

# Supplementary Information

## **Ensemble-Imbalance-based classification for amyotrophic lateral sclerosis prognostic prediction: identifying short-survival patients at diagnosis**

Fabiano Papaiz <sup>1,2,3\*</sup>, Mario Emílio Teixeira Dourado Jr. <sup>1,4</sup>,  
Ricardo Alexsandro de Medeiros Valentim <sup>1</sup>, Rafael Pinto <sup>1,5</sup>,  
Antonio Higor Freire de Moraes <sup>1,3</sup>, Joel Perdiz Arrais<sup>2</sup>

<sup>1</sup>Laboratory of Technological Innovation in Health (LAIS), Federal University of Rio Grande do Norte, Natal, Brazil.

<sup>2</sup>Centre for Informatics and Systems, Department of Informatics Engineering, University of Coimbra, Coimbra,

Portugal. <sup>3</sup>Advanced Nucleus of Technological Innovation (NAVI), Federal Institute of Rio Grande do Norte, Natal,

Brazil. <sup>4</sup>Department of Internal Medicine, Federal University of Rio Grande do Norte, Natal, Brazil.

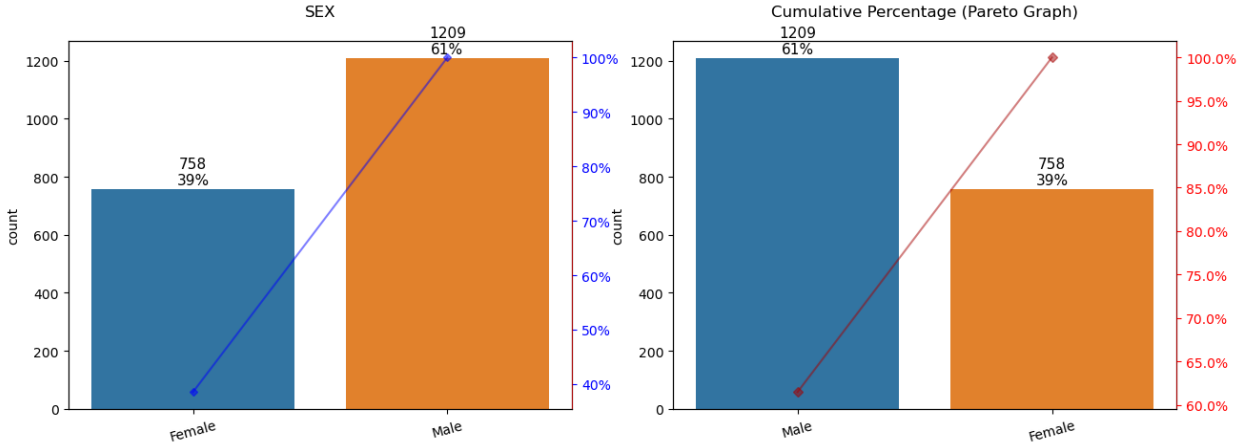
<sup>5</sup>Information Systems Coordination, Federal Institute of Rio Grande do Norte, Natal, Brazil.

### Table of Contents

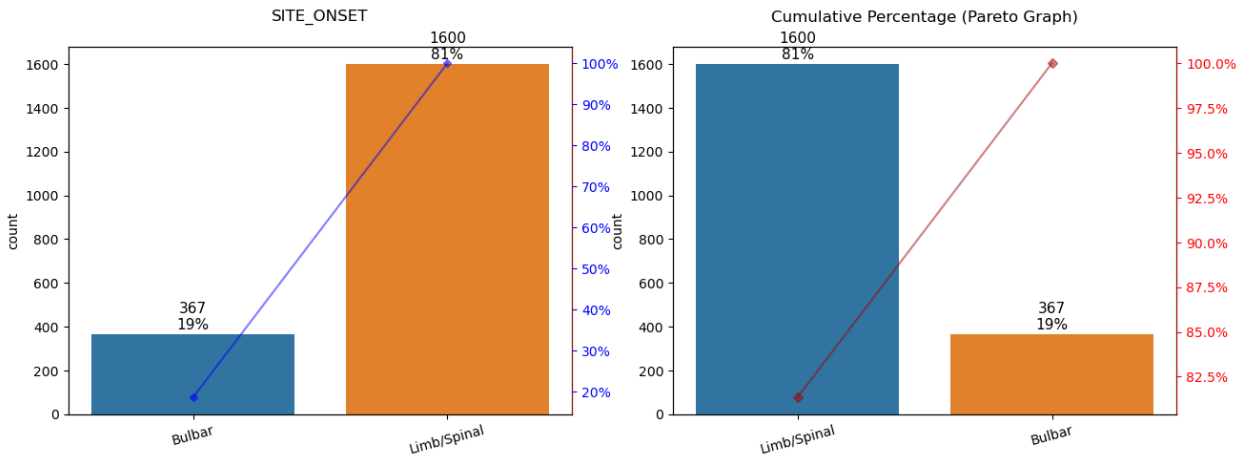
<b><i>Input and Output Variable Distributions</i></b> .....	<b>2</b>
<b><i>Compare the variable distributions for the 2 Survival Groups: Short and Non-Short</i></b> .....	<b>10</b>
<b><i>Compare the output variable distribution for the Training and Validation subsets used to train and validate the machine learning models</i></b> .....	<b>18</b>

# Input and Output Variable Distributions.

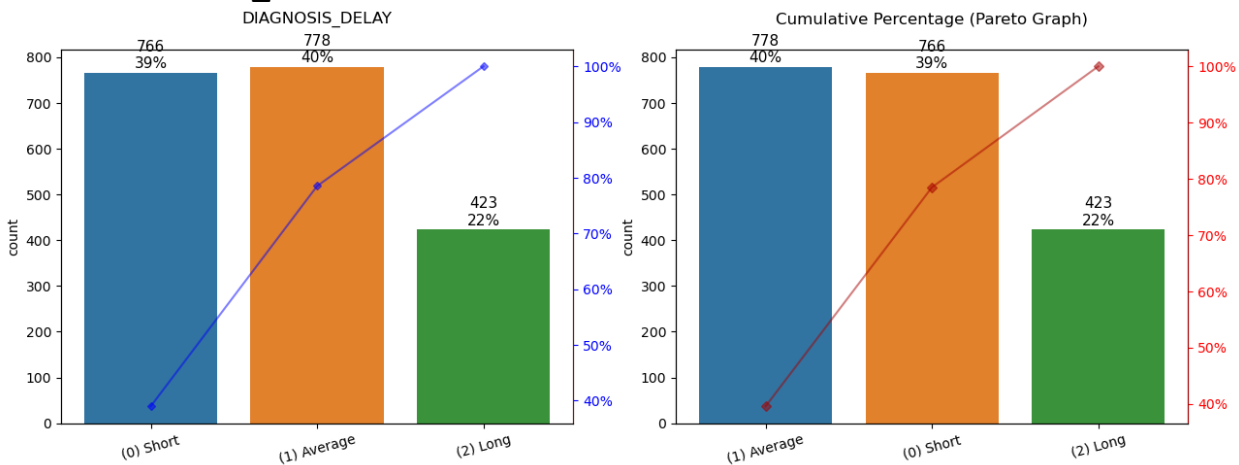
## Column Sex



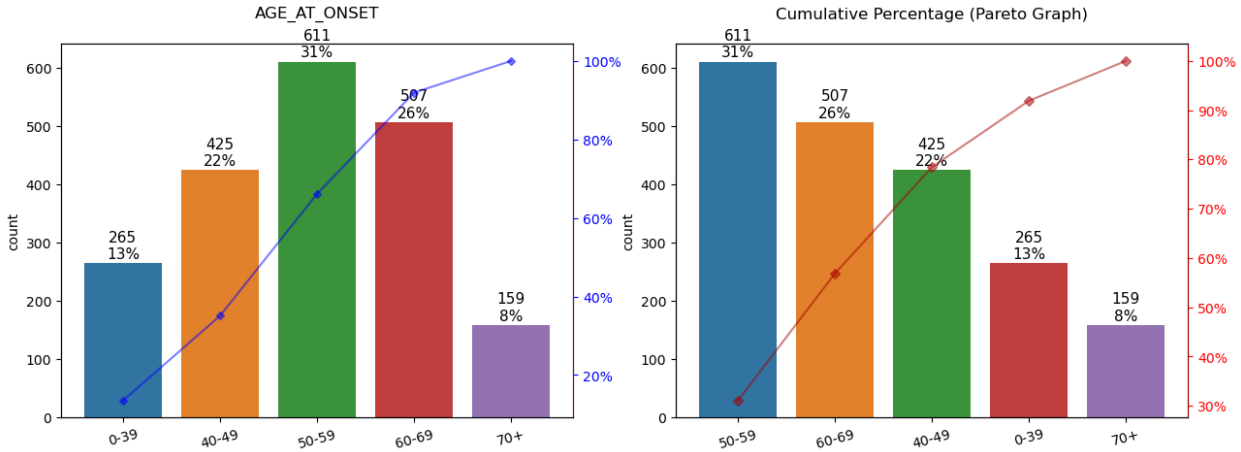
## Column Site\_Onset



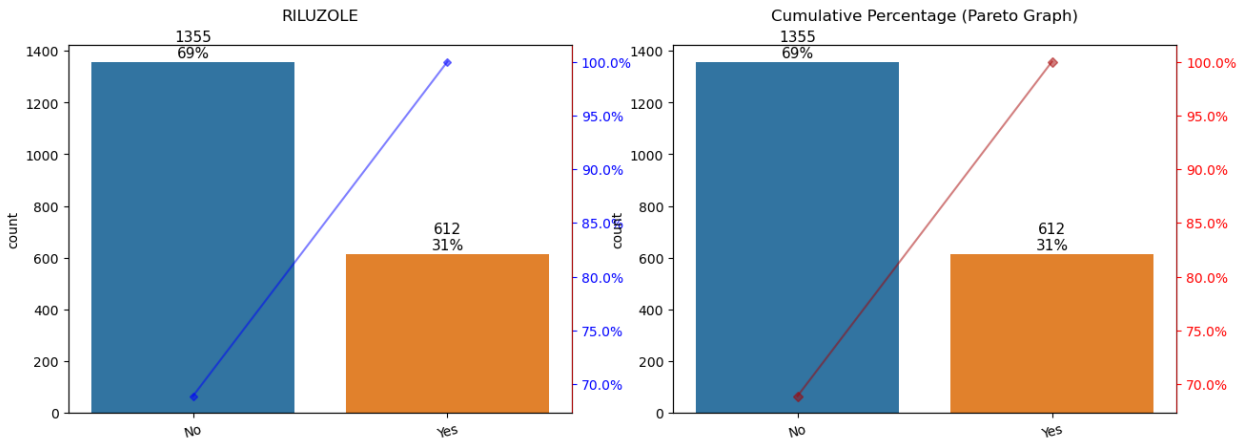
## Column Diagnosis\_Delay



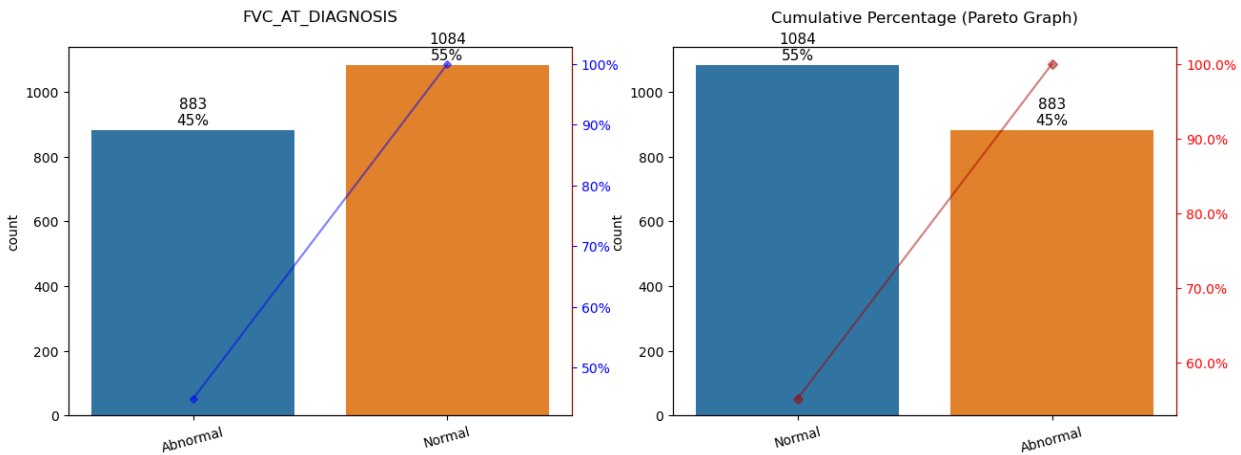
**Column Age\_at\_Onset**



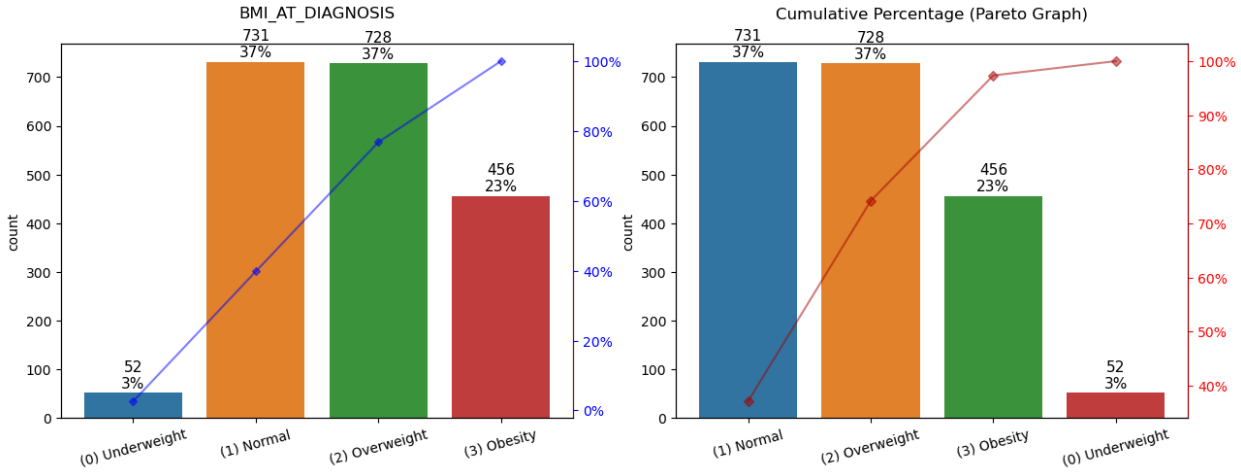
**Column Riluzole**



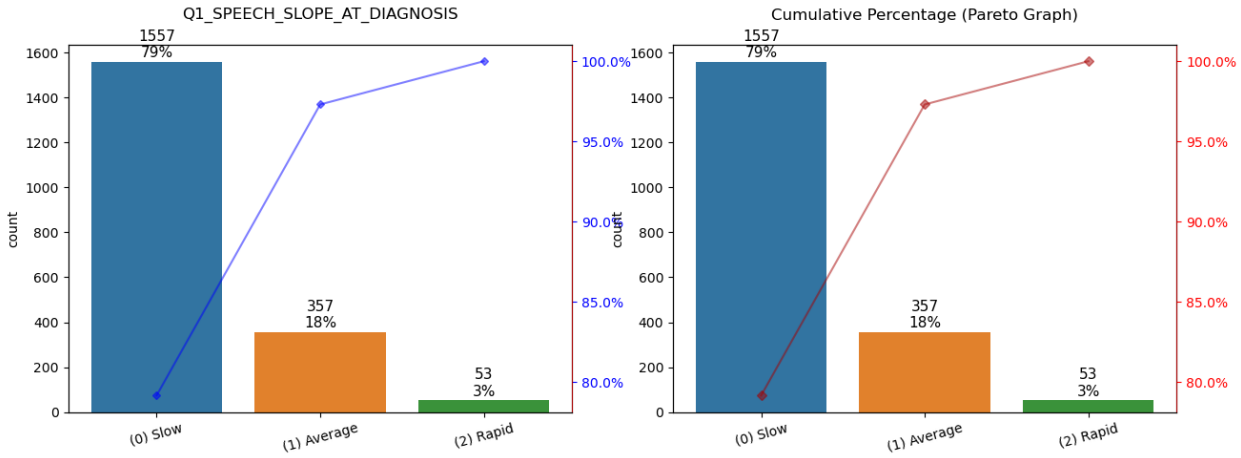
**Column FVC\_at\_Diagnosis**



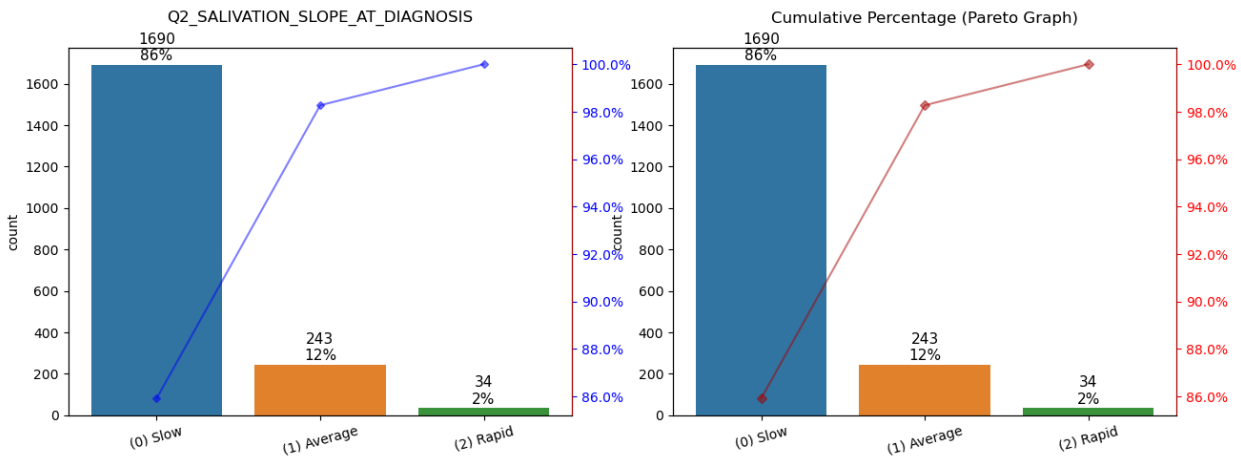
### Column BMI\_at\_Diagnosis



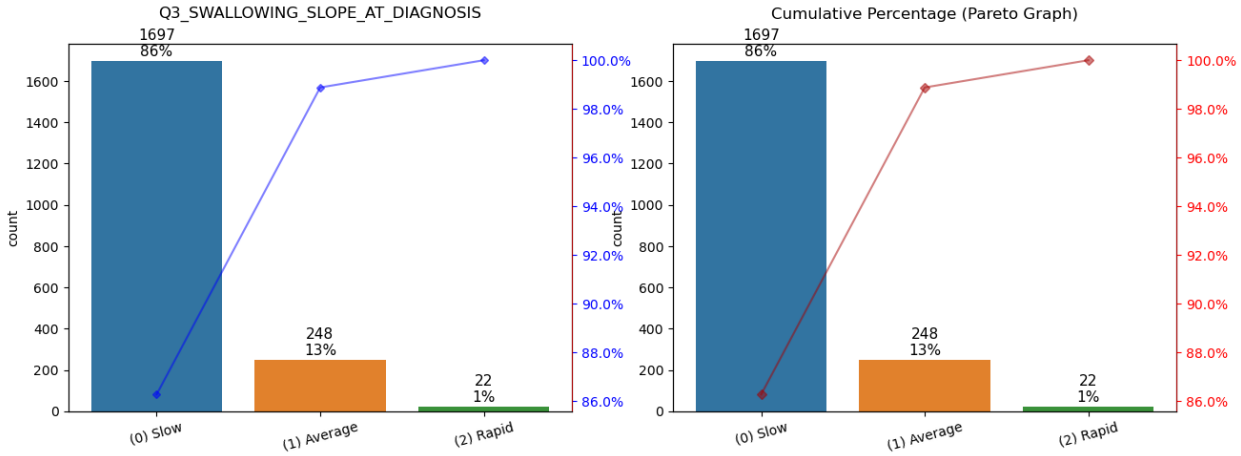
### Column Q1\_Speech\_slope\_at\_Diagnosis



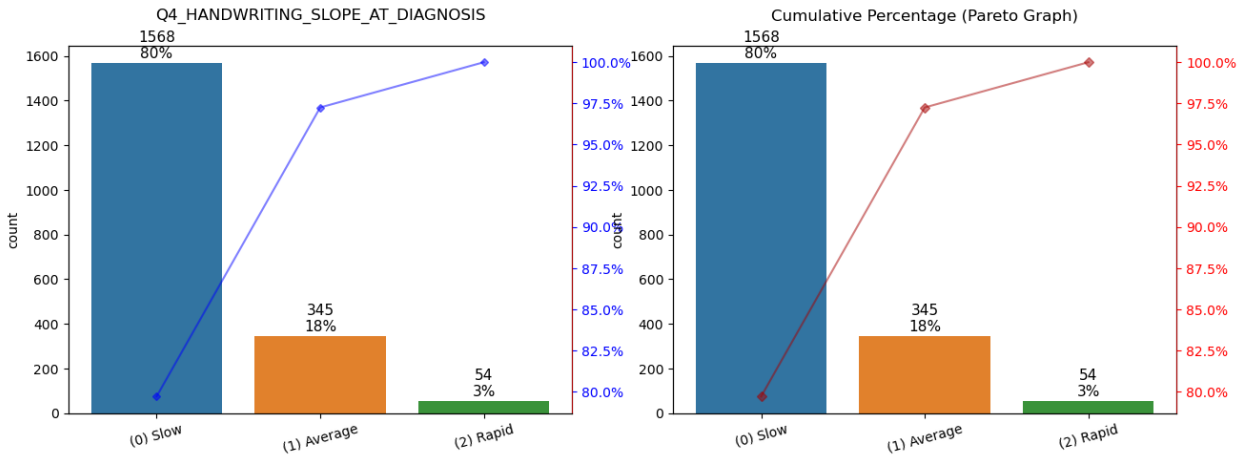
### Column Q2\_Salivation\_slope\_at\_Diagnosis



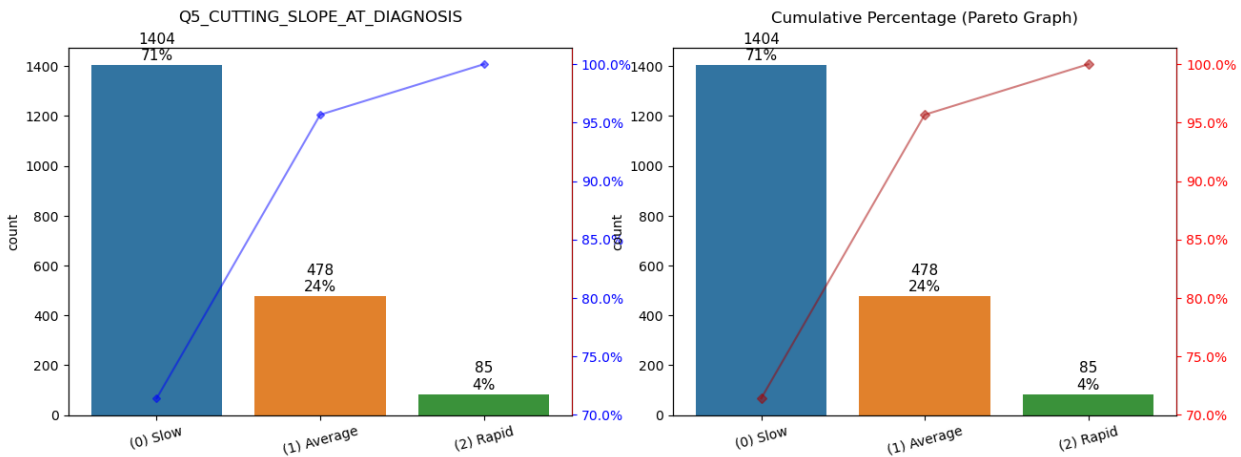
### Column Q3\_Swallowing\_slope\_at\_Diagnosis



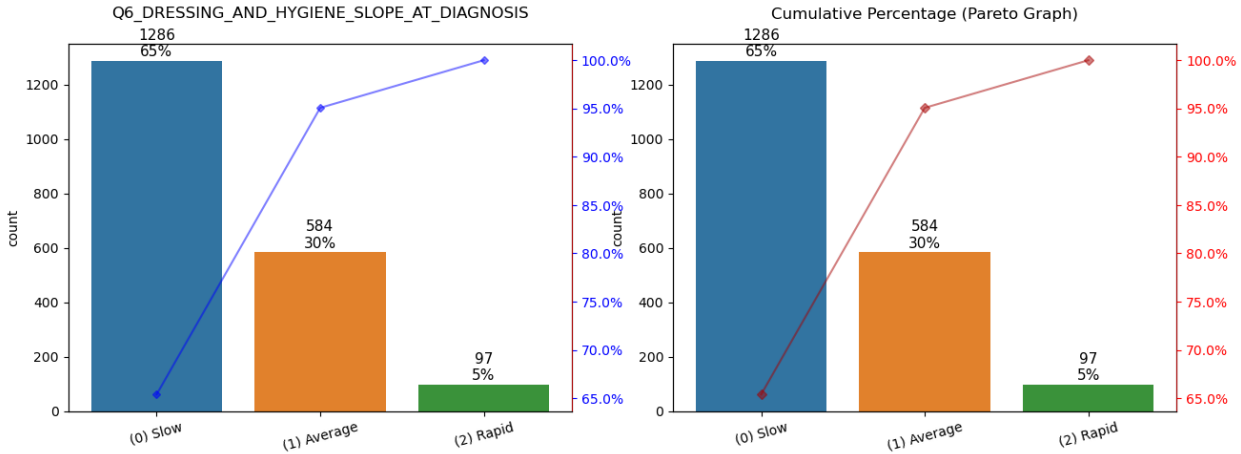
### Column Q4\_Handwriting\_slope\_at\_Diagnosis



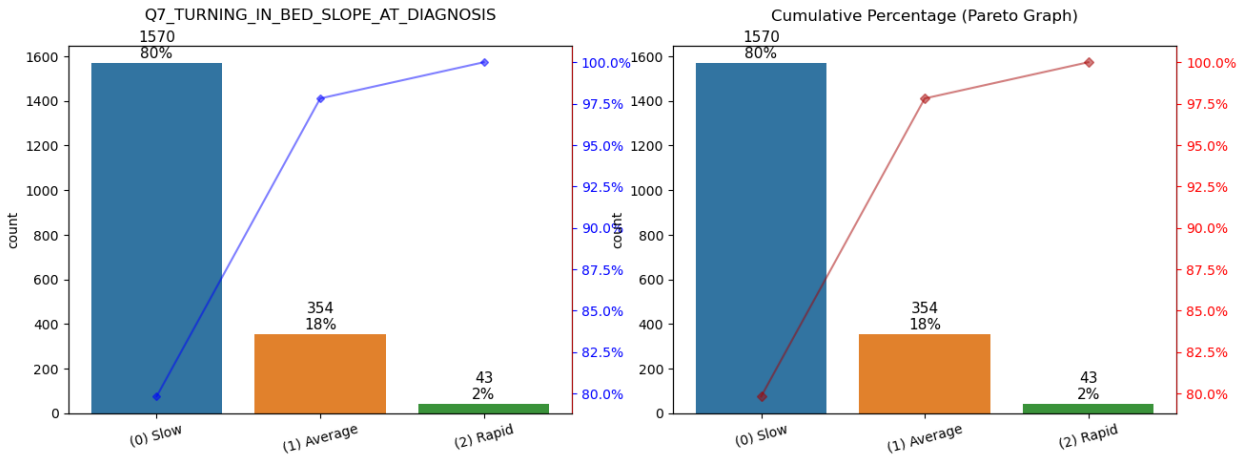
### Column Q5\_Cutting\_slope\_at\_Diagnosis



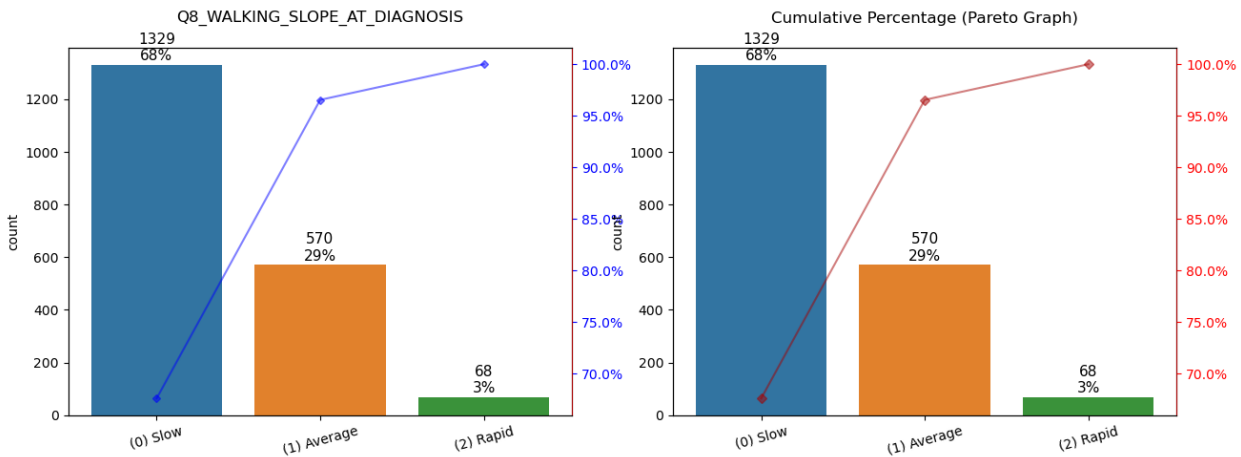
**Column Q6\_Dressing\_and\_Hygiene\_slope\_at\_Diagnosis**



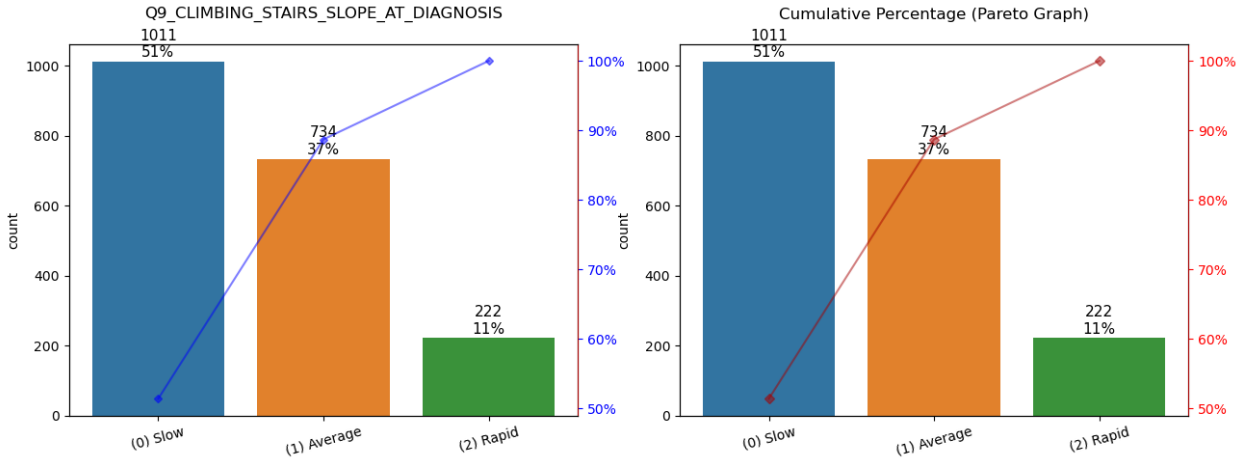
**Column Q7\_Turning\_in\_Bed\_slope\_at\_Diagnosis**



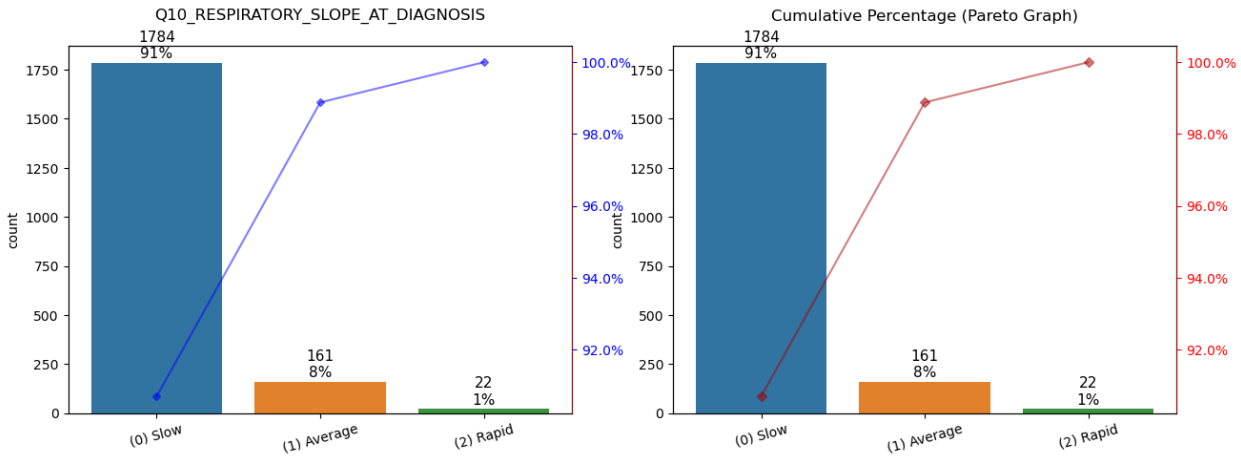
**Column Q8\_Walking\_slope\_at\_Diagnosis**



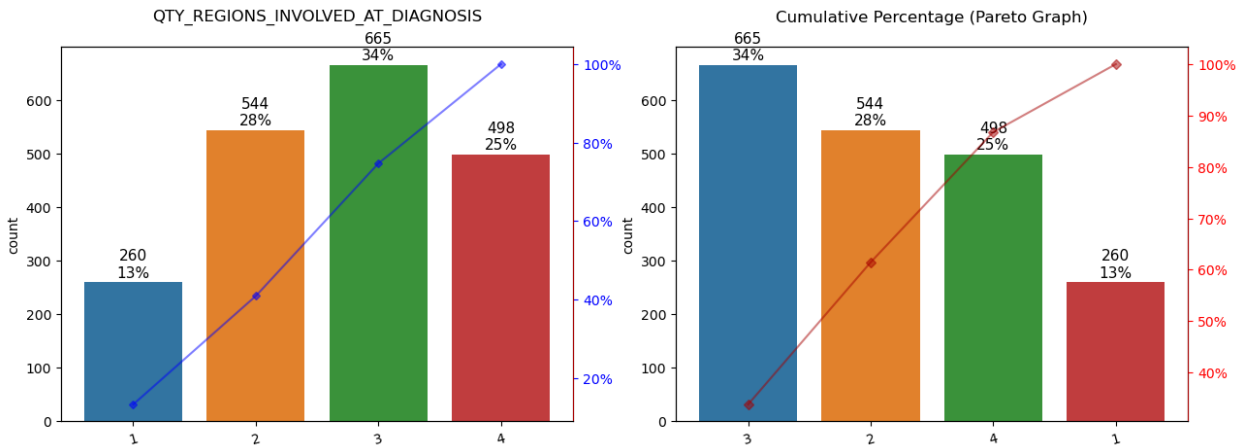
**Column Q9\_Climbing\_Stairs\_slope\_at\_Diagnosis**



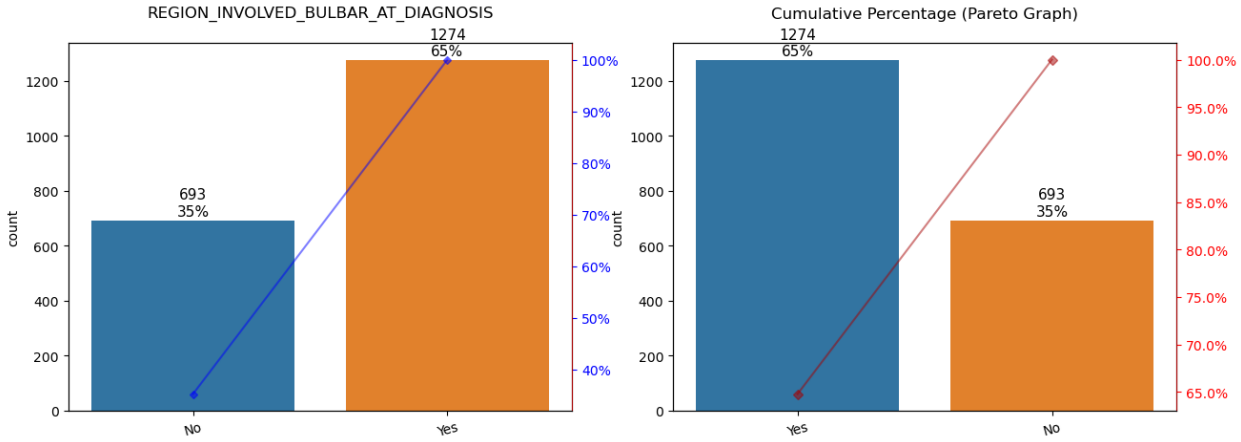
**Column Q10\_Respiratory\_slope\_at\_Diagnosis**



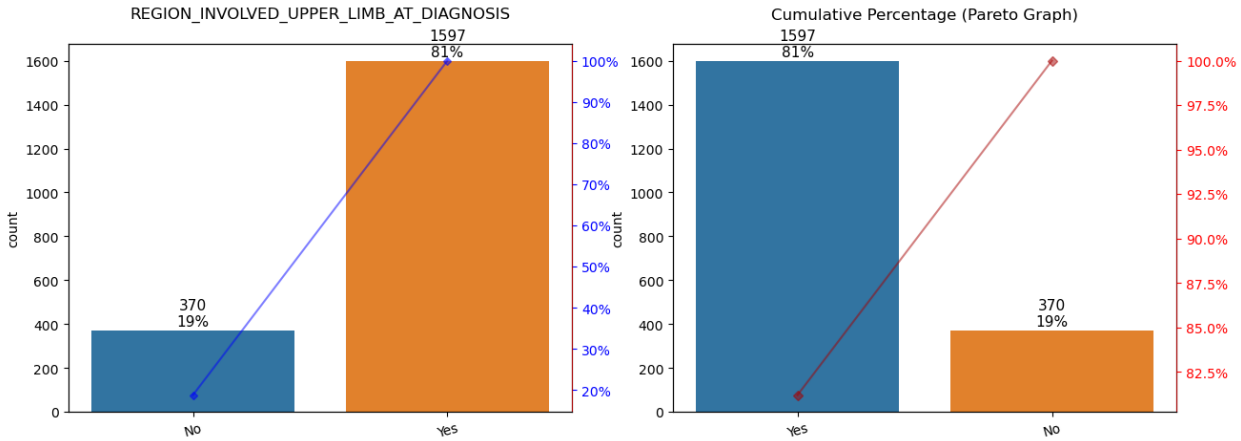
**Column Qty\_Regions\_Involved\_at\_Diagnosis**



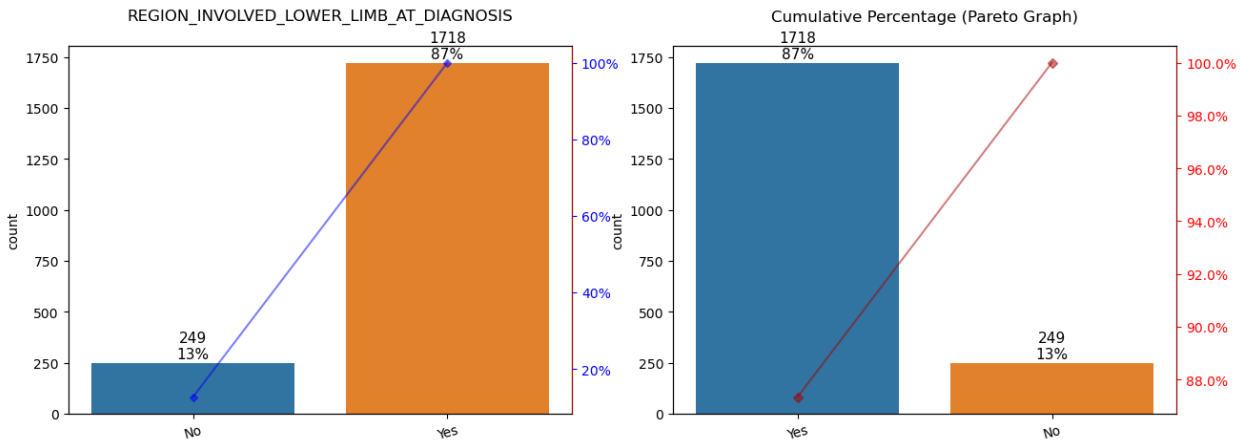
### Column Region\_Involved\_Bulbar\_at\_Diagnosis



### Column Region\_Involved\_Upper\_Limb\_at\_Diagnosis

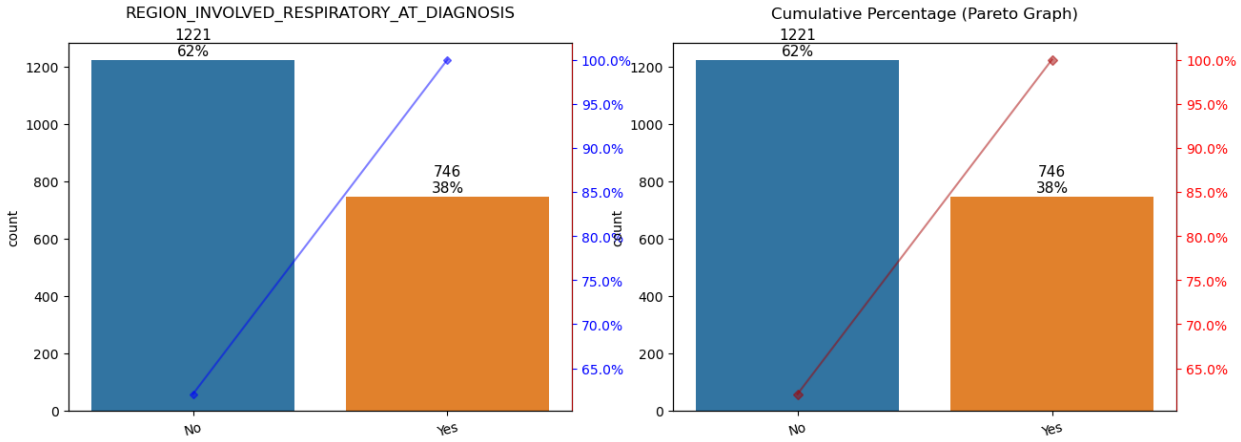


### Column Region\_Involved\_Lower\_Limb\_at\_Diagnosis

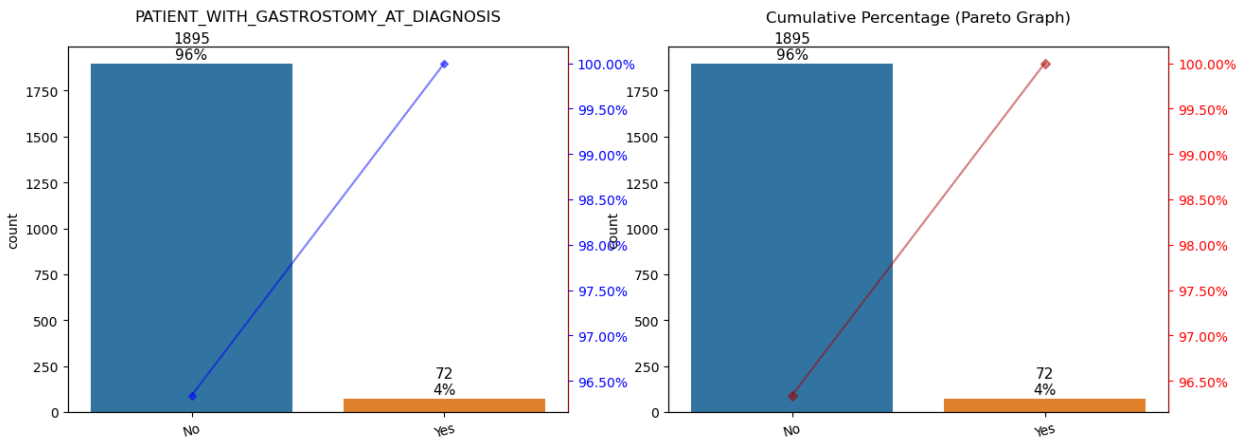




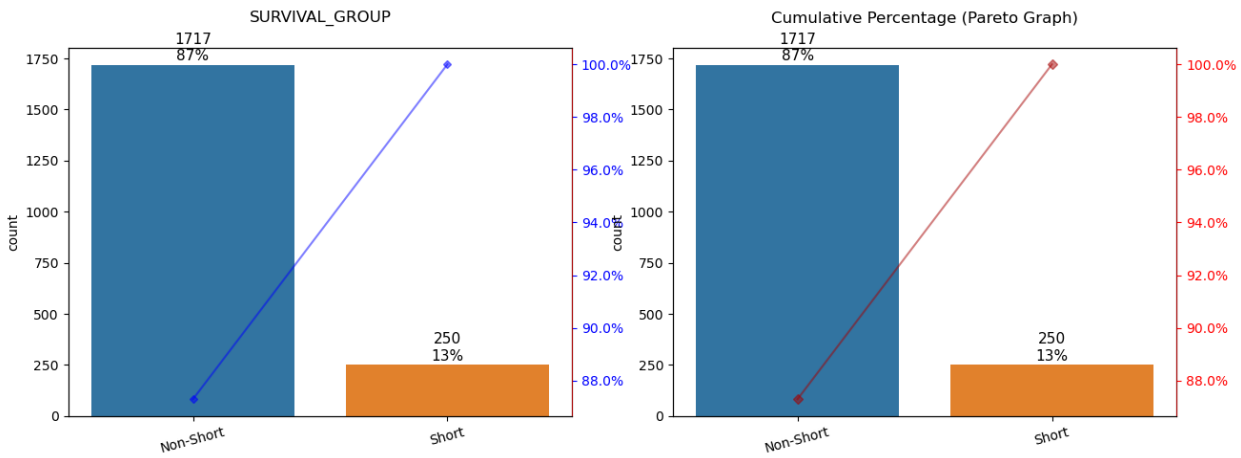
**Column Region\_Involved\_Respiratory\_at\_Diagnosis**



**Column Patient\_with\_Gastrostomy\_at\_Diagnosis**

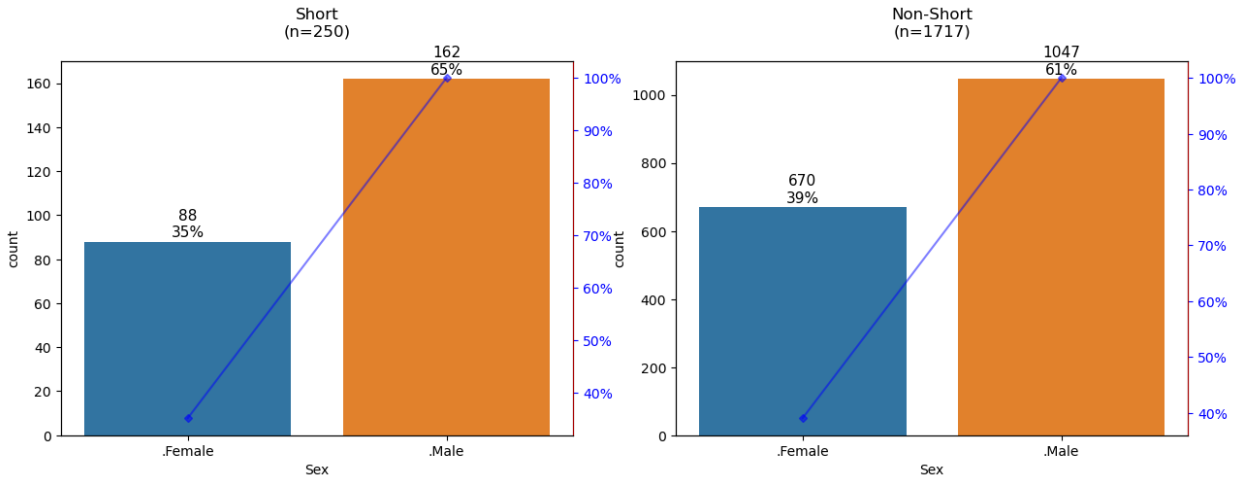


**OUTPUT Column Survival\_Group**

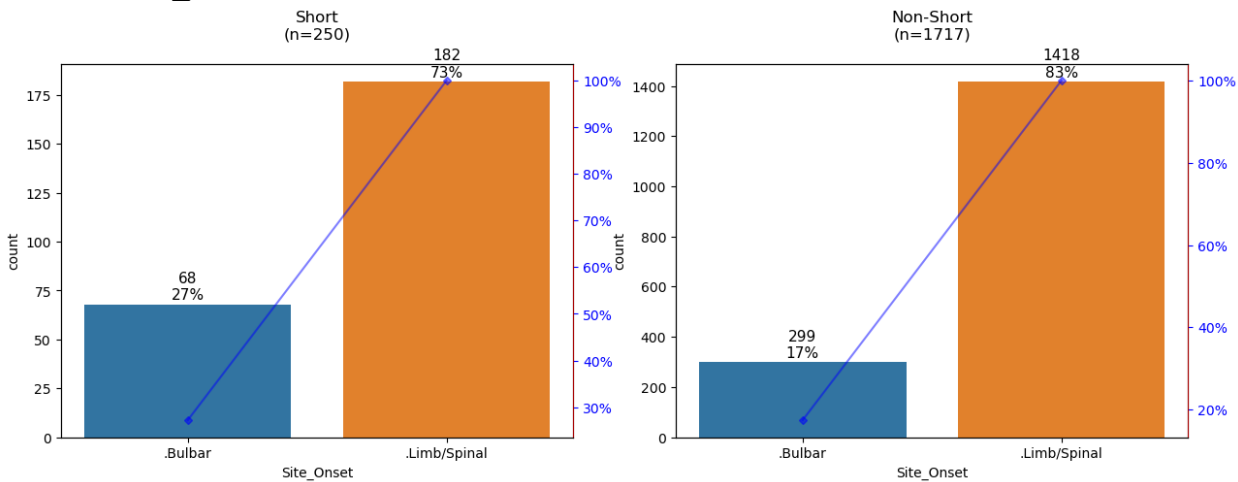


# Compare the variable distributions for the 2 Survival Groups: *Short* and *Non-Short*.

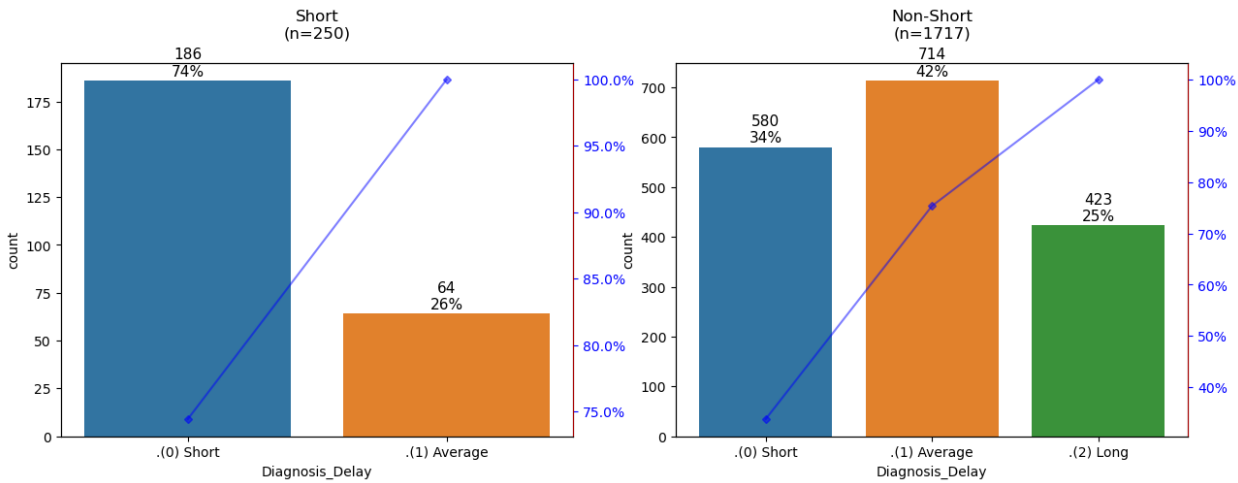
## Column Sex



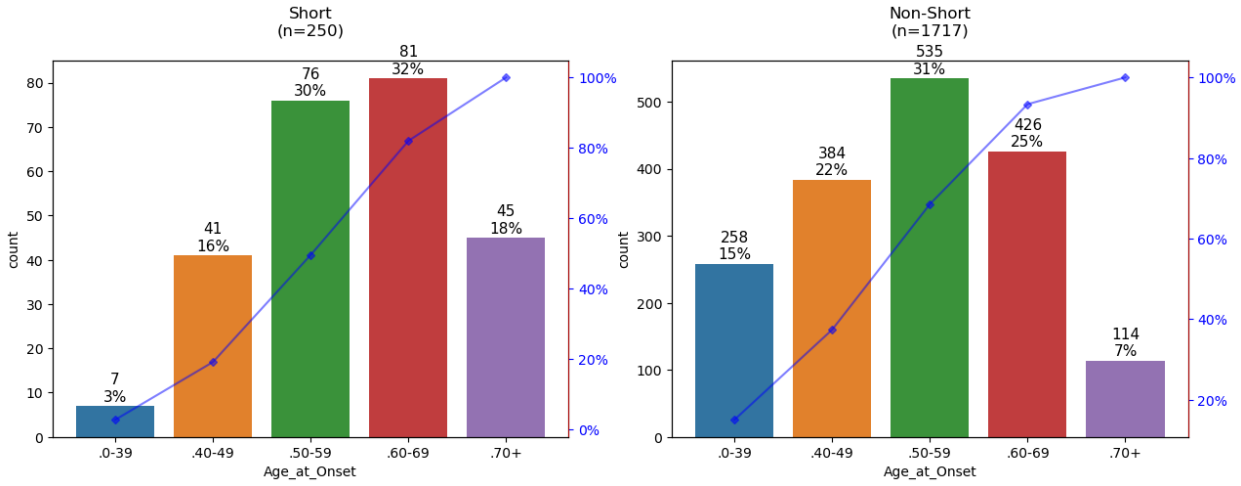
## Column Site\_Onset



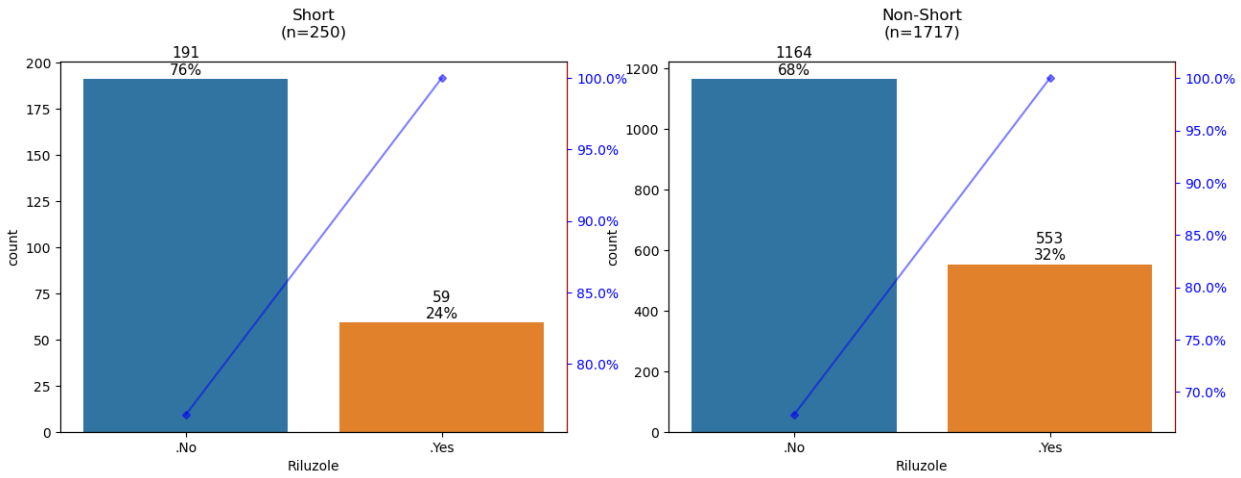
## Column Diagnosis\_Delay



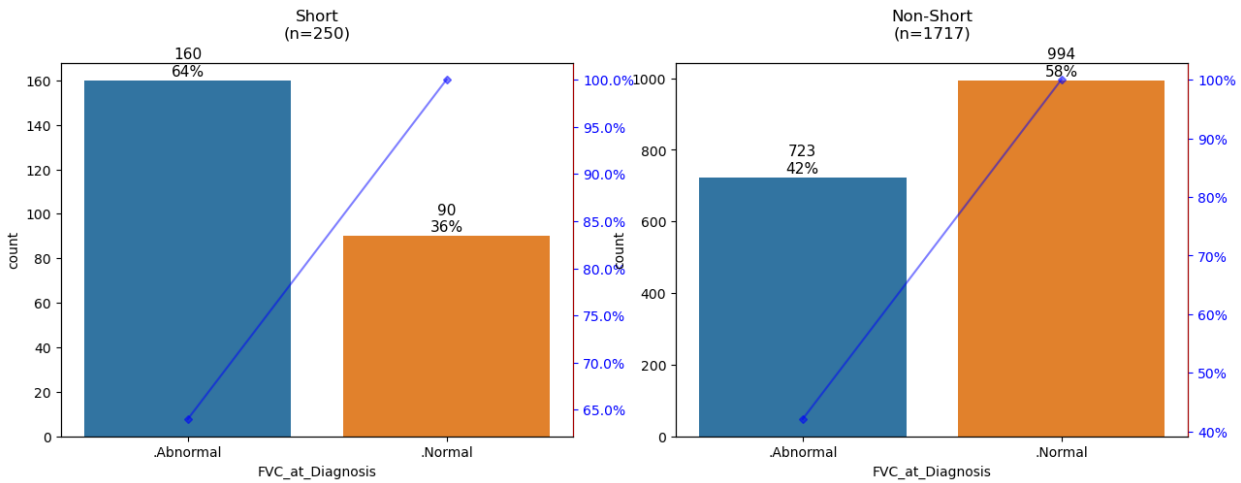
**Column Age\_at\_Onset**



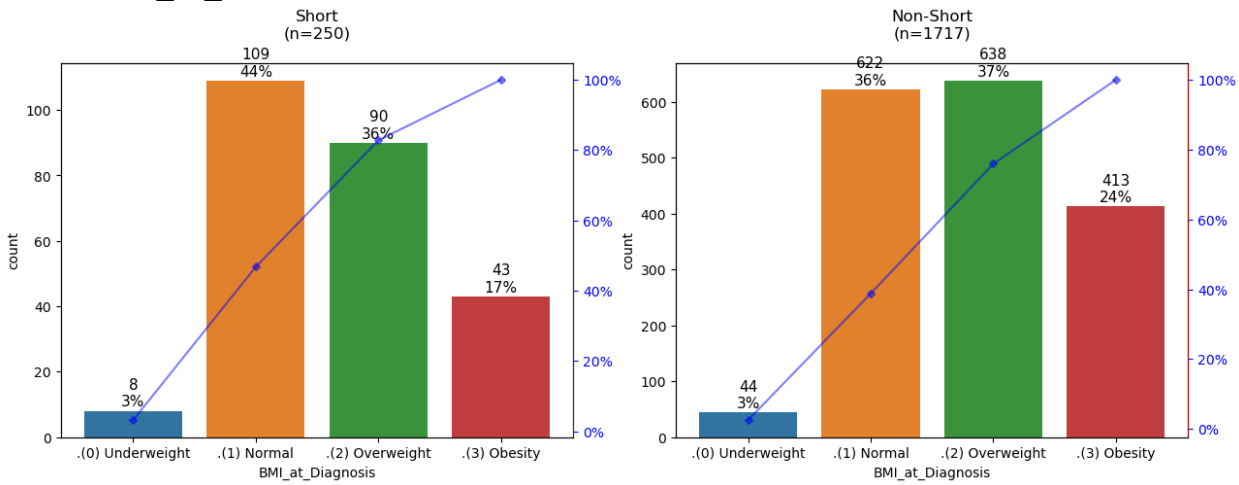
**Column Riluzole**



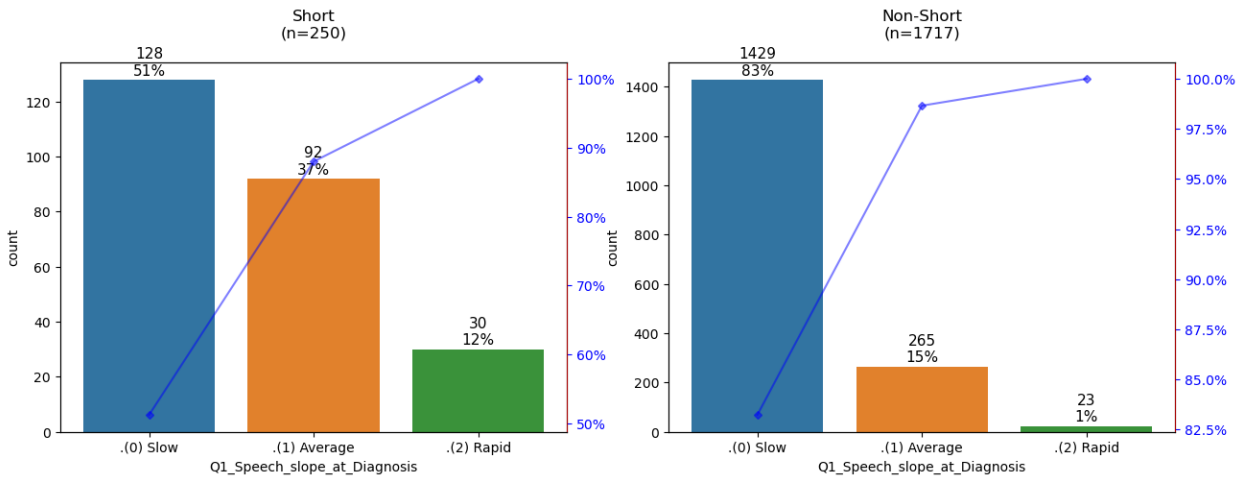
**Column FVC\_at\_Diagnosis**



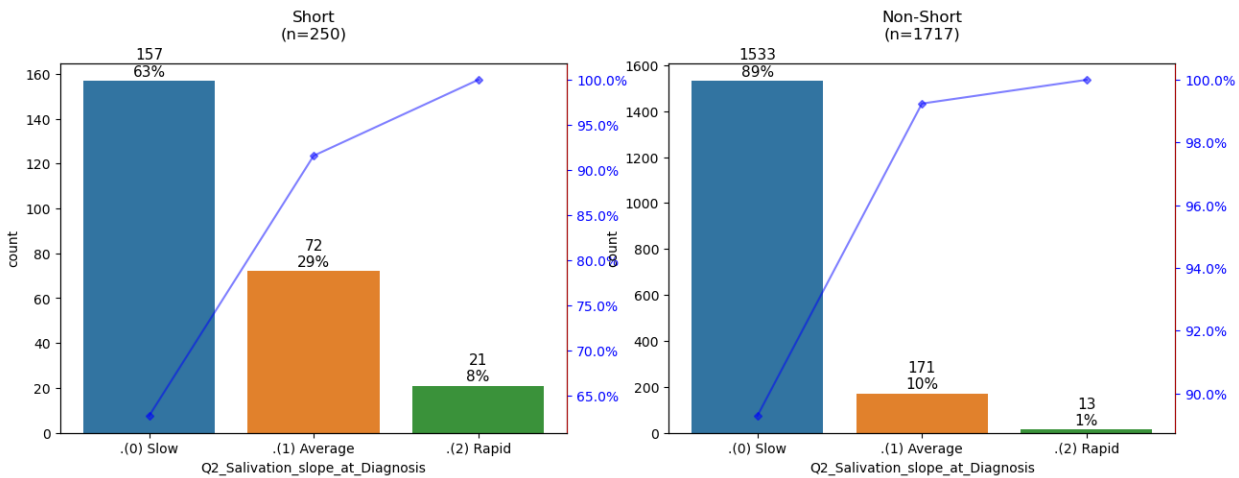
### Column BMI\_at\_Diagnosis



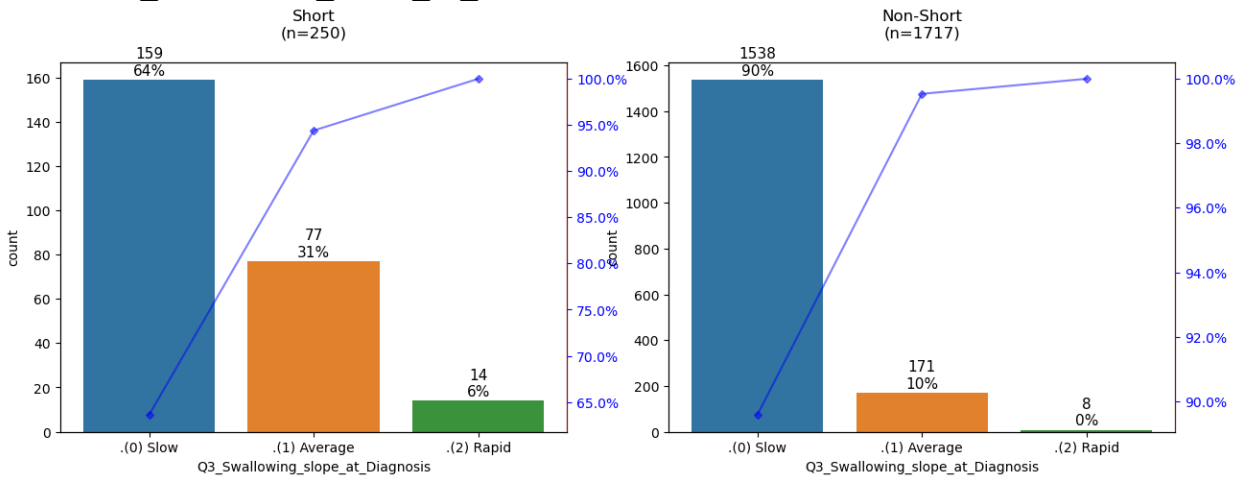
### Column Q1\_Speech\_slope\_at\_Diagnosis



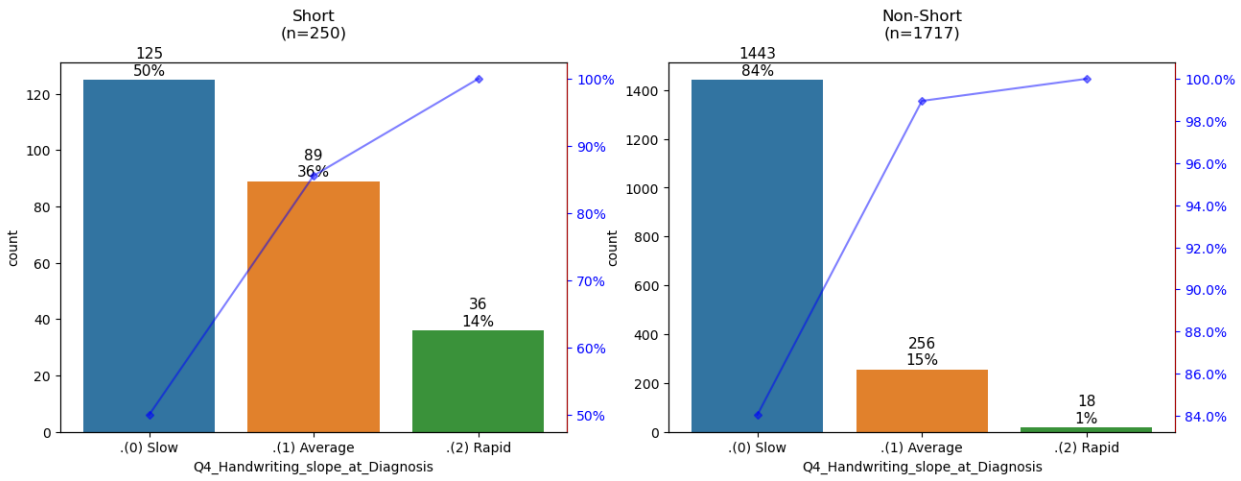
### Column Q2\_Salivation\_slope\_at\_Diagnosis



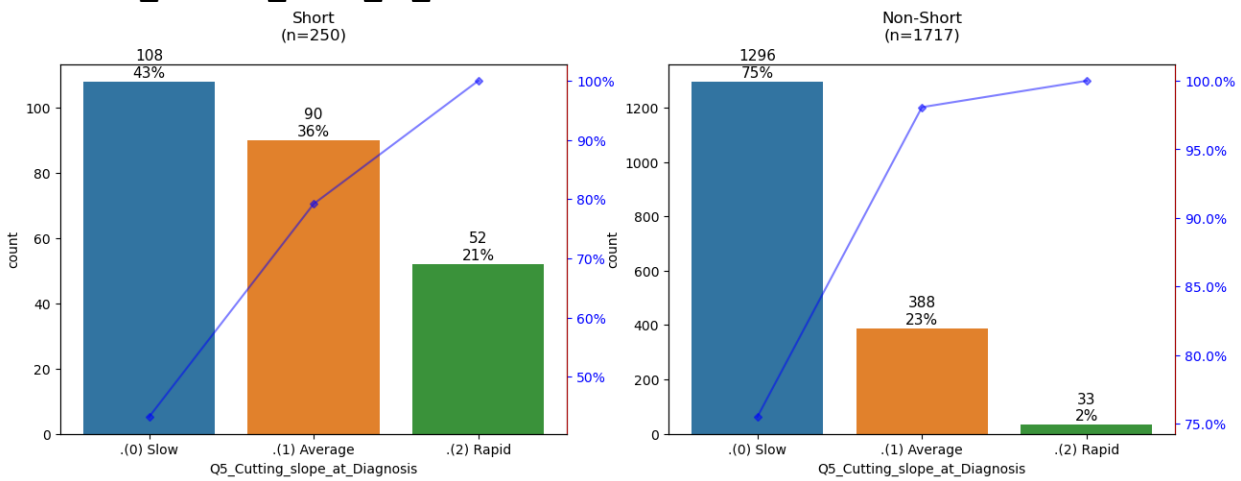
### Column Q3\_Swallowing\_slope\_at\_Diagnosis



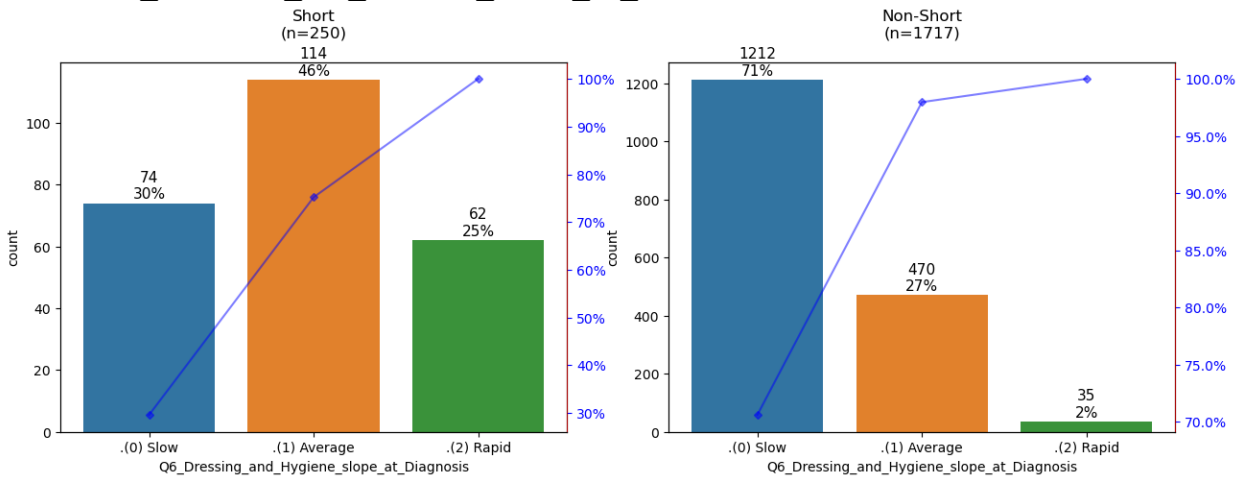
### Column Q4\_Handwriting\_slope\_at\_Diagnosis



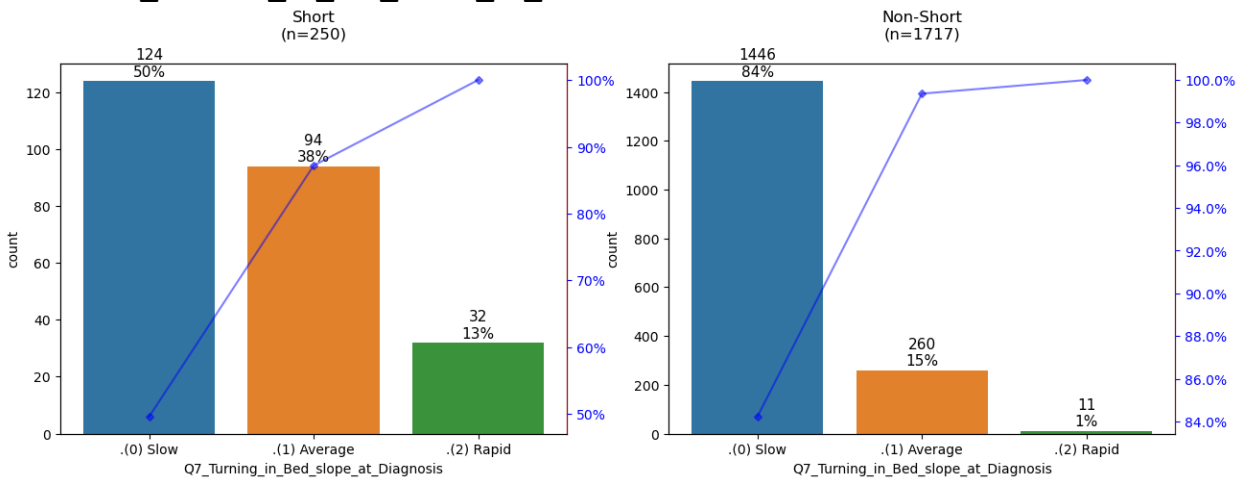
### Column Q5\_Cutting\_slope\_at\_Diagnosis



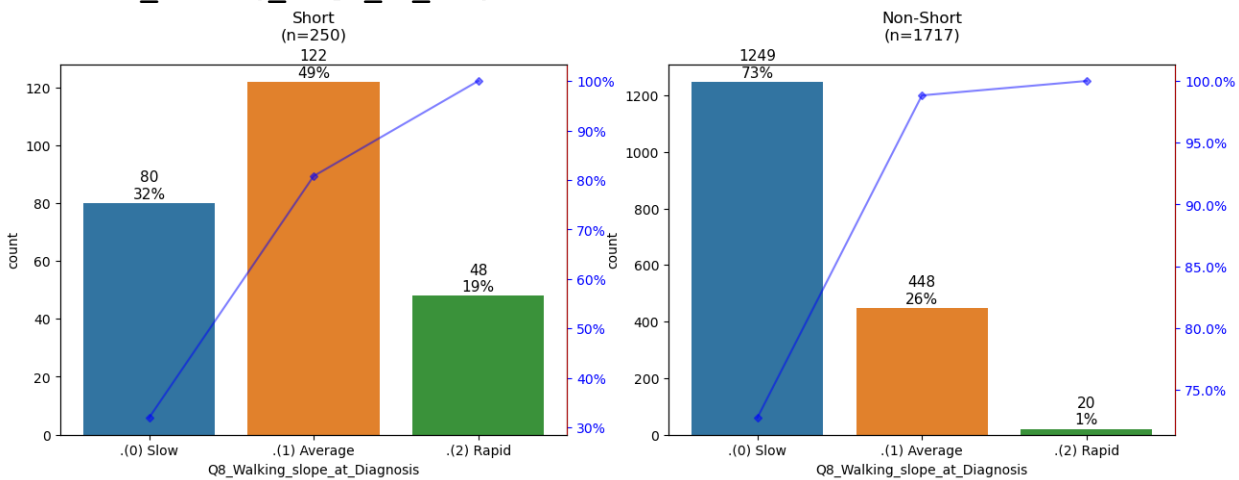
**Column Q6\_Dressing\_and\_Hygiene\_slope\_at\_Diagnosis**



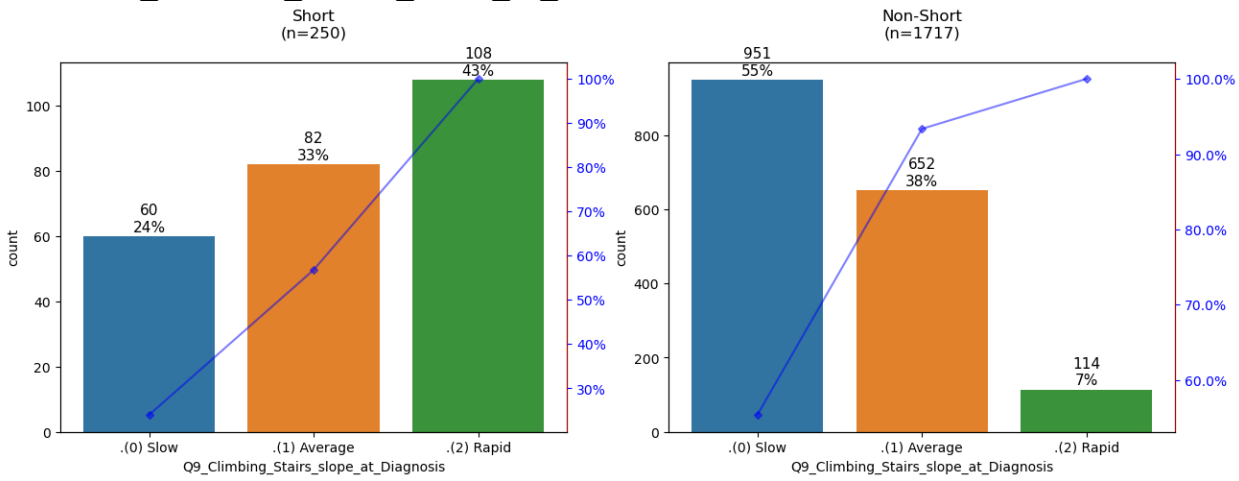
**Column Q7\_Turning\_in\_Bed\_slope\_at\_Diagnosis**



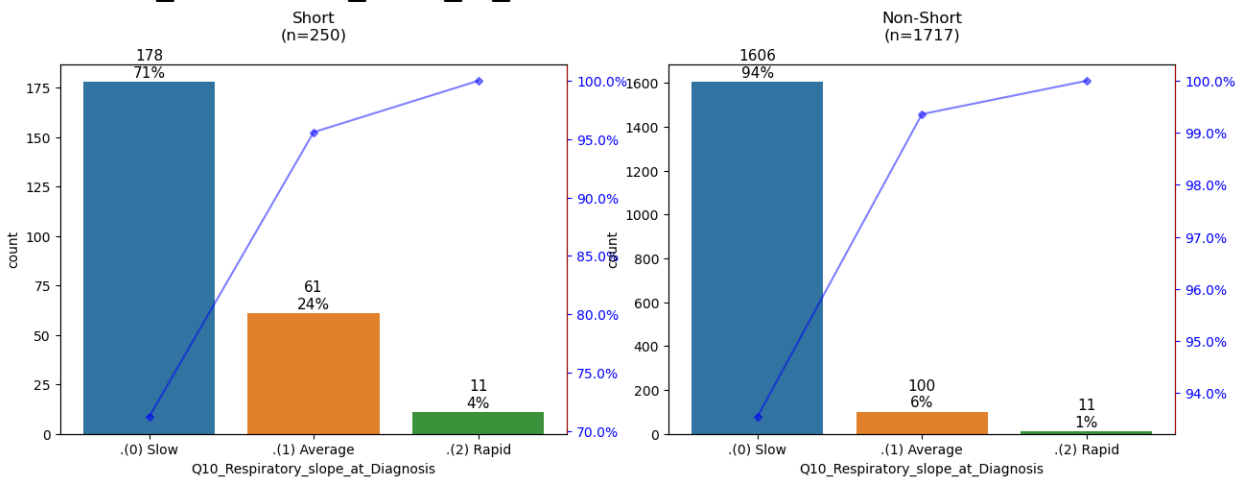
**Column Q8\_Walking\_slope\_at\_Diagnosis**



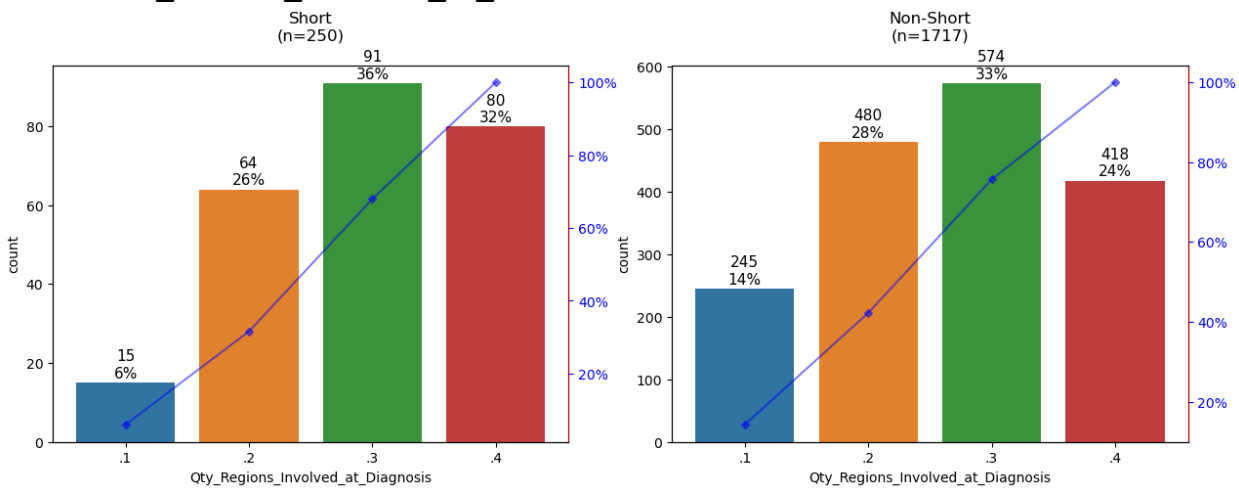
Column Q9\_Climbing\_Stairs\_slope\_at\_Diagnosis



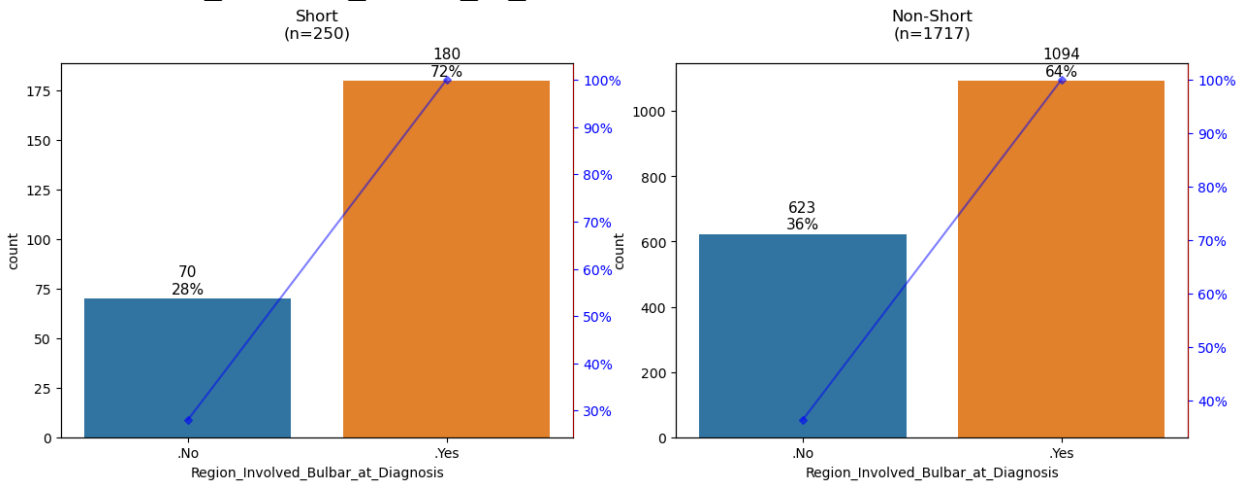
Column Q10\_Respiratory\_slope\_at\_Diagnosis



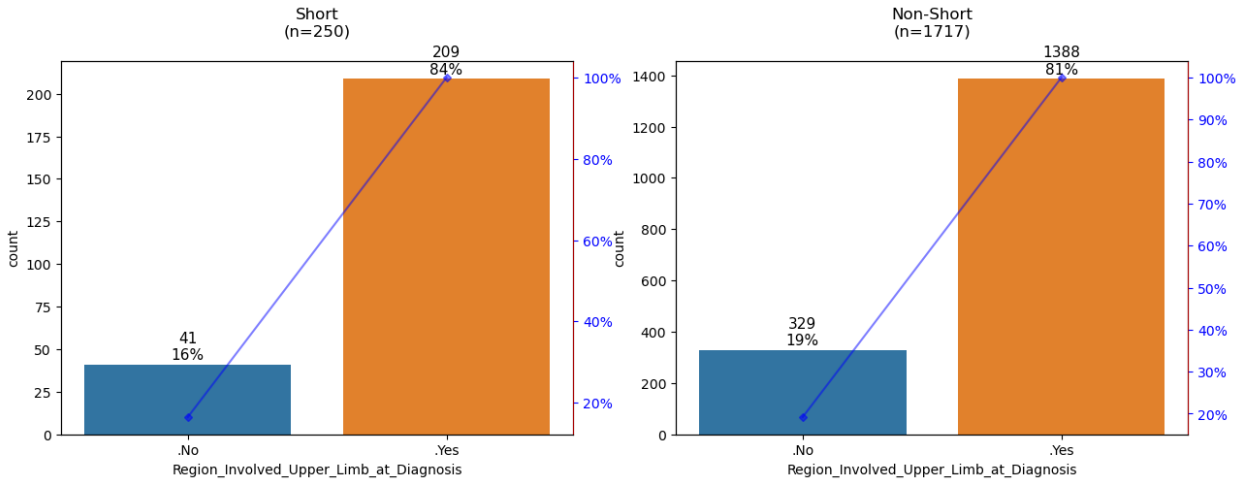
Column Qty\_Regions\_Involved\_at\_Diagnosis



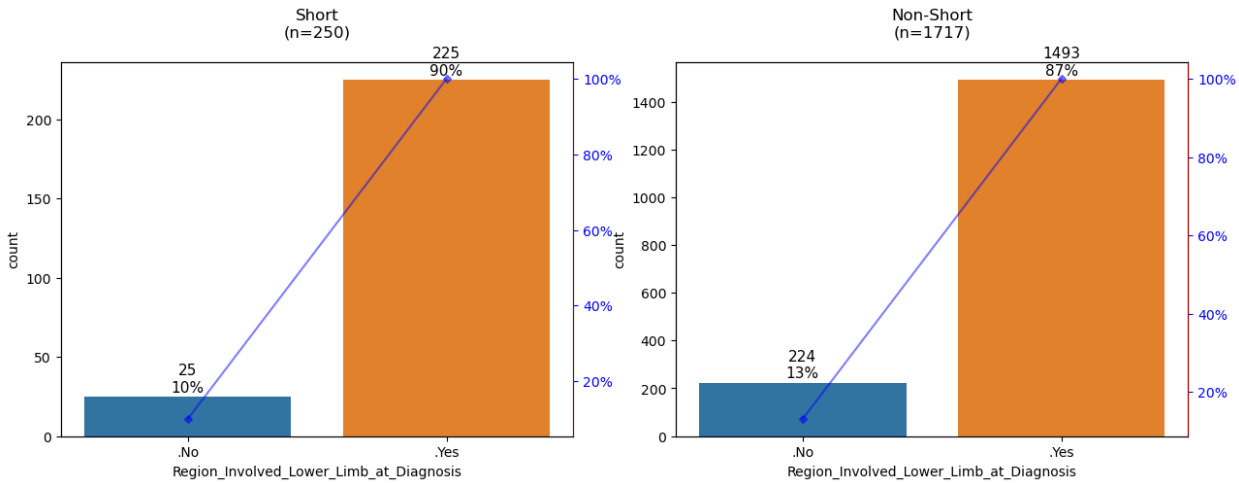
**Column Region\_Involved\_Bulbar\_at\_Diagnosis**



**Column Region\_Involved\_Upper\_Limb\_at\_Diagnosis**

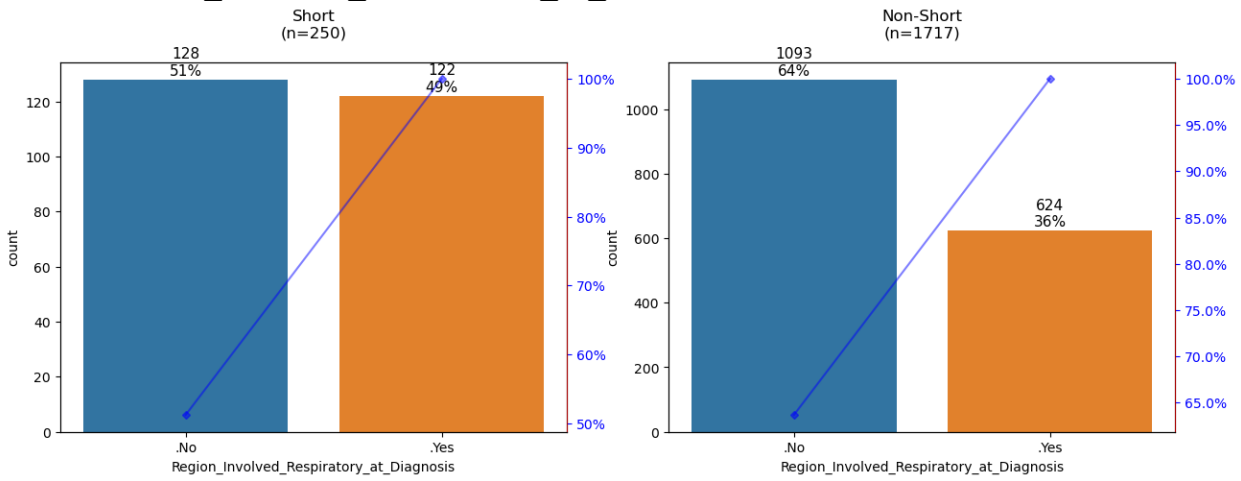


**Column Region\_Involved\_Lower\_Limb\_at\_Diagnosis**

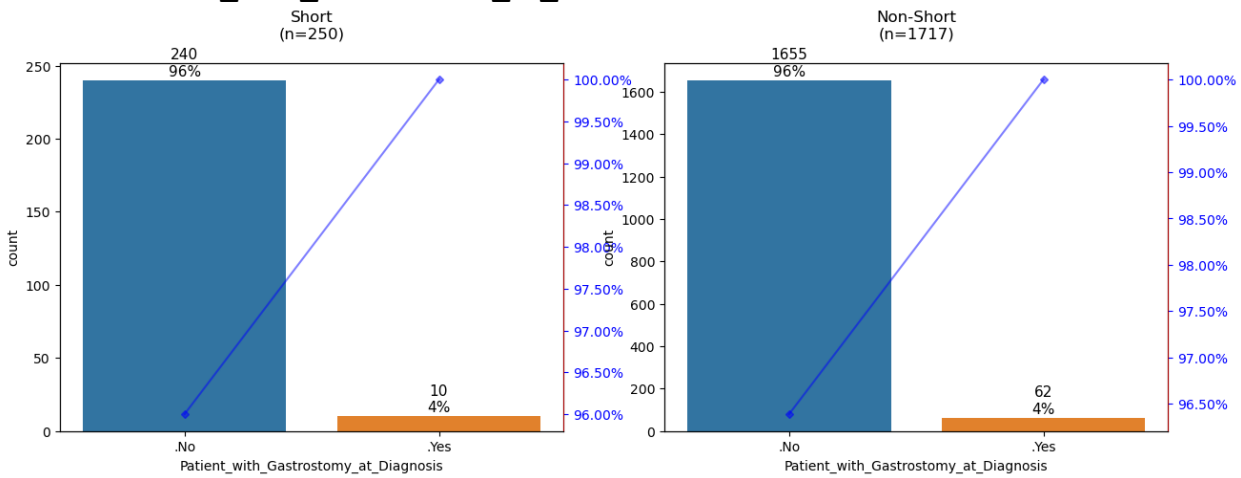




**Column Region\_Involved\_Respiratory\_at\_Diagnosis**



**Column Patient\_with\_Gastrostomy\_at\_Diagnosis**



# Compare the output variable distribution for the *Training* and *Validation* subsets used to train and validate the machine learning models

