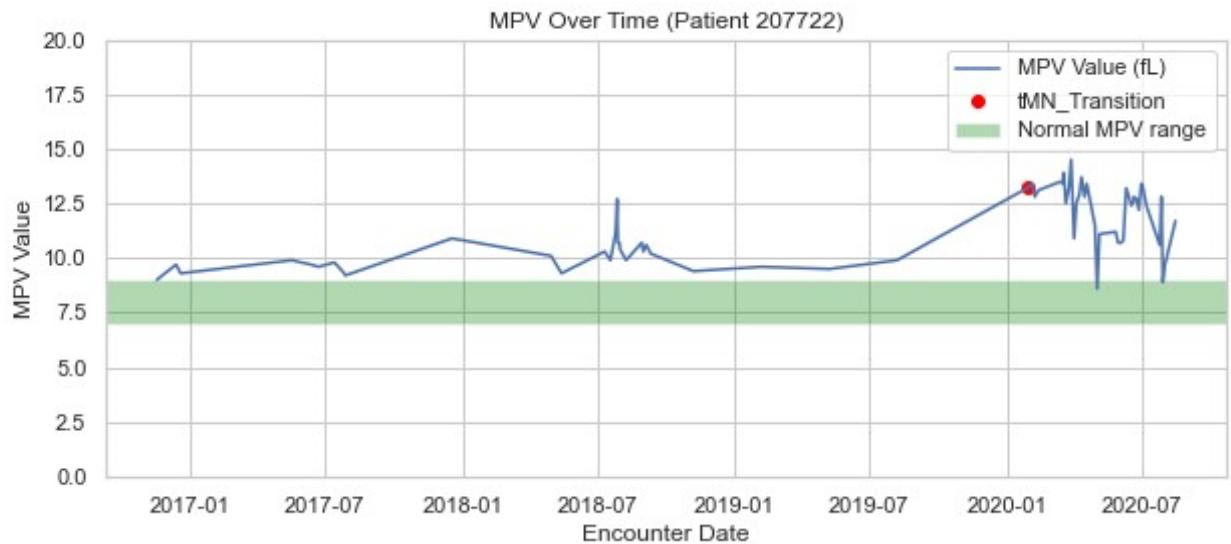


Data Patterns:

- Time-interval data distribution are generally left skewed with large value outliers
- But -2 time interval are right skewed with extreme small values

MPV - Individual Graphs

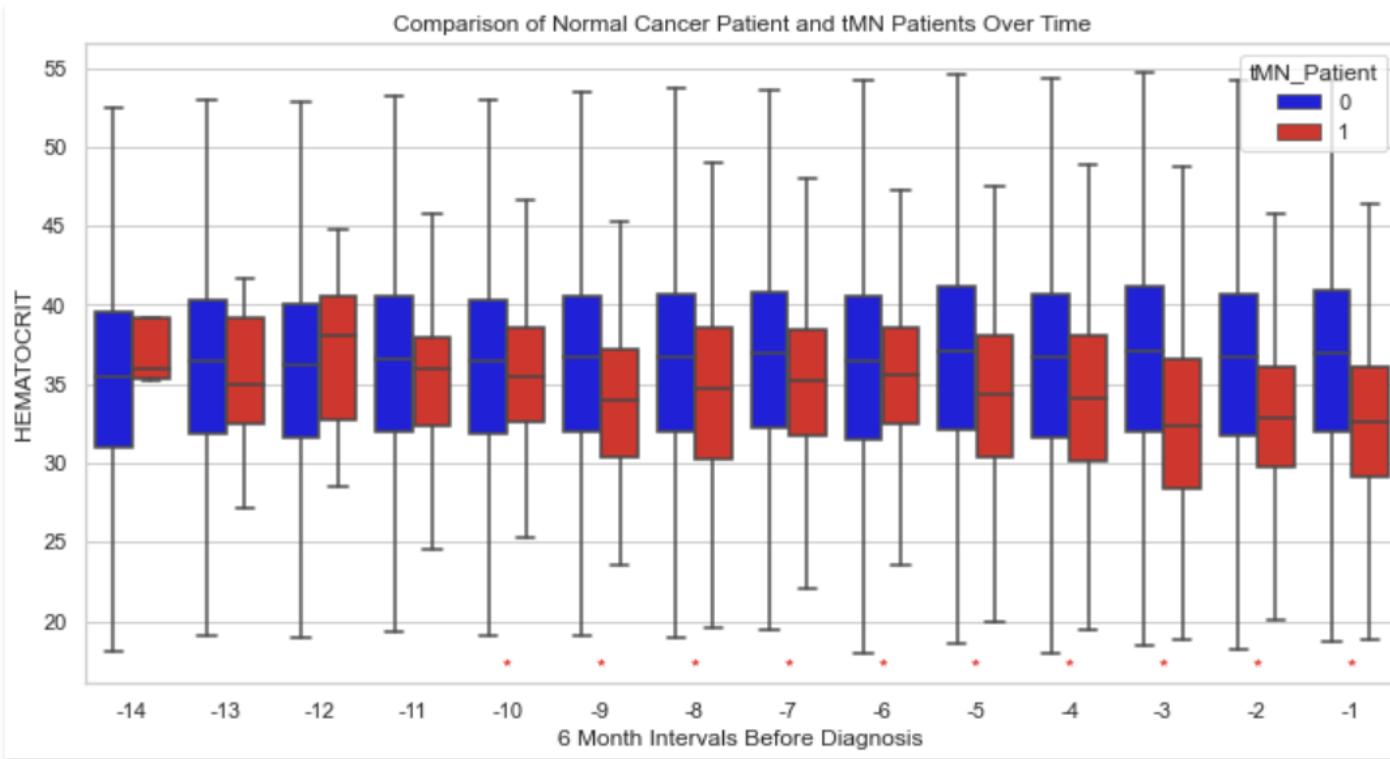
- Overall trend:
 - Statistically significant higher values 2 year before the tMN transition
 - Stable and for staying the above of control group most of the time
- Conclusion:
 - Statistically strong predictor



HEMATOCRIT (HCT)

	Interval	Statistic	P-value	Rounded P-value	Significance
0	-1	1.061194e+09	9.057116e-87	0.00	significant
1	-2	9.702961e+08	1.166947e-63	0.00	significant
2	-3	7.660627e+08	3.074351e-69	0.00	significant
3	-4	7.793637e+08	4.661585e-25	0.00	significant
4	-5	6.574009e+08	1.600426e-19	0.00	significant
5	-6	6.881454e+08	4.736281e-05	0.00	significant
6	-7	3.795775e+08	2.061915e-05	0.00	significant
7	-8	2.643728e+08	3.218270e-06	0.00	significant
8	-9	2.327515e+08	1.838927e-09	0.00	significant
9	-10	1.936406e+08	8.742779e-03	0.01	significant
10	-11	1.140230e+08	1.140396e-01	0.11	not significant
11	-12	1.583058e+07	6.675243e-01	0.67	not significant
12	-13	1.593123e+07	5.964301e-01	0.60	not significant
13	-14	6.163138e+06	6.016844e-01	0.60	not significant

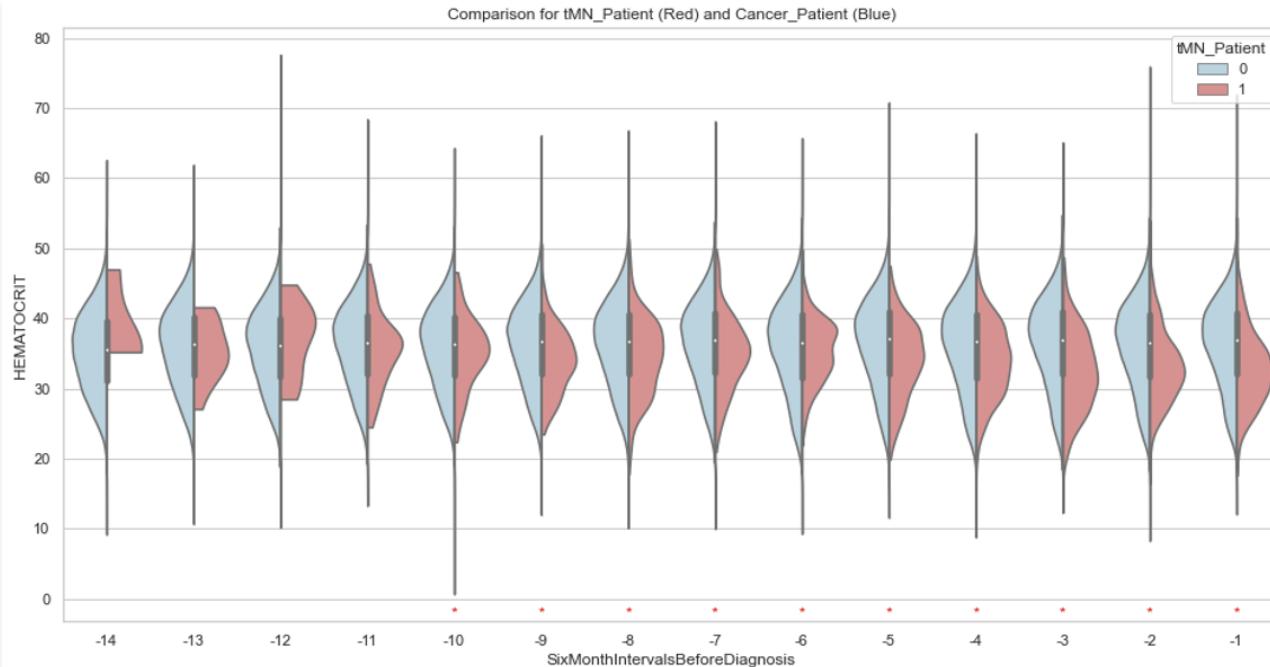
HCT – General Trend



tMN patients:

- HCT level is statistically significant lower dated back to 5 years
- Decreasing trend is consistently dated back to 2.5 years ago

HCT – General Trend

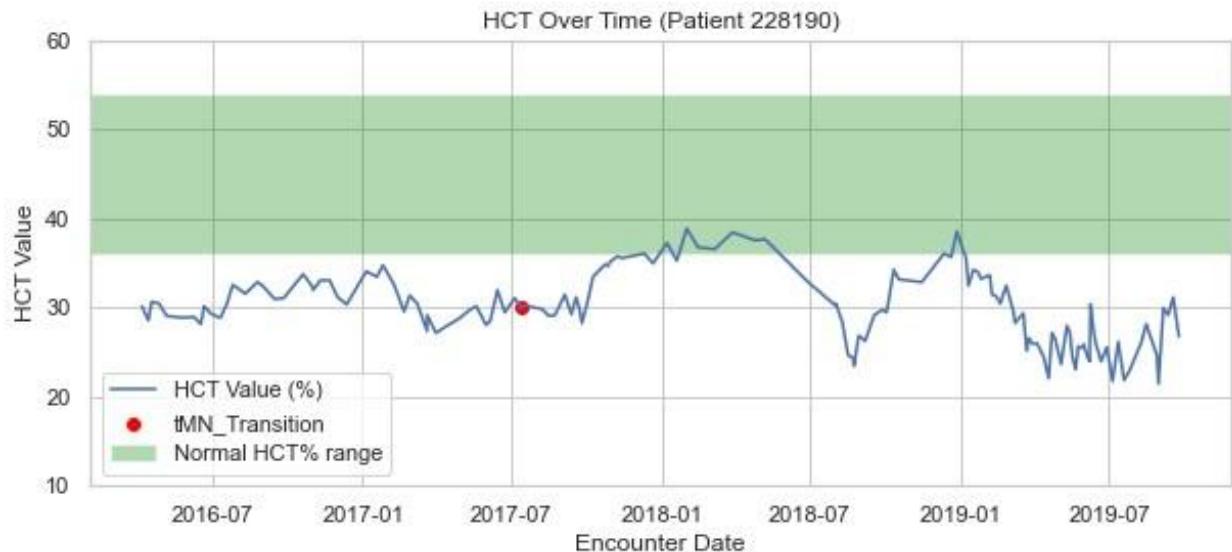


Data Patterns:

- General presence of outliers that lead to extended tails
- Multimodal distribution at -6 time interval, suggesting the potential existence of distinct subgroups or patterns within the data
- The presence of two peaks hints at different characteristics or behaviors within the tMN patient group and needs further analysis

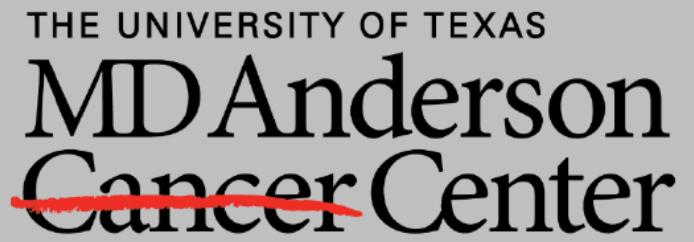
HCT - Individual Graphs

- Overall trend:
 - Statistical lower values 2.5 year before the tMN transition but stays no statistically significant differences within 1 year.
- Fluctuation Probable Causes:
 - Medication Side Effects: Certain medications used in cancer treatment can have direct effects on red blood cell production and survival, influencing hematocrit levels.
 - Inflammatory Response: Cancer and its treatments can trigger inflammatory responses in the body, affecting red blood cell production and survival, which may influence hematocrit levels.
- Conclusion:
 - Relatively strong performing parameter but needs the combined analysis with other parameters to solve the multi-model pattern problem



Conclusion

- This analysis highlights significant divergences in 19 hematological features between cancer patients and those who eventually develop therapy-related myeloid neoplasms (tMN). These differences underscore the potential of routine blood biomarkers as early predictive indicators for tMN risk stratification.
- Notably, mean corpuscular volume (MCV) and red cell distribution width (RDW) emerged as strong individual predictors, demonstrating consistent separation across patient classes in pre-diagnostic time windows. In contrast, absolute neutrophil count (NEUT ABS) outperformed relative neutrophil percentage, suggesting that raw count metrics may preserve more predictive signal than normalized or computed values—possibly due to noise introduced during downstream calculation pipelines.
- Moderate or correlated indicators (e.g., HGB, HCT, WBC, PLT) were integrated into machine learning models for further investigation. To address multicollinearity, we evaluated model weights under regularization schemes such as L1 (Lasso) and L2 (Ridge), ensuring robustness and interpretability.
- Initial machine learning models—including logistic regression, random forest, and XGBoost classifiers—were trained using temporal windows leading up to the transition point. Feature importance analysis revealed both expected and novel biomarkers contributing to early detection. Cross-validation and AUC metrics confirmed the discriminatory power of the selected features.
- These findings establish a foundation for deploying ML-based predictive tools to flag at-risk patients prior to clinical manifestation of tMN—where early intervention could be most effective in delaying or preventing onset.



Making Cancer History®