Health Care Data Analytic and Data Mining  
Final Exam

short line

Group One

Biniyam Abebe

Peter Mukiibi

Doreen Accorley

Yash Puri  
11h March, 2025

[**Introduction 1**](#_exakenlpyb6w)

[**TOOLS USED 3**](#_924vrs83xl3t)

[**The Market Share 3**](#_qvp033cahz0g)

[**Difference Between Government and Commercial Care 12**](#_j66wchk1la1f)

[**Conclusion 15**](#_oleijnnaq78k)

[**CODES 16**](#_7rofhjeb8kv4)

# Introduction

The healthcare landscape is rapidly evolving, driven by advances in technology and a push towards more complex medical procedures. The CBNC report highlights that larger hospitals, with their ability to offer sophisticated and technology-intensive care, seem to have a competitive advantage. This study aims to investigate whether these substantial hospitals also attract a higher proportion of commercially insured patients—ones who bring in more revenue—while seeing fewer government-funded Medicare beneficiaries.

Using the Major Diagnostic Categories (MDCs) to classify procedures, we focus on identifying those that are technologically intensive and complex. By analyzing these high-tech procedures, we can observe the market share of larger, resource-rich hospitals in comparison to all other hospitals in Vermont. With access to the state's All-Payer Claims Database (APCD), we are equipped to conduct a comprehensive analysis.

Furthermore, we aim to explore the geographical referral patterns within Vermont, as patient choice and mobility often influence hospital selection. Understanding these dynamics is crucial, especially when commercial insurance agreements tend to cover the hospital's negotiated price, despite potential cost disparities within the same market area. This analysis seeks to uncover insights into patient behavior and referral trends, offering a clearer understanding of the hospital landscape in Vermont.

* Research Objective: Investigate if larger hospitals with advanced technology attract more commercially insured patients and fewer Medicare beneficiaries.
* Methodology: Analyze technologically intensive and complex procedures classified by MDCs using Vermont’s APCD to compare market share of larger hospitals.
* Geographical Analysis: Explore patient referral patterns within Vermont to understand the influence of patient choice and mobility on hospital selection.

# TOOLS USED

* SAS enterprise, Sas base
* Excel
* Tableau

# 

# 

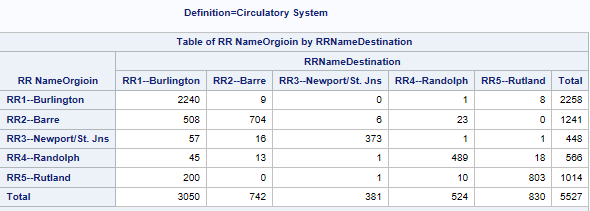
# 

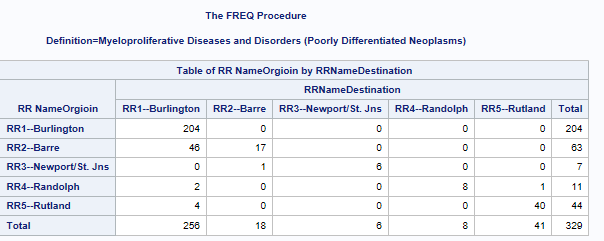
# 

# 

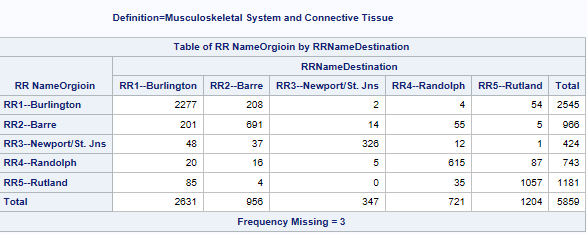
# The Market Share

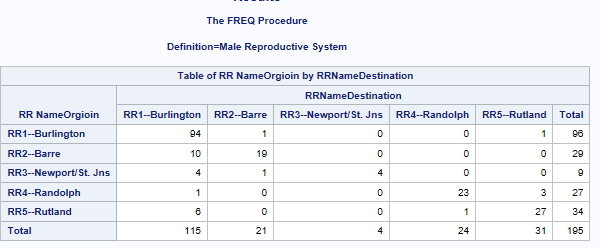
The OD matrix for High End, Circulatory System and Myeloproliferative Disease,





From Low End Care, Musculoskeletal and Male reproductive System





First less evaluate a market between Circulatory system from high end and musculoskeletal from low end. This is the percentage and the market share.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Definition=Circulatory System | | |  |  |  |
| RRName | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 99.2 | 0.4 | 0 | 0.04 | 0.35 |
| RR2--Barre | 40.93 | 56.73 | 0.48 | 1.85 | 0 |
| RR3--Newport/St. Jns | 12.72 | 3.57 | 83.26 | 0.22 | 0.22 |
| RR4--Randolph | 7.95 | 2.3 | 0.18 | 86.4 | 3.18 |
| RR5--Rutland | 19.72 | 0 | 0.1 | 0.99 | 79.19 |
| Market Share | 55.18 | 13.43 | 6.89 | 9.48 | 15.02 |

Definition=Musculoskeletal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 89.47 | 8.17 | 0.08 | 0.16 | 2.12 |
| RR2--Barre | 20.81 | 71.53 | 1.45 | 5.69 | 0.52 |
| RR3--Newport/St. Jns | 11.32 | 8.73 | 76.89 | 2.83 | 0.24 |
| RR4--Randolph | 2.69 | 2.15 | 0.67 | 82.77 | 11.71 |
| RR5--Rutland | 7.2 | 0.34 | 0 | 2.96 | 89.5 |
| Market Share | 44.91 | 16.32 | 5.92 | 12.31 | 20.55 |

As shown above, in 99 percent of circulatory cardiac patients from the high-end from Burlington, they go to region one Burlington. In region 2 Barre, more than 40 percent of patients travel to region 1 Burlington. On the other hand, from the lower end, musculoskeletal, 89 percent of patients from Burlington go to Burlington, but only 20 percent of patients from Barre go to Burlington.

Now less compare Myeloproliferative from the High end care and Male reproductive service from the low end.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Definition=Myeloproliferative Diseases and Disorders (Poorly Differentiated Neoplasms) | | | | | |
|  |  |  |  |  |  |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 100 | 0 | 0 | 0 | 0 |
| RR2--Barre | 73.02 | 26.98 | 0 | 0 | 0 |
| RR3--Newport/St. Jns | 0 | 14.29 | 85.71 | 0 | 0 |
| RR4--Randolph | 18.18 | 0 | 0 | 72.73 | 9.09 |
| RR5--Rutland | 9.09 | 0 | 0 | 0 | 90.91 |
| Market Share | 77.81 | 5.47 | 1.82 | 2.43 | 12.46 |

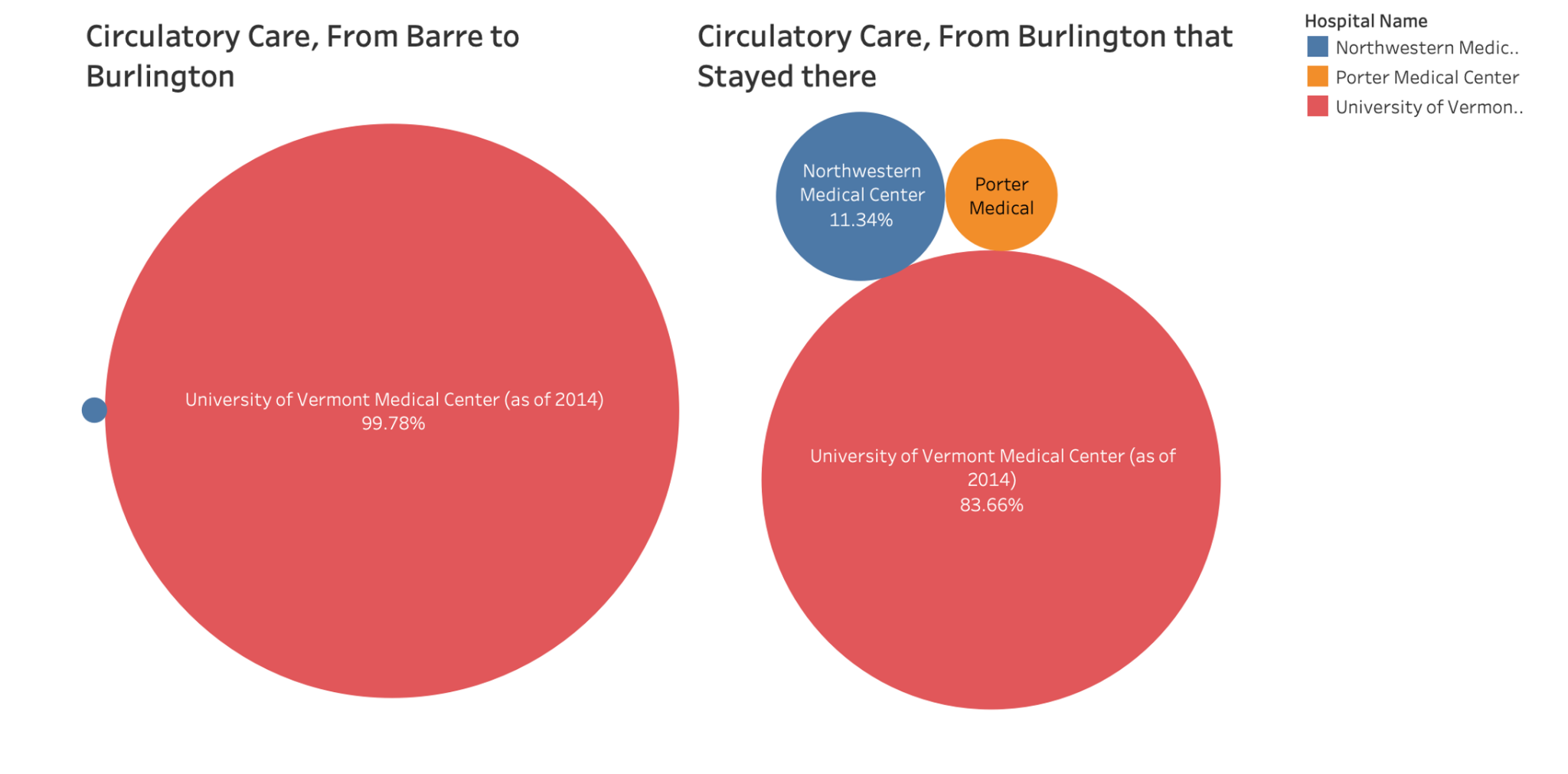
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Definition=Male Reproductive System | | | |  |  |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 97.92 | 1.04 | 0 | 0 | 1.04 |
| RR2--Barre | 34.48 | 65.52 | 0 | 0 | 0 |
| RR3--Newport/St. Jns | 44.44 | 11.11 | 44.44 | 0 | 0 |
| RR4--Randolph | 3.7 | 0 | 0 | 85.19 | 11.11 |
| RR5--Rutland | 17.65 | 0 | 0 | 2.94 | 79.41 |
| Market share | 58.97 | 10.77 | 2.05 | 12.31 | 15.9 |

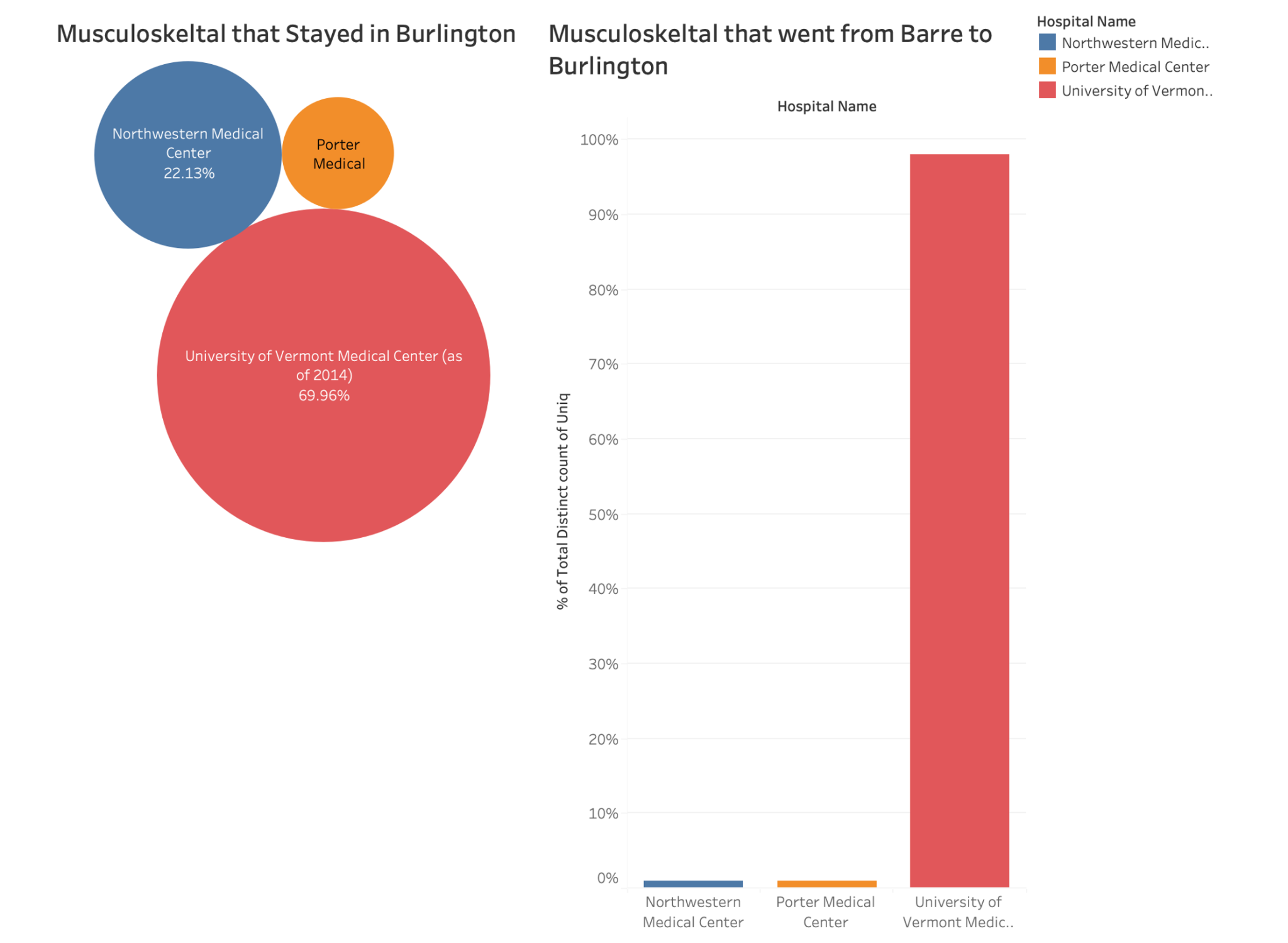
100 Percent patients from Burlington are still in Burlington for treatment. 73% of patients travel from Barre to Burlington for this service. On the other hand, from the low end of the spectrum, the male reproductive system, 34.5% of patients travel from Barre to Burlington, and 98% of patients from Burlington remain there for treatment.

In all cases, the market share is highly controlled by Burlington. For circulatory 55%, for musculoskeletal 45%, for myeloproliferative 77%, and for male reproductive 59%.

Where do patients go if they went to Region One Burlington? What percentage of patients are admitted there?

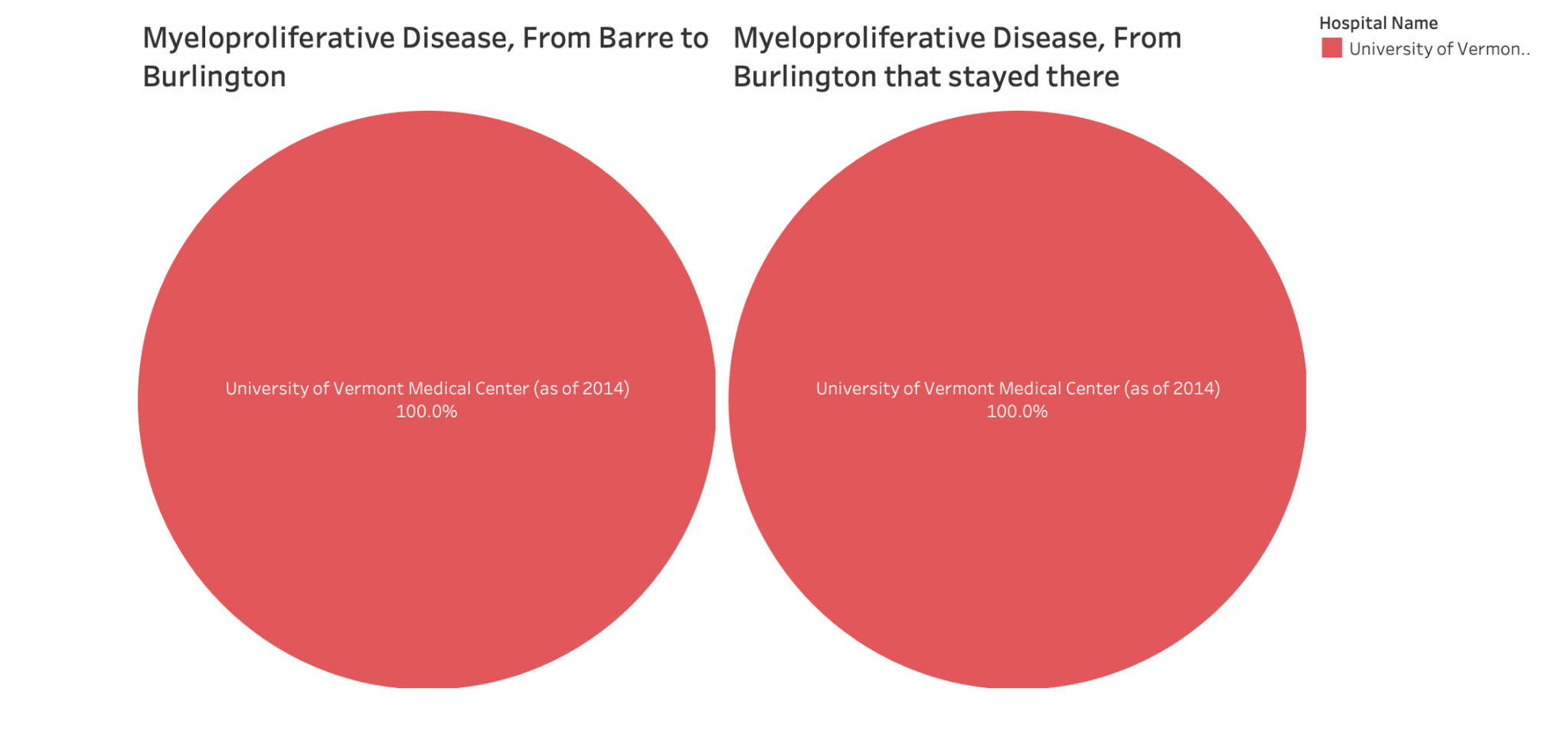
Since most patients are from high-end, circulatory, and myeloproliferative disorders, these are evaluated first.





Compared to the Musculoskeletal, the University of Vermont has the highest percentage of circulatory patients, both those who relocated to the region from Bare or those who were already residents.

The pattern is the same for Myeloproliferative Disease as well. But here the other Hospitals do not even have to compete; it is an absolute monopoly.



# Difference Between Government and Commercial Care

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Level\_Of\_Care=Low END Payment=Government | | | | |  |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 89.2 | 8.34 | 0.16 | 0.24 | 2.07 |
| RR2--Barre | 13.98 | 79.65 | 1.59 | 4.07 | 0.71 |
| RR3--Newport/St. Jns | 7.67 | 6.67 | 82.33 | 3.33 | 0 |
| RR4--Randolph | 2.02 | 2.02 | 0.61 | 84.85 | 10.51 |
| RR5--Rutland | 5.09 | 0.28 | 0 | 2.34 | 92.3 |
| Market Share | 38.02 | 17.54 | 7.8 | 14.14 | 22.5 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Level\_Of\_Care=Low END Payment=Private | | | |  |  |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 86.74 | 10.42 | 0 | 0.13 | 2.7 |
| RR2--Barre | 24.02 | 66.54 | 1.57 | 7.48 | 0.39 |
| RR3--Newport/St. Jns | 12.33 | 12.33 | 71.23 | 2.74 | 1.37 |
| RR4--Randolph | 1.91 | 2.55 | 1.27 | 78.34 | 15.92 |
| RR5--Rutland | 5.56 | 0.74 | 0 | 4.44 | 89.26 |
| Market Share | 49.77 | 17.31 | 3.79 | 10.25 | 18.88 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Level\_Of\_Care=High END Payment=Government | | | | |  |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 99.23 | 0.35 | 0 | 0.07 | 0.35 |
| RR2--Barre | 32.99 | 64.02 | 0.69 | 2.3 | 0 |
| RR3--Newport/St. Jns | 9.8 | 2.88 | 86.74 | 0.29 | 0.29 |
| RR4--Randolph | 5.98 | 1.84 | 0.23 | 88.97 | 2.99 |
| RR5--Rutland | 13.05 | 0 | 0.13 | 0.67 | 86.15 |
| Market Share | 48.6 | 15.14 | 8.07 | 10.81 | 17.38 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Level\_Of\_Care=High END Payment=Private | | | |  |  |
| RR Name | RR1--Burlington | RR2--Barre | RR3--Newport/St. Jns | RR4--Randolph | RR5--Rutland |
| RR1--Burlington | 99.23 | 0.57 | 0 | 0 | 0.19 |
| RR2--Barre | 59.36 | 40.18 | 0 | 0.46 | 0 |
| RR3--Newport/St. Jns | 15.69 | 3.92 | 80.39 | 0 | 0 |
| RR4--Randolph | 15.63 | 4.69 | 0 | 75 | 4.69 |
| RR5--Rutland | 37.89 | 0 | 0 | 2.48 | 59.63 |
| Market Share | 71.48 | 9.44 | 4.03 | 5.21 | 9.83 |

The High End and Private have the highest Market Share.

The High End and Private has 71.48 percent of the market Share. From this 99 percent of patients from region one stays there for treatment. Almost 60 percent of patients travel from Barre to Burling to. This can be compared with High end Government in which region one has 48 percent of the market share. Where, only almost 33 percent of patients travel from Barre to Burlington. When you compare the low end Service, Low End private have Almost 50 percent of the market share while the low End Government has 38 percent of the market share. When we compare people that travel from Barre to Burlington 24 percent and 13 percent respectively travel for the service. From this we can come to the conclusion High End and Private have the highest share.

# 

# 

# 

# 

# 

# 

# 

# 

# Conclusion

Hospitals have more access to the following, which gives them more power and influence in the healthcare system:

- Ability to assess the status of high-end health care services vs routine, standard care which can be acquired readily in other facilities.

- Patients who have substantial health insurance and disposable income.

- The monopolistic practices of a few large US hospitals were proven in a CNBC documentary.

short dash

# CODES

**PROC** **SQL**;

CREATE TABLE WORK.QUERY\_FOR\_VTINP16\_UPD AS

SELECT t1.hnum2,

t1.PPAY,

t1.CHRGS,

t1.hsa,

t1.UNIQ,

t1.MDC,

t2.'HSA(Patient''s Health Service Ar'n,

t2.'HSA Name'n,

t2.'RR Name'n,

t2.Name,

t2.'RR Collapsed Referral Region'n,

t3.hnum2 AS hnum21,

t3.HospitalName,

t3.RR,

t3.RRName

FROM WORK.VTINP16\_UPD t1

LEFT JOIN WORK.CASESTUDY\_O\_D\_HOSPMONOPOLY\_\_0000 t2 ON (t1.hsa = t2.'HSA(Patient''s Health Service Ar'n)

LEFT JOIN WORK.CASESTUDY\_O\_D\_HOSPMONOPOLY\_\_0001 t3 ON (t1.hnum2 = t3.hnum2)

LEFT JOIN WORK.CASESTUDY\_O\_D\_HOSPMONOPOLY\_MOODL t4 ON (t1.MDC = t4.MDC);

**QUIT**;

**PROC** **SORT**

DATA=WORK.QUERY\_FOR\_VTINP16\_UPD\_0000(KEEP=Definition RRNameDestination "RR NameOrgioin"n)

OUT=WORK.SORTTempTableSorted

;

BY Definition;

**RUN**;

TITLE;

TITLE1 "Table Analysis";

TITLE2 "Results";

FOOTNOTE;

FOOTNOTE1 "Generated by SAS (&\_SASSERVERNAME, &SYSSCPL) on %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) at %TRIM(%QSYSFUNC(TIME(), NLTIMAP25.))";

**PROC** **FREQ** DATA = WORK.SORTTempTableSorted

ORDER=INTERNAL

;

BY Definition;

TABLES "RR NameOrgioin"n \* RRNameDestination /

NOROW

NOCOL

NOPERCENT

NOCUM

SCORES=TABLE

ALPHA=**0.05**;

/\* -------------------------------------------------------------------

End of task code

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;

%***\_eg\_conditional\_dropds***(WORK.SORTTempTableSorted);

TITLE; FOOTNOTE;

**PROC** **SQL**;

CREATE TABLE WORK.QUERY\_FOR\_VTINP16\_UPD\_0000 AS

SELECT t1.hnum2,

t1.PPAY,

t1.CHRGS,

t1.hsa,

t1.UNIQ,

t1.MDC,

t1.'HSA(Patient''s Health Service Ar'n,

t1.'HSA Name'n,

t1.'RR NameOrgioin'n,

t1.Name,

t1.'RR Collapsed Referral Region'n,

t1.HospitalName,

t1.RR,

t1.RRNameDestination,

t1.Definition

FROM WORK.QUERY\_FOR\_VTINP16\_UPD t1

WHERE t1.'RR NameOrgioin'n NOT CONTAINS 'Z\_OutState';

**QUIT**;

**PROC** **SQL**;

CREATE TABLE WORK.QUERY\_FOR\_VTINP16\_UPD\_58 AS

SELECT t1.hnum2,

t1.PPAY,

t1.CHRGS,

t1.hsa,

t1.UNIQ,

t1.MDC,

t1.'HSA(Patient''s Health Service Ar'n,

t1.'HSA Name'n,

t1.'RR NameOrgioin'n,

t1.Name,

t1.'RR Collapsed Referral Region'n,

t1.HospitalName,

t1.RR,

t1.RRNameDestination,

t1.Definition,

/\* New variable Payment \*/

CASE

WHEN t1.PPAY = **1** THEN 'Government'

WHEN t1.PPAY IN (**6**, **7**) THEN 'Private'

ELSE ''

END AS Payment,

/\* New variable Level\_Of\_Care \*/

CASE

WHEN t1.MDC = **5** THEN 'High END'

WHEN t1.MDC = **8** THEN 'Low END'

ELSE 'NONE'

END AS Level\_Of\_Care

FROM WORK.QUERY\_FOR\_VTINP16\_UPD\_0000 t1

WHERE t1.PPAY IN (**1**, **6**, **7**) /\* Keep only specific PPAY values \*/

OR t1.MDC IN (**5**, **8**); /\* Keep only specific MDC values \*/

**QUIT**;

For the second table

**PROC** **SORT**

DATA=WORK.QUERY\_FOR\_VTINP16\_UPD\_58\_0003(KEEP=Level\_Of\_Care Payment "RR NameOrgioin"n RRNameDestination)

OUT=WORK.SORTTempTableSorted

;

BY Level\_Of\_Care Payment;

**RUN**;

TITLE;

TITLE1 "Table Analysis";

TITLE2 "Results";

FOOTNOTE;

FOOTNOTE1 "Generated by SAS (&\_SASSERVERNAME, &SYSSCPL) on %TRIM(%QSYSFUNC(DATE(), NLDATE20.)) at %TRIM(%QSYSFUNC(TIME(), NLTIMAP25.))";

**PROC** **FREQ** DATA = WORK.SORTTempTableSorted

ORDER=INTERNAL

;

BY Level\_Of\_Care Payment;

TABLES "RR NameOrgioin"n \* RRNameDestination /

NOROW

NOCOL

NOFREQ

NOCUM

SCORES=TABLE

ALPHA=**0.05**;

/\* -------------------------------------------------------------------

End of task code

------------------------------------------------------------------- \*/

**RUN**; **QUIT**;