In [18]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
plt.rc('font', size = 12)


health = pd.read_csv('Healthcare_data_updated.csv', header = 0)  #this is the updated file with the more useable Age_Category

#get basic info on this df
print(health.shape)
print(list(health.columns))

#One column has a comma in the middle of its name so I altered that col name (it was causing problems when I plotted)
health=health.rename(columns = {'Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx':'Comorb_Encntr_For_General_Exam_W_O_Complaint_Susp_Or_Reprtd_Dx'})
```

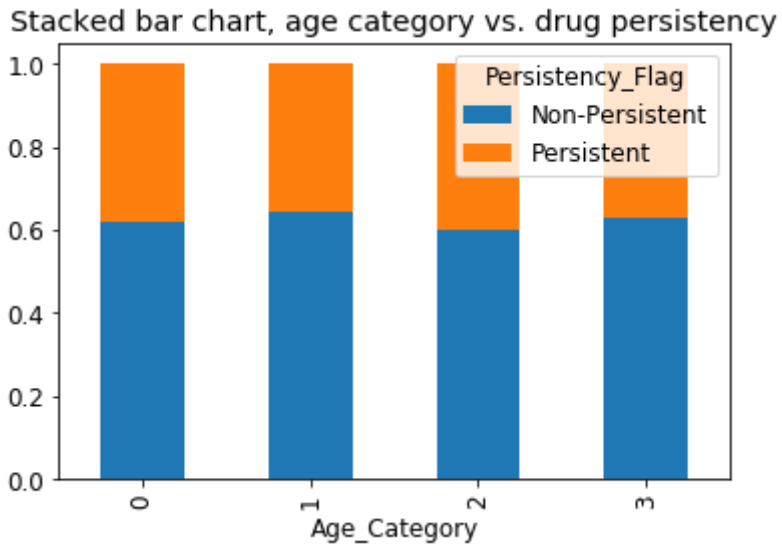
```
(3424, 71)
['Unnamed: 0', 'Ptid', 'Persistency_Flag', 'Gender', 'Race', 'Ethnicity', 'Region', 'Age_Category', 'Age
_Bucket', 'Ntm_Speciality', 'Ntm_Specialist_Flag', 'Ntm_Speciality_Bucket', 'Gluco_Record_Prior_Ntm', 'G
luco_Record_During_Rx', 'Dexa_Freq_During_Rx', 'Dexa_During_Rx', 'Frag_Frac_Prior_Ntm', 'Frag_Frac_Durin
g_Rx', 'Risk_Segment_Prior_Ntm', 'Tscore_Bucket_Prior_Ntm', 'Risk_Segment_During_Rx', 'Tscore_Bucket_Dur
ing_Rx', 'Change_T_Score', 'Change_Risk_Segment', 'Adherent_Flag', 'Idn_Indicator', 'Injectable_Experien
ce_During_Rx', 'Comorb_Encounter_For_Screening_For_Malignant_Neoplasms', 'Comorb_Encounter_For_Immunizat
ion', 'Comorb_Encntr_For_General_Exam_W_O_Complaint,_Susp_Or_Reprtd_Dx', 'Comorb_Vitamin_D_Deficiency',
'Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified', 'Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Sus
pected_Or_Reprtd_Dx', 'Comorb_Long_Term_Current_Drug_Therapy', 'Comorb_Dorsalgia', 'Comorb_Personal_Hist
ory_Of_Other_Diseases_And_Conditions', 'Comorb_Other_Disorders_Of_Bone_Density_And_Structure', 'Comorb_D
isorders_of_lipoprotein_metabolism_and_other_lipidemias', 'Comorb_Osteoporosis_without_current_pathologi
cal_fracture', 'Comorb_Personal_history_of_malignant_neoplasm', 'Comorb_Gastro_esophageal_reflux_disease
', 'Concom_Cholesterol_And_Triglyceride_Regulating_Preparations', 'Concom_Narcotics', 'Concom_Systemic_C
orticosteroids_Plain', 'Concom_Anti_Depressants_And_Mood_Stabilisers', 'Concom_Fluoroquinolones', 'Conco
m_Cephalosporins', 'Concom_Macrolides_And_Similar_Types', 'Concom_Broad_Spectrum_Penicillins', 'Concom_A
naesthetics_General', 'Concom_Viral_Vaccines', 'Risk_Type_1_Insulin_Dependent_Diabetes', 'Risk_Osteogene
sis_Imperfecta', 'Risk_Rheumatoid_Arthritis', 'Risk_Untreated_Chronic_Hyperthyroidism', 'Risk_Untreated_
Chronic_Hypogonadism', 'Risk_Untreated_Early_Menopause', 'Risk_Patient_Parent_Fractured_Their_Hip', 'Ris
k_Smoking_Tobacco', 'Risk_Chronic_Malnutrition_Or_Malabsorption', 'Risk_Chronic_Liver_Disease', 'Risk_Fa
mily_History_Of_Osteoporosis', 'Risk_Low_Calcium_Intake', 'Risk_Vitamin_D_Insufficiency', 'Risk_Poor_Hea
lth_Frailty', 'Risk_Excessive_Thinness', 'Risk_Hysterectomy_Oophorectomy', 'Risk_Estrogen_Deficiency', '
Risk_Immobilization', 'Risk_Recurring_Falls', 'Count_Of_Risks']
```

In [19]:
```
#the data is unbalanced between the two possible values for Persistency_Flag - this will be dealt with l
ater when applying the ML classification algorithm
health['Persistency_Flag'].value_counts()
```

Out[19]:
```
Non-Persistent    2135
Persistent        1289
Name: Persistency_Flag, dtype: int64
```

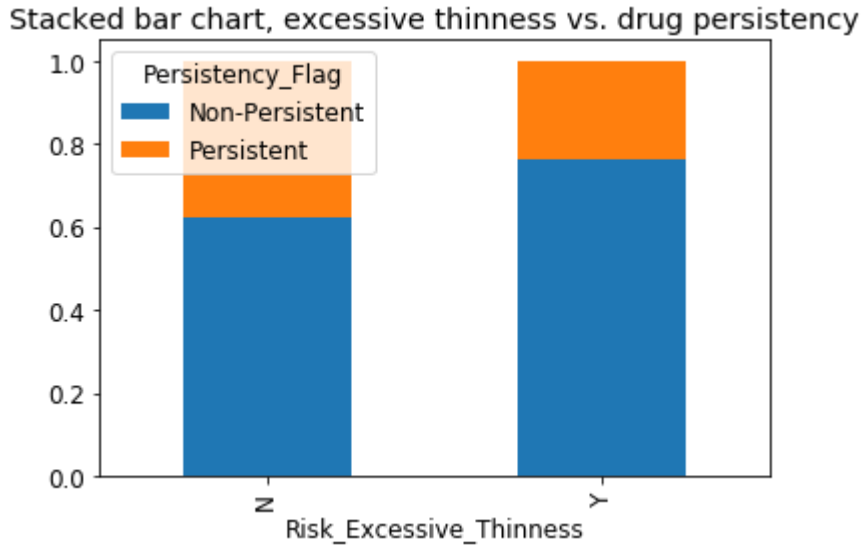
In [20]:
```
#I created bar charts to do a visual check for any possible relationship between a variable (column) and
the target drug persistency variable
#Even if the bar chart indicates a relationship, if the numbers are too skewed it might not be helpful,
so I also printed out value counts for that indep variable


table = pd.crosstab(health.Age_Category, health.Persistency_Flag)
table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
plt.title('Stacked bar chart, age category vs. drug persistency')
#age category does not seem a strong predictor for drug persistency
health['Age_Category'].value_counts()

#No evident information from this independent variable (it sheds little or no light on persistency based
on the bar chart)
```

Out[20]:
```
3    1438
2    1086
1     733
0     167
Name: Age_Category, dtype: int64
```



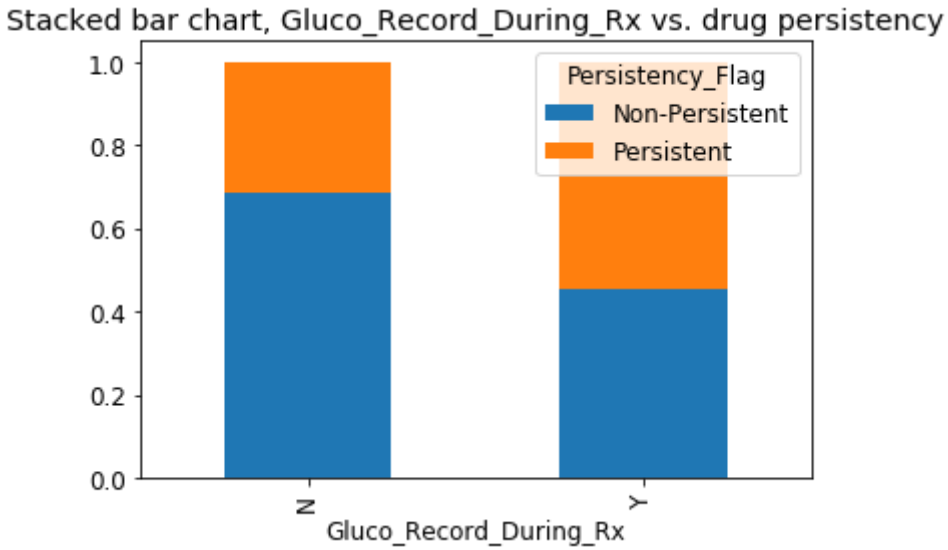
In [21]:
```
table = pd.crosstab(health.Risk_Excessive_Thinness, health.Persistency_Flag)
table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
plt.title('Stacked bar chart, excessive thinness vs. drug persistency')
health['Risk_Excessive_Thinness'].value_counts()
#excessive thinness seems a slightly better predictor but the numbers are skewed (so not so helpful)
```

Out[21]:
```
N    3357
Y      67
Name: Risk_Excessive_Thinness, dtype: int64
```



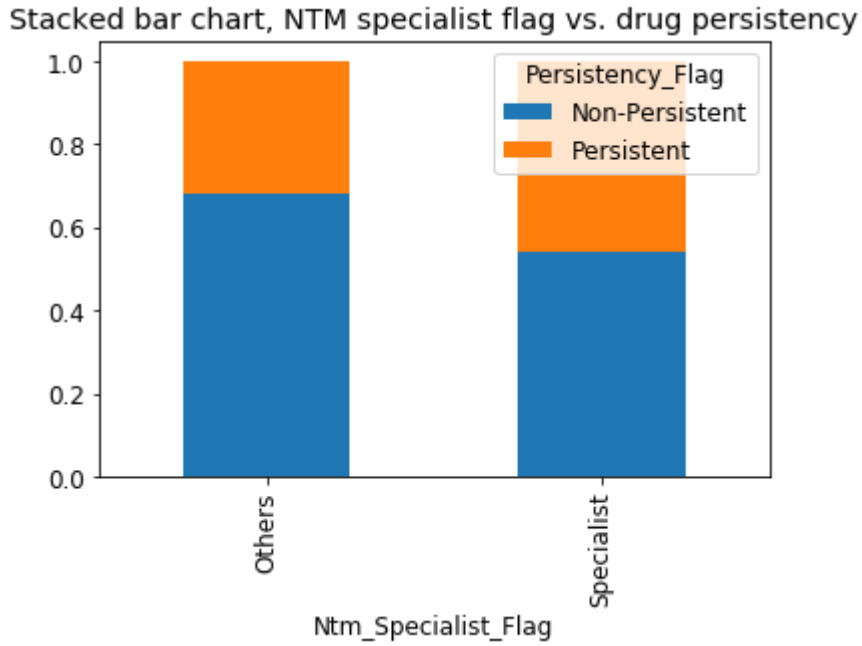
In [15]:
```
table = pd.crosstab(health.Gluco_Record_During_Rx, health.Persistency_Flag)
table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
plt.title('Stacked bar chart, Gluco_Record_During_Rx vs. drug persistency')
health['Gluco_Record_During_Rx'].value_counts()
```

Out[15]:
```
N    2522
Y     902
Name: Gluco_Record_During_Rx, dtype: int64
```



In [11]:
```
table = pd.crosstab(health.Ntm_Specialist_Flag, health.Persistency_Flag)
table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
plt.title('Stacked bar chart, NTM specialist flag vs. drug persistency')
health['Ntm_Specialist_Flag'].value_counts()
```

Out[11]:
```
Others        2013
Specialist    1411
Name: Ntm_Specialist_Flag, dtype: int64
```

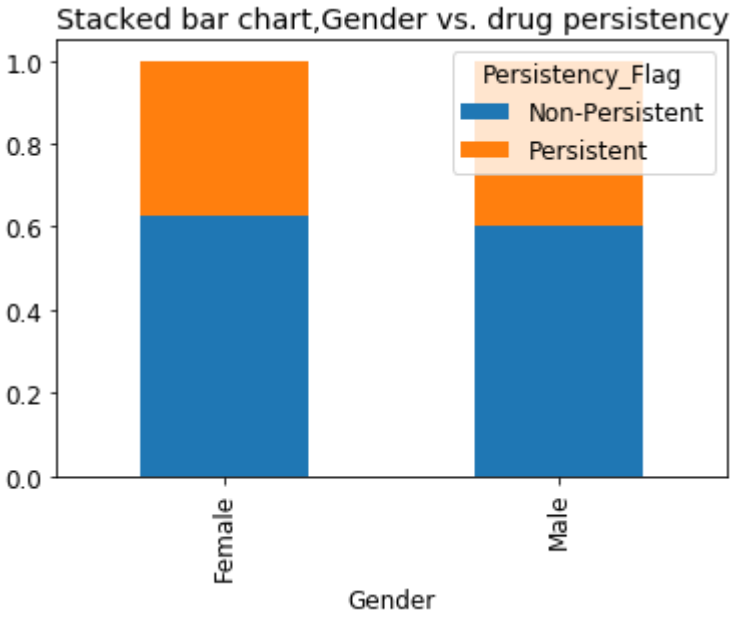


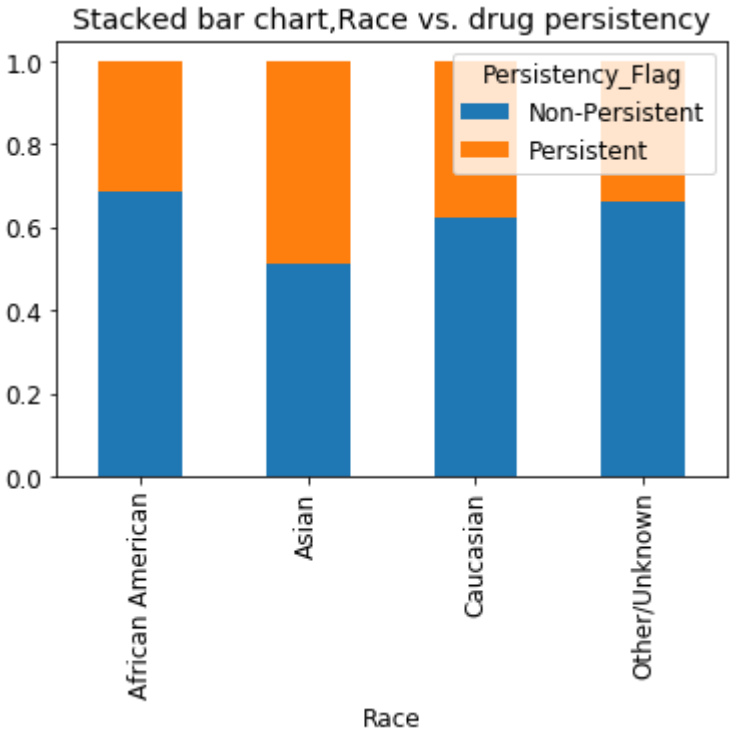
In [12]:
```
#Now to check all the variables:
#first create list for demographic variables
#then create lists for clinical variables

#including age_category and not the other, age_bucket
list1 = health.columns[3:8]
list2 = health.columns[9:12]
demographic = list1.to_list() + list2.to_list() #the produced list1, list2 were series so I converted to
list format



#I similarly created three clinical lists
list3 = health.columns[12:26]
clinical1 = list3.to_list()
list4 = health.columns[26:50]
clinical2 = list4.to_list()
list5 = health.columns[50:71]
clinical3 = list5.to_list()
```

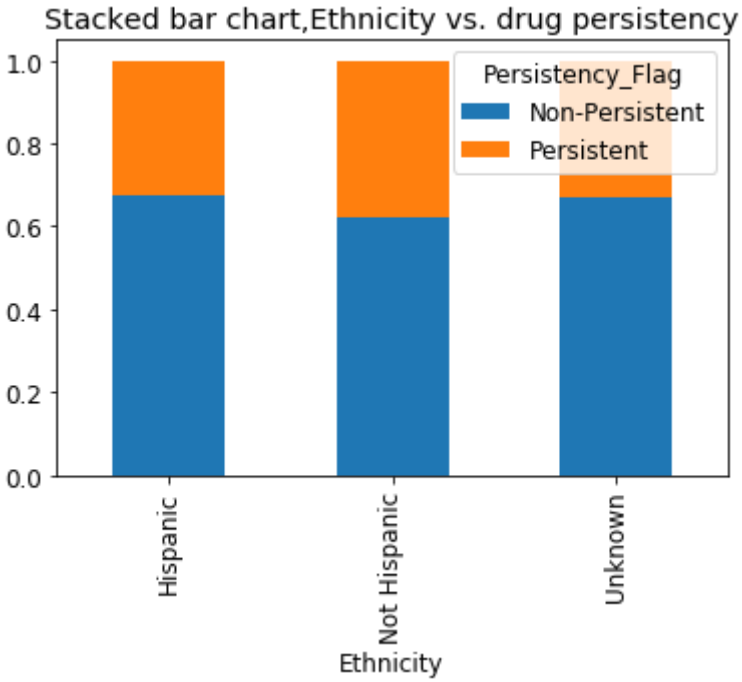
In [22]:
```
#using a loop, look at all the independent demographic variables against the dependent variable drug per
sistency
for item in demographic:
    table = pd.crosstab(health[item], health.Persistency_Flag)
    table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
    plt.title('Stacked bar chart,'+ item + ' vs. drug persistency')
    print(health[item].value_counts())
    print('---')
```

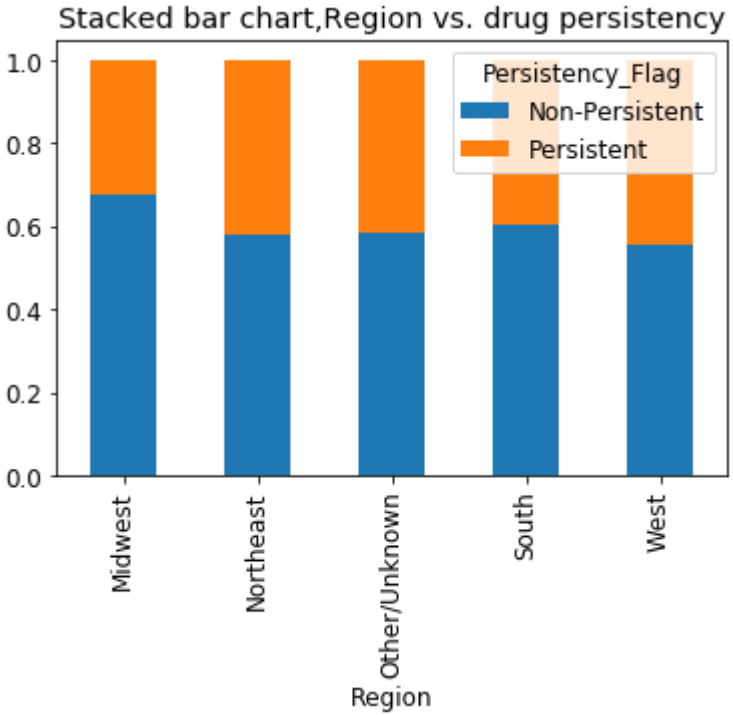
```
Female    3230
Male       194
Name: Gender, dtype: int64
---
Caucasian           3148
Other/Unknown         97
African American      95
Asian                 84
Name: Race, dtype: int64
---
Not Hispanic    3235
Hispanic          98
Unknown           91
Name: Ethnicity, dtype: int64
---
Midwest          1383
South            1247
West              502
Northeast         232
Other/Unknown      60
Name: Region, dtype: int64
---
3    1438
2    1086
1     733
0     167
Name: Age_Category, dtype: int64
---
GENERAL PRACTITIONER                                1535
RHEUMATOLOGY                                         604
ENDOCRINOLOGY                                        458
Unknown                                              310
ONCOLOGY                                             225
OBSTETRICS AND GYNECOLOGY                             90
UROLOGY                                               33
ORTHOPEDIC SURGERY                                    30
CARDIOLOGY                                            22
PATHOLOGY                                             16
HEMATOLOGY & ONCOLOGY                                 14
OTOLARYNGOLOGY                                        14
```

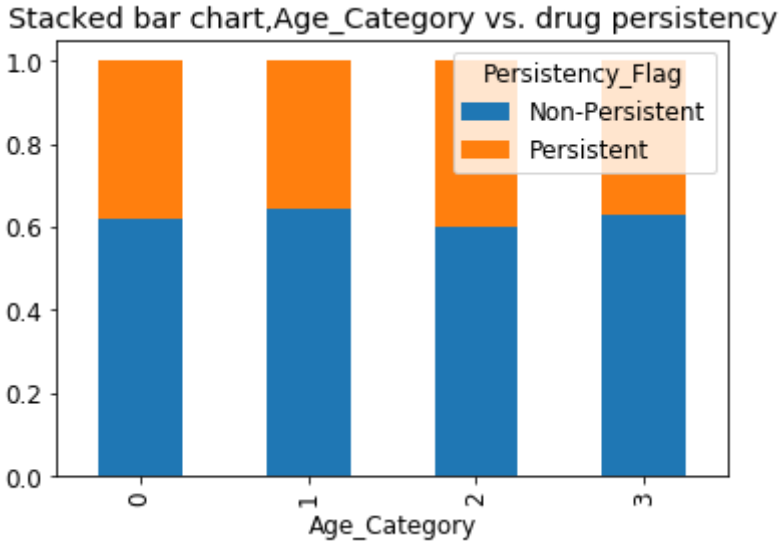
```
PEDIATRICS                                                          13
PHYSICAL MEDICINE AND REHABILITATION                                11
PULMONARY MEDICINE                                                   8
SURGERY AND SURGICAL SPECIALTIES                                     8
PSYCHIATRY AND NEUROLOGY                                             4
ORTHOPEDICS                                                          3
NEPHROLOGY                                                           3
GERIATRIC MEDICINE                                                   2
VASCULAR SURGERY                                                     2
TRANSPLANT SURGERY                                                   2
GASTROENTEROLOGY                                                     2
HOSPICE AND PALLIATIVE MEDICINE                                      2
PLASTIC SURGERY                                                      2
CLINICAL NURSE SPECIALIST                                            1
PAIN MEDICINE                                                        1
EMERGENCY MEDICINE                                                   1
OBSTETRICS & OBSTETRICS & GYNECOLOGY & OBSTETRICS & GYNECOLOGY       1
HOSPITAL MEDICINE                                                    1
OPHTHALMOLOGY                                                        1
PODIATRY                                                             1
NEUROLOGY                                                            1
NUCLEAR MEDICINE                                                     1
RADIOLOGY                                                            1
OCCUPATIONAL MEDICINE                                                1
Name: Ntm_Speciality, dtype: int64
---
Others        2013
Specialist    1411
Name: Ntm_Specialist_Flag, dtype: int64
---
OB/GYN/Others/PCP/Unknown    2104
Endo/Onc/Uro                  716
Rheum                         604
Name: Ntm_Speciality_Bucket, dtype: int64
---
```

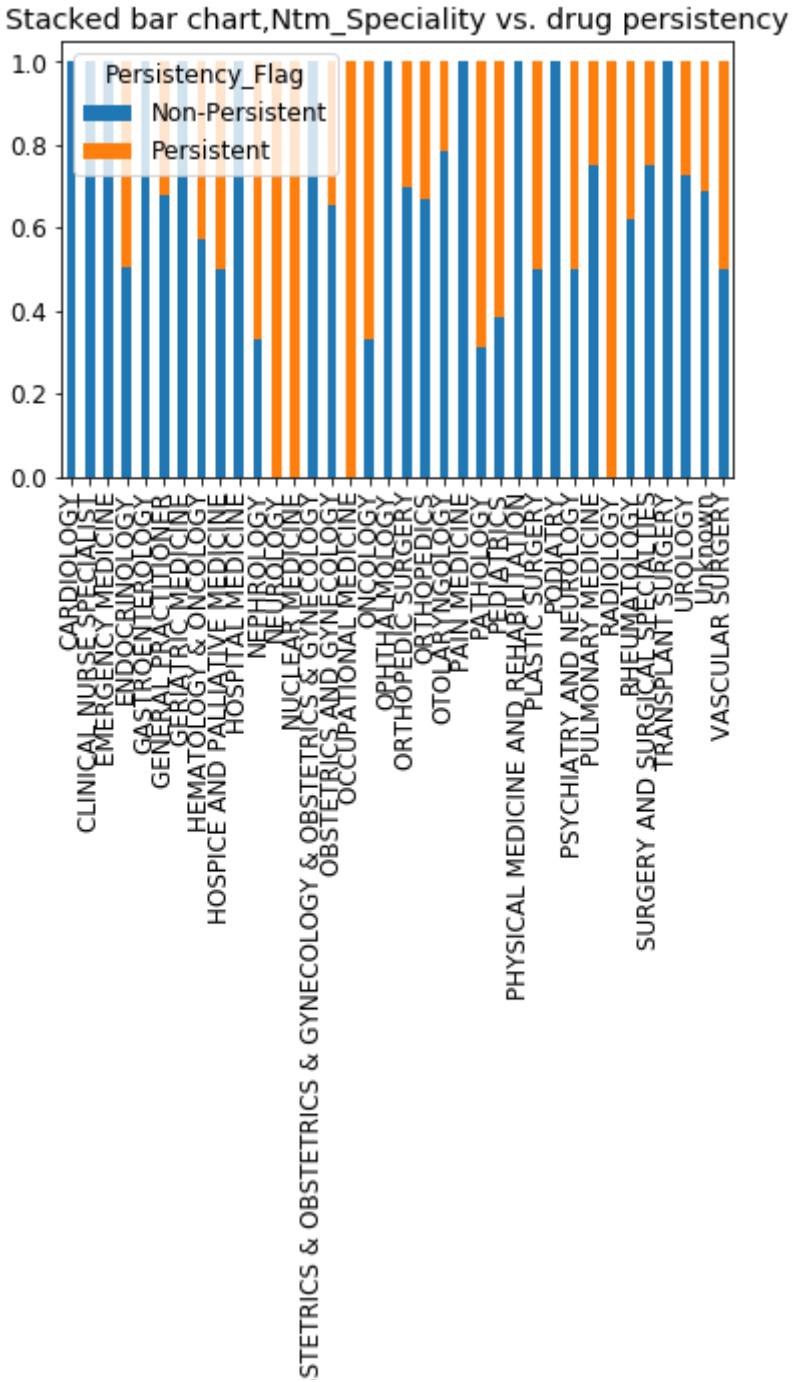

Stacked bar chart,Gender vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Female
Male
Gender

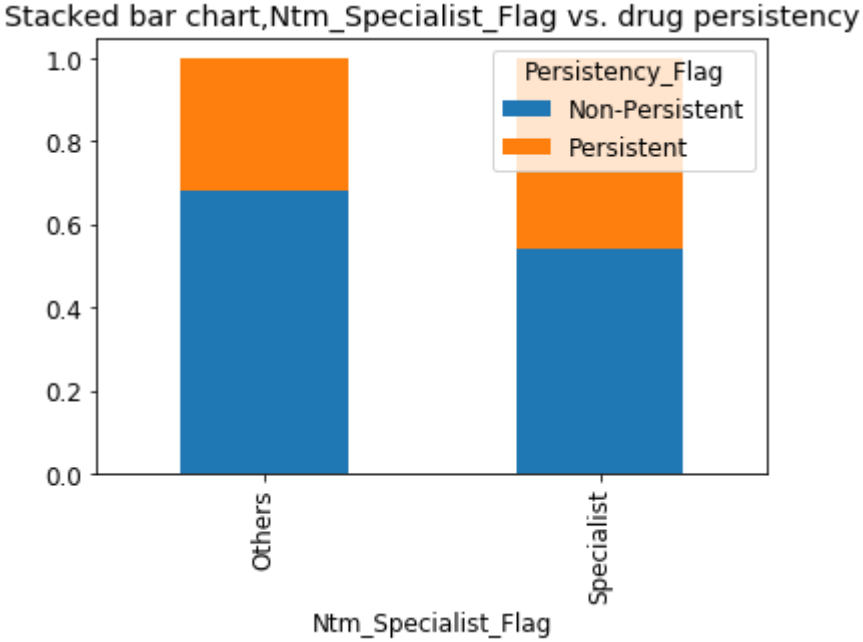

Stacked bar chart,Race vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
African American
Asian
Caucasian
Other/Unknown
Race

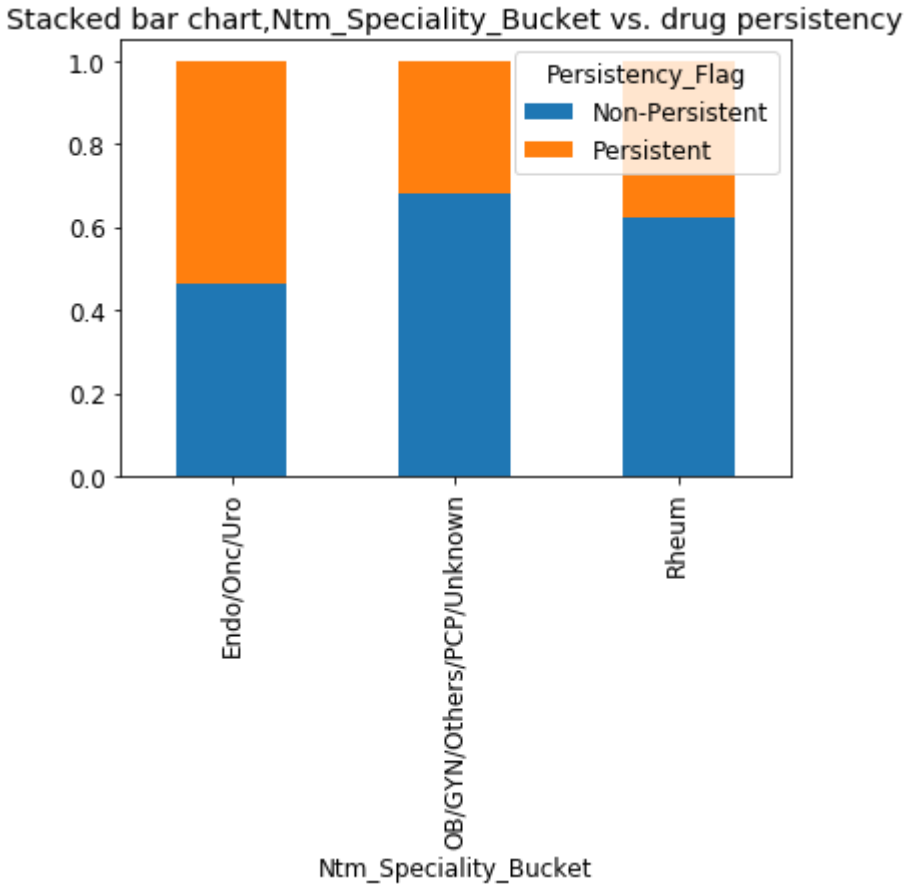

Stacked bar chart,Ethnicity vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Hispanic
Not Hispanic
Unknown
Ethnicity


Stacked bar chart,Region vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Midwest
Northeast
Other/Unknown
South
West
Region


Stacked bar chart,Age_Category vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
Age_Category


Stacked bar chart,Ntm_Speciality vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
CARDIOLOGY
CLINICAL NURSE SPECIALIST
EMERGENCY MEDICINE
ENDOCRINOLOGY
GASTROENTEROLOGY
GENERAL PRACTITIONER
GERIATRIC MEDICINE
HEMATOLOGY & ONCOLOGY
HOSPICE AND PALLIATIVE MEDICINE
NEPHROLOGY
NEUROLOGY
NUCLEAR MEDICINE
OBSTETRICS & GYNECOLOGY
OCCUPATIONAL MEDICINE
ONCOLOGY
OPHTHALMOLOGY
ORTHOPEDIC SURGERY
ORTHOPEDICS
OTOLARYNGOLOGY
PAIN MEDICINE
PATHOLOGY
PEDIATRICS
PHYSICAL MEDICINE AND REHABILITATION
PLASTIC SURGERY
PODIATRY
PSYCHIATRY AND NEUROLOGY
PULMONARY MEDICINE
RADIOLOGY
RHEUMATOLOGY
SURGERY AND SURGICAL SPECIALTIES
TRANSPLANT SURGERY
UROLOGY
Unknown
VASCULAR SURGERY


Stacked bar chart,Ntm_Specialist_Flag vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Others
Specialist
Ntm_Specialist_Flag

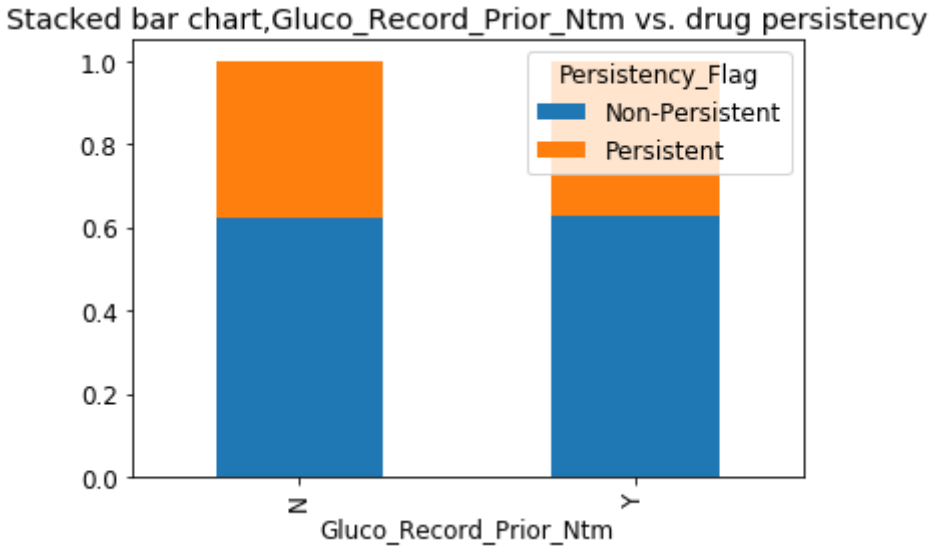

Stacked bar chart,Ntm_Speciality_Bucket vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Endo/Onc/Uro
OB/GYN/Others/PCP/Unknown
Rheum
Ntm_Speciality_Bucket

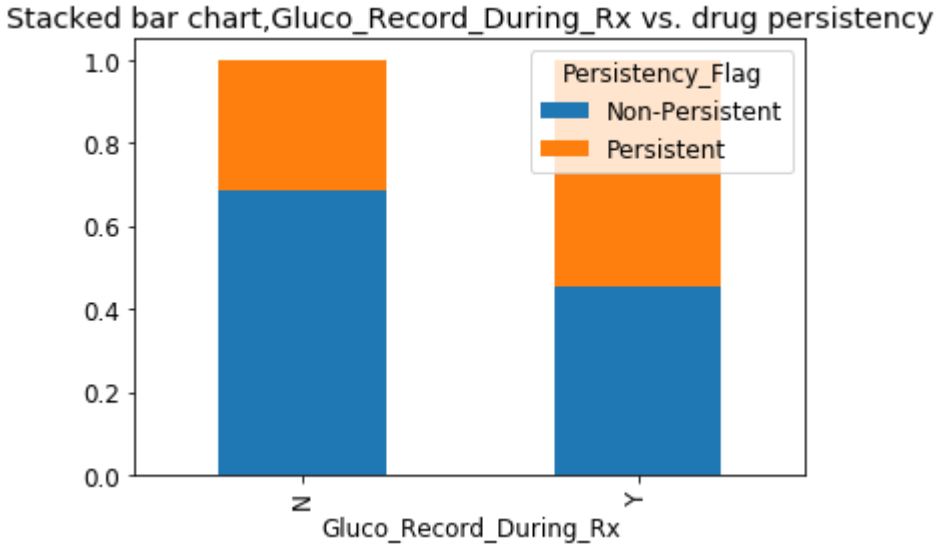
In [23]:
```
#look at the first list of clinical variables against the dependent variable drug persistency

for item in clinical1:
    table = pd.crosstab(health[item], health.Persistency_Flag)
    table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
    plt.title('Stacked bar chart,'+ item + ' vs. drug persistency')
    print(health[item].value_counts())
    print('---')
```

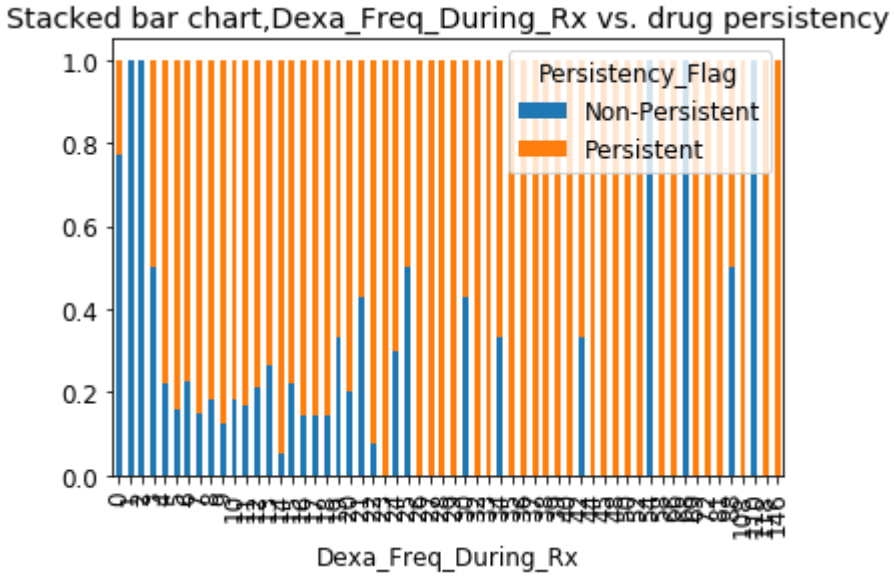
```
N    2619
Y     805
Name: Gluco_Record_Prior_Ntm, dtype: int64
---
N    2522
Y     902
Name: Gluco_Record_During_Rx, dtype: int64
---
0      2488
5       114
6       107
7        93
8        71
4        68
10       55
12       52
3        46
14       38
9        32
11       30
1        24
2        24
13       19
20       15
16       14
18       14
22       13
26       10
24       10
15        9
17        7
30        7
21        7
28        7
36        5
19        3
32        3
34        3
42        3
25        2
```

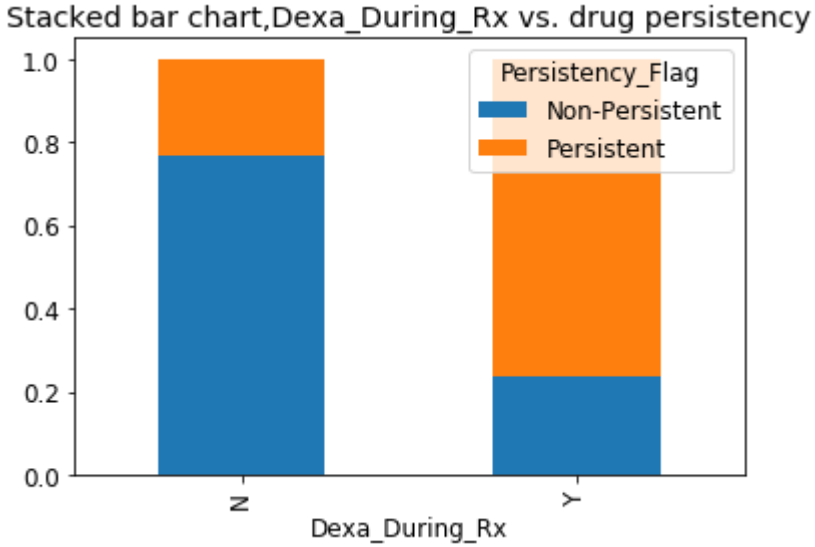
```
39        2
58        2
52        2
48        2
88        2
38        1
81        1
40        1
146       1
45        1
37        1
35        1
33        1
29        1
27        1
23        1
118       1
44        1
110       1
108       1
72        1
68        1
69        1
54        1
50        1
66        1
Name: Dexa_Freq_During_Rx, dtype: int64
---
N    2488
Y     936
Name: Dexa_During_Rx, dtype: int64
---
N    2872
Y     552
Name: Frag_Frac_Prior_Ntm, dtype: int64
---
N    3007
Y     417
Name: Frag_Frac_During_Rx, dtype: int64
---
VLR_LR    1931
```

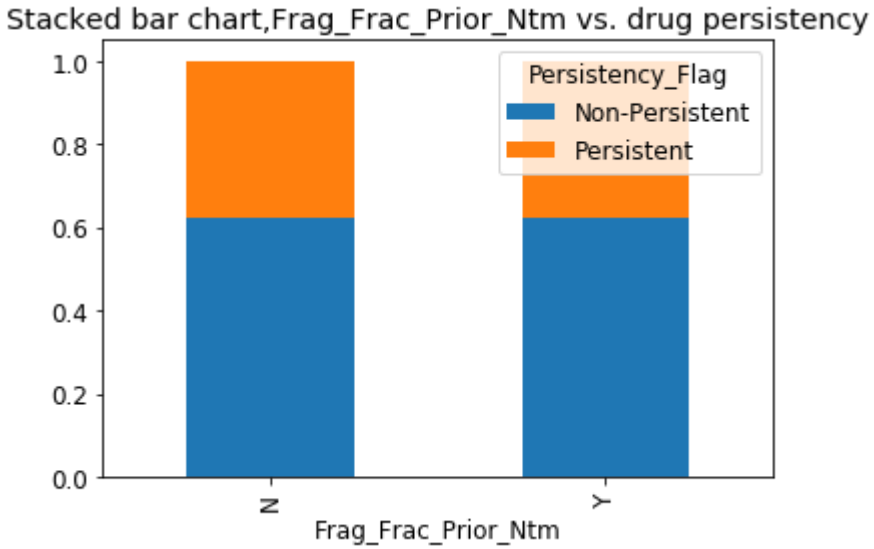
```
HR_VHR    1493
Name: Risk_Segment_Prior_Ntm, dtype: int64
---
>-2.5     1951
<=-2.5    1473
Name: Tscore_Bucket_Prior_Ntm, dtype: int64
---
Unknown    1497
HR_VHR      965
VLR_LR      962
Name: Risk_Segment_During_Rx, dtype: int64
---
Unknown    1497
<=-2.5     1017
>-2.5       910
Name: Tscore_Bucket_During_Rx, dtype: int64
---
No change    1660
Unknown      1497
Worsened      173
Improved       94
Name: Change_T_Score, dtype: int64
---
Unknown      2229
No change    1052
Worsened      121
Improved       22
Name: Change_Risk_Segment, dtype: int64
---
Adherent        3251
Non-Adherent     173
Name: Adherent_Flag, dtype: int64
---
Y    2557
N     867
Name: Idn_Indicator, dtype: int64
---
```

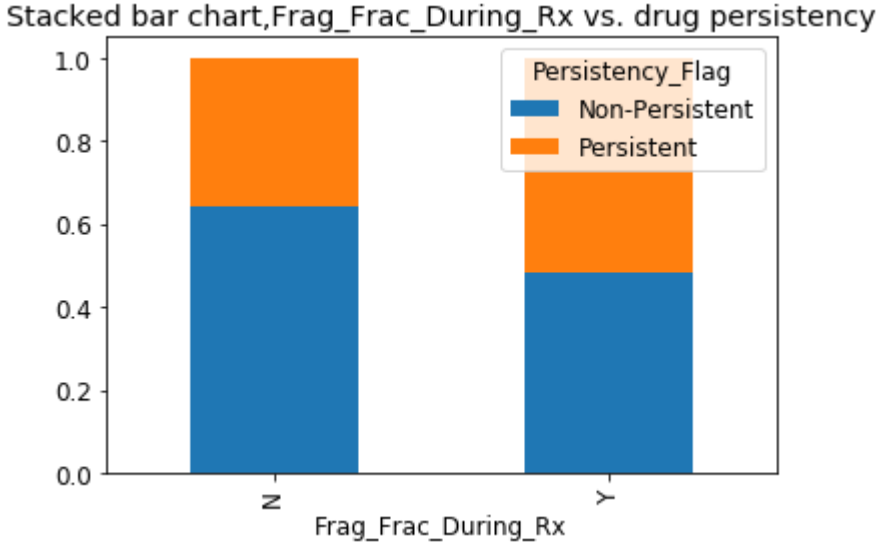

Stacked bar chart,Gluco_Record_Prior_Ntm vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Gluco_Record_Prior_Ntm

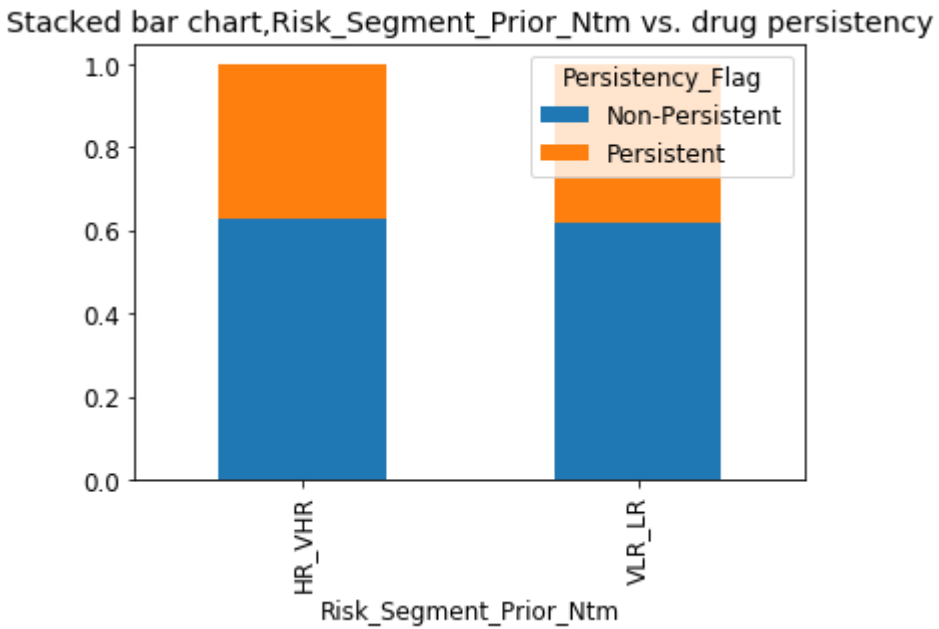

Stacked bar chart,Gluco_Record_During_Rx vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Gluco_Record_During_Rx

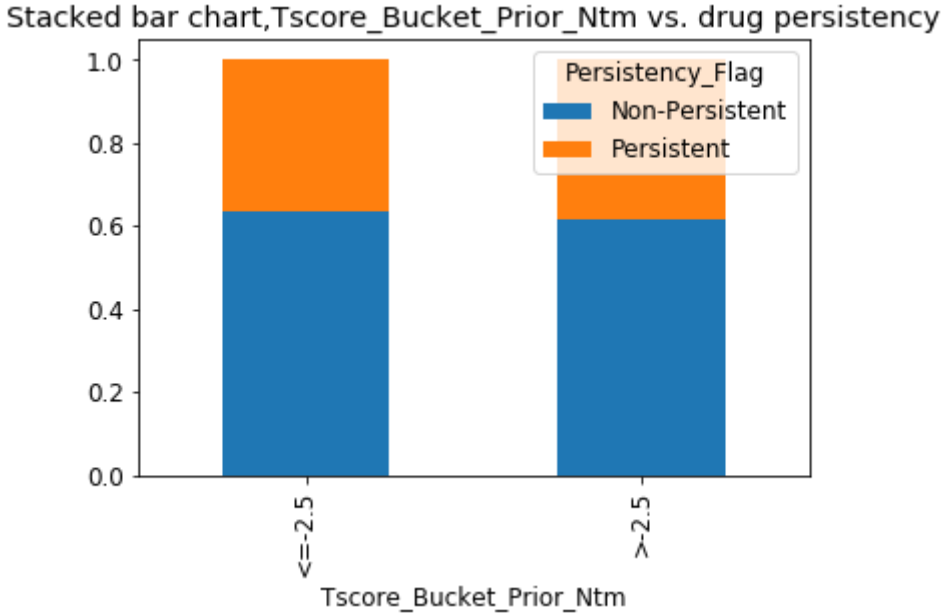

Stacked bar chart,Dexa_Freq_During_Rx vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Dexa_Freq_During_Rx

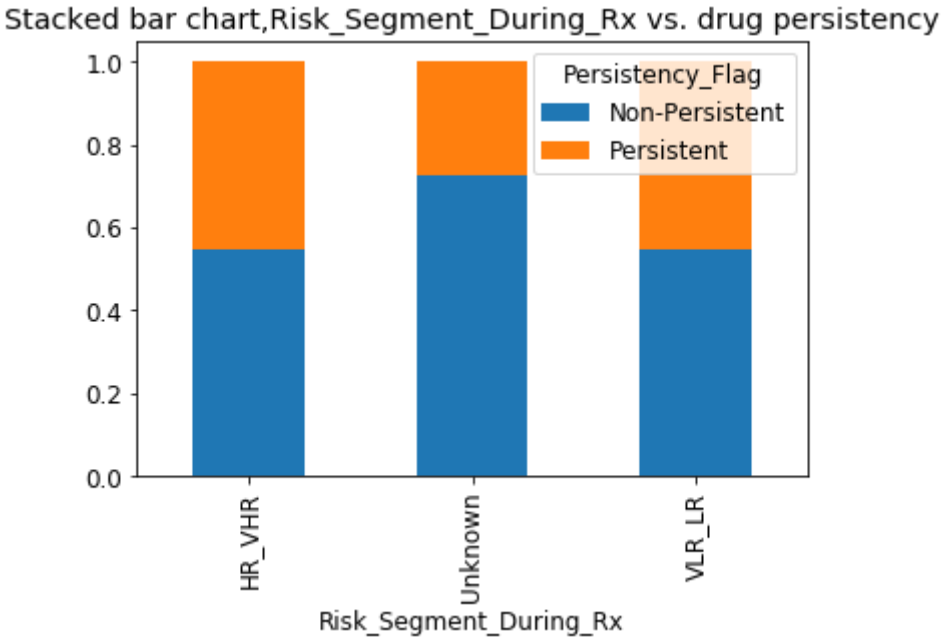

Stacked bar chart,Dexa_During_Rx vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Dexa_During_Rx

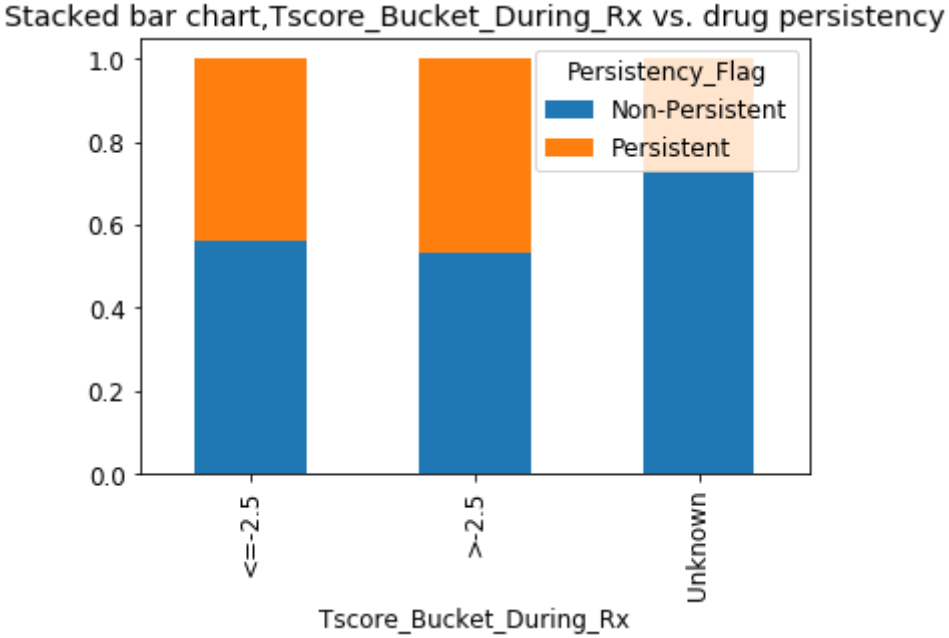

Stacked bar chart,Frag_Frac_Prior_Ntm vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Frag_Frac_Prior_Ntm

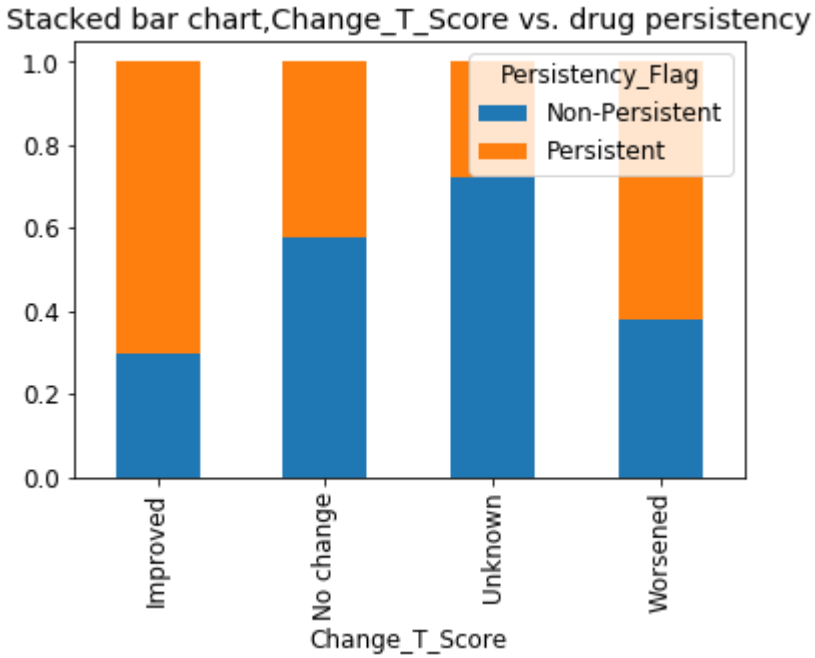

Stacked bar chart,Frag_Frac_During_Rx vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Frag_Frac_During_Rx

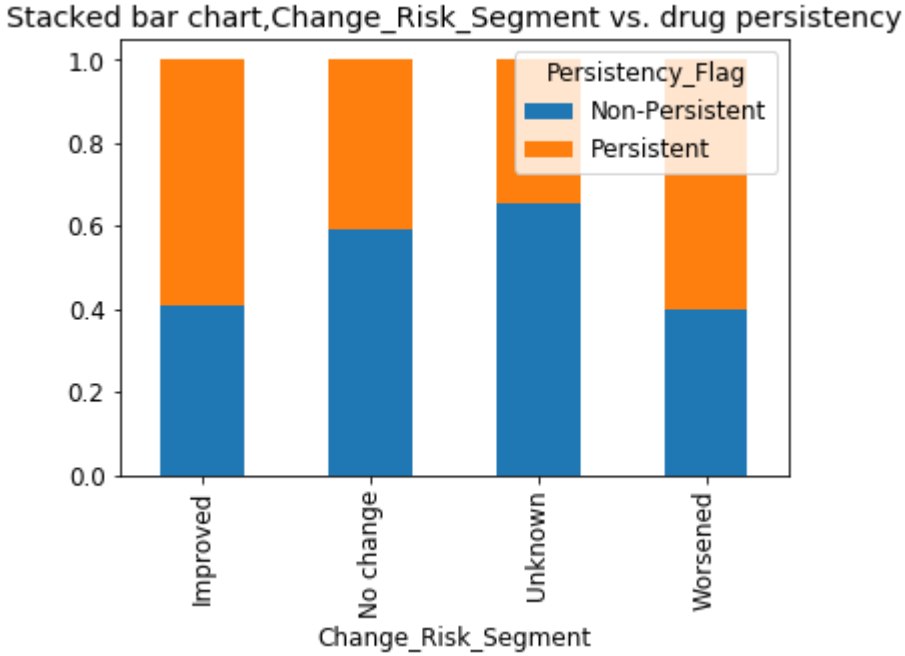

Stacked bar chart,Risk_Segment_Prior_Ntm vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
HR_VHR
VLR_LR
Risk_Segment_Prior_Ntm

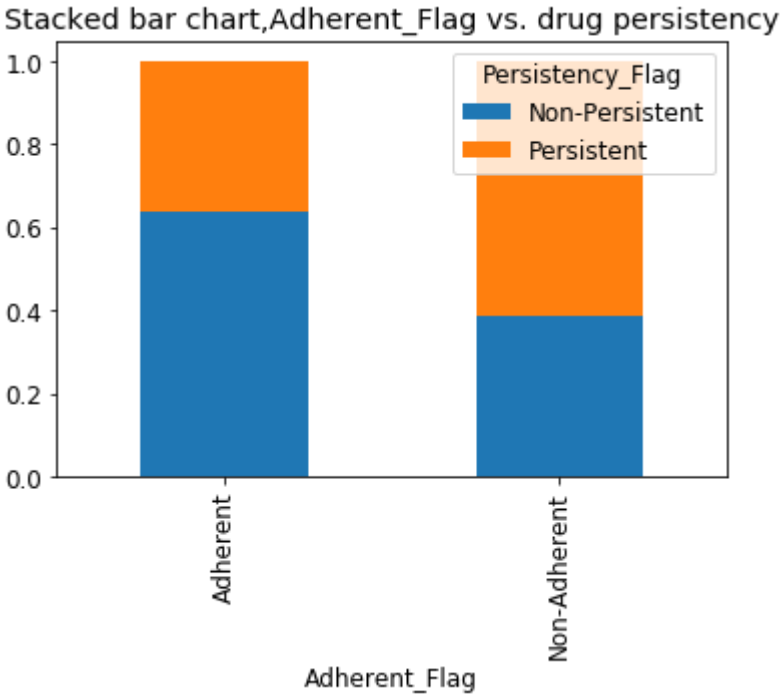

Stacked bar chart,Tscore_Bucket_Prior_Ntm vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
<=-2.5
>-2.5
Tscore_Bucket_Prior_Ntm

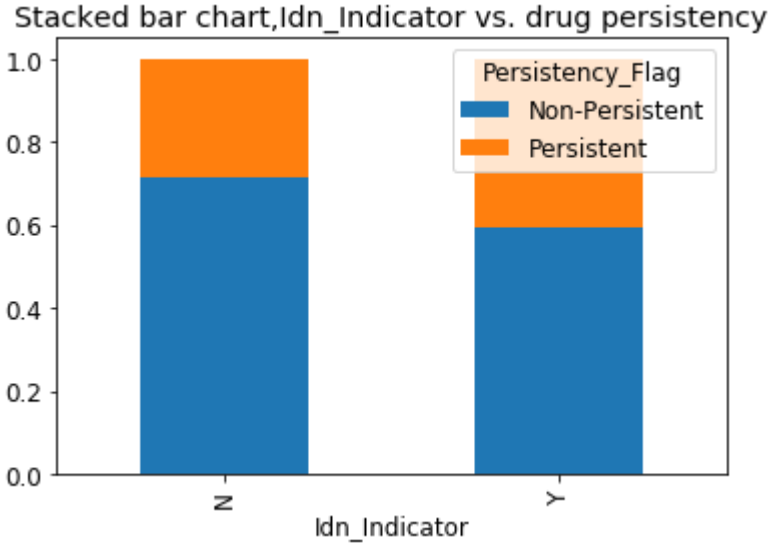

Stacked bar chart,Risk_Segment_During_Rx vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
HR_VHR
Unknown
VLR_LR
Risk_Segment_During_Rx


Stacked bar chart,Tscore_Bucket_During_Rx vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
<=-2.5
>-2.5
Unknown
Tscore_Bucket_During_Rx


Stacked bar chart,Change_T_Score vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Improved
No change
Unknown
Worsened
Change_T_Score


Stacked bar chart,Change_Risk_Segment vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Improved
No change
Unknown
Worsened
Change_Risk_Segment


Stacked bar chart,Adherent_Flag vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
Adherent
Non-Adherent
Adherent_Flag

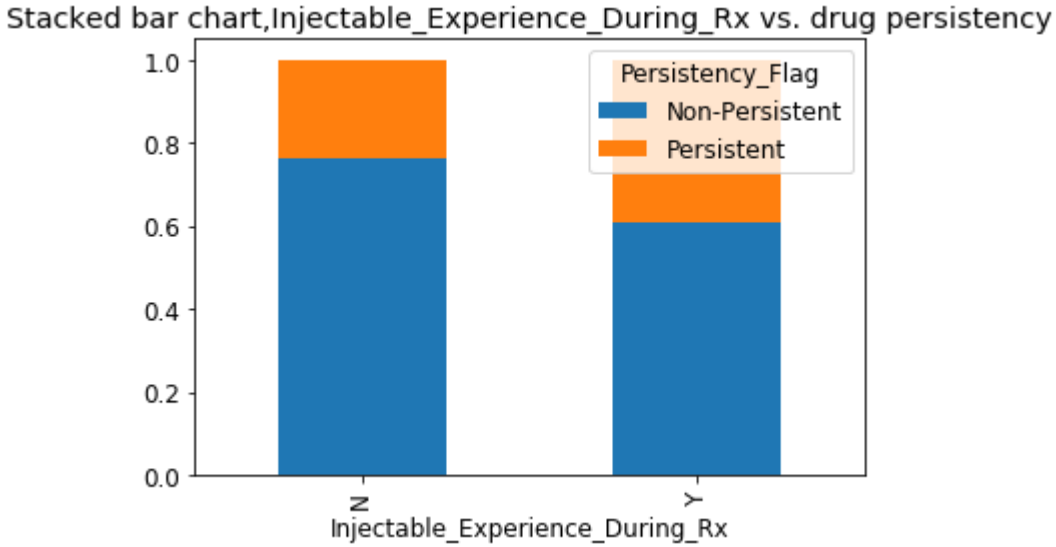

Stacked bar chart,Idn_Indicator vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Idn_Indicator

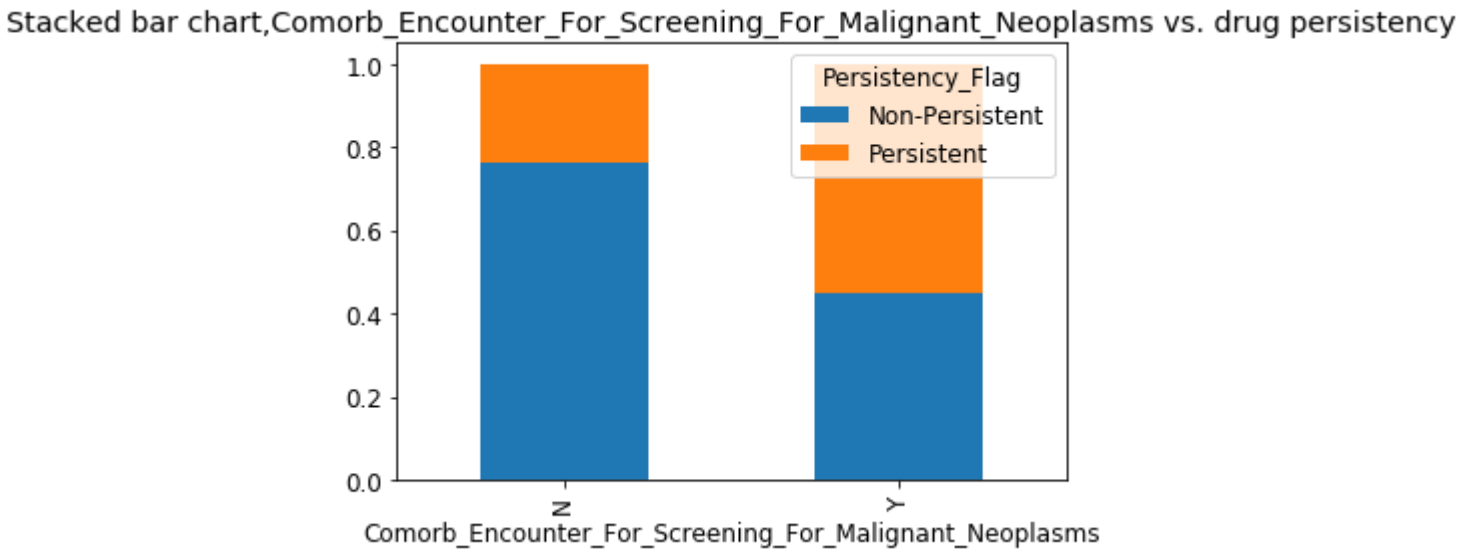
In [24]:
```
#look at indep vars in second clinical group in same manner

for item in clinical2:
    table = pd.crosstab(health[item], health.Persistency_Flag)
    table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
    plt.title('Stacked bar chart,'+ item + ' vs. drug persistency')
    print(health[item].value_counts())
    print('---')
```

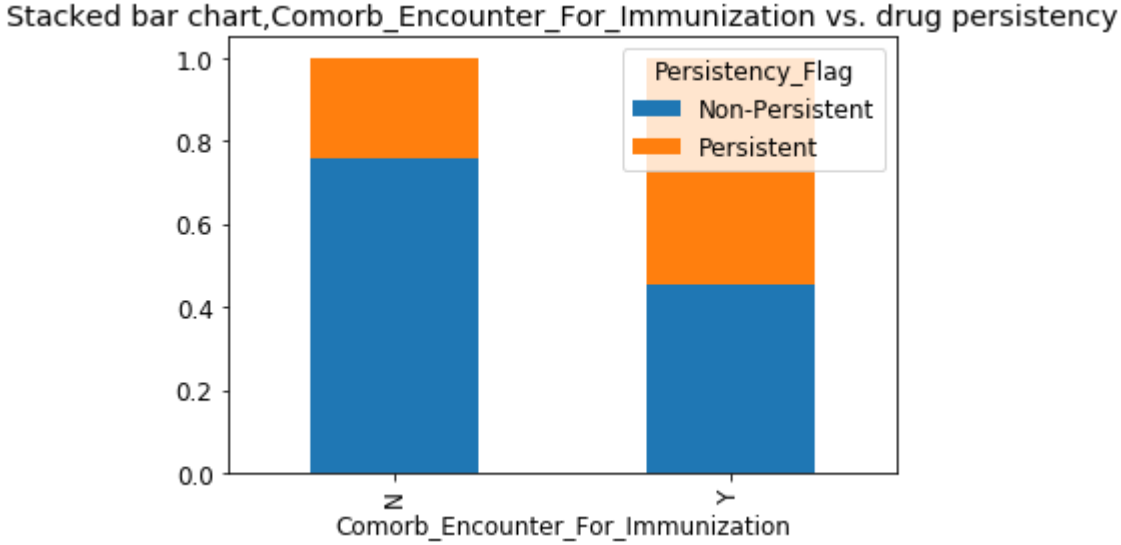
```
Y    3056
N     368
Name: Injectable_Experience_During_Rx, dtype: int64
---
N    1891
Y    1533
Name: Comorb_Encounter_For_Screening_For_Malignant_Neoplasms, dtype: int64
---
N    1911
Y    1513
Name: Comorb_Encounter_For_Immunization, dtype: int64
---
N    2072
Y    1352
Name: Comorb_Encntr_For_General_Exam_W_O_Complaint_Susp_Or_Reprtd_Dx, dtype: int64
---
N    2331
Y    1093
Name: Comorb_Vitamin_D_Deficiency, dtype: int64
---
N    2425
Y     999
Name: Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified, dtype: int64
---
N    2633
Y     791
Name: Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx, dtype: int64
---
N    2607
Y     817
Name: Comorb_Long_Term_Current_Drug_Therapy, dtype: int64
---
N    2645
Y     779
Name: Comorb_Dorsalgia, dtype: int64
---
N    2747
Y     677
Name: Comorb_Personal_History_Of_Other_Diseases_And_Conditions, dtype: int64
---
```

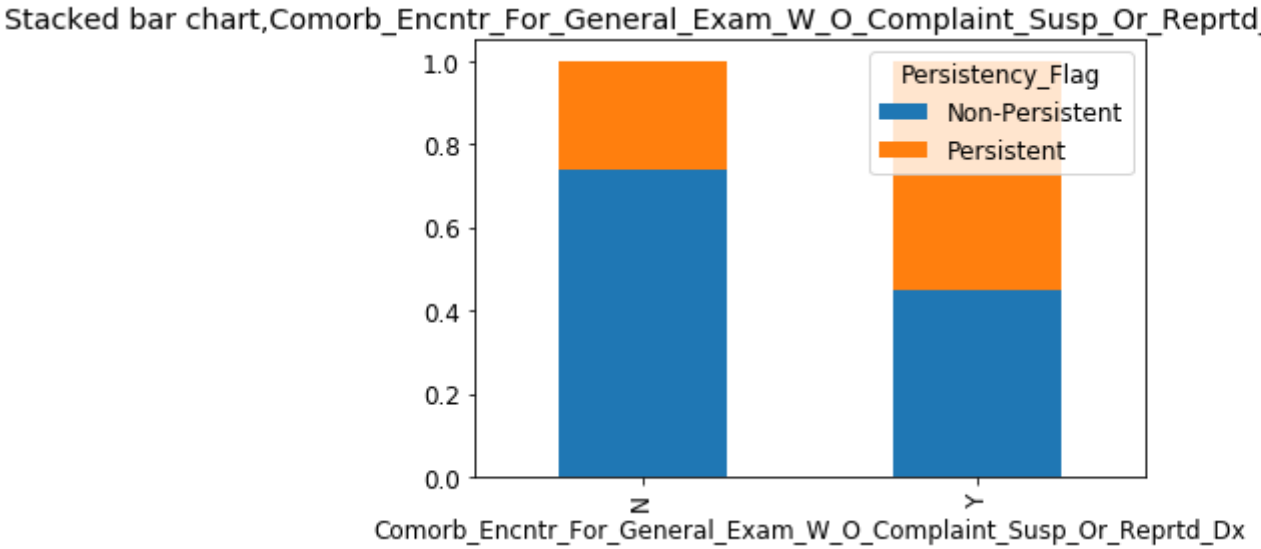
```
N    2906
Y     518
Name: Comorb_Other_Disorders_Of_Bone_Density_And_Structure, dtype: int64
---
Y    1765
N    1659
Name: Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias, dtype: int64
---
N    2507
Y     917
Name: Comorb_Osteoporosis_without_current_pathological_fracture, dtype: int64
---
N    2775
Y     649
Name: Comorb_Personal_history_of_malignant_neoplasm, dtype: int64
---
N    2794
Y     630
Name: Comorb_Gastro_esophageal_reflux_disease, dtype: int64
---
N    2242
Y    1182
Name: Concom_Cholesterol_And_Triglyceride_Regulating_Preparations, dtype: int64
---
N    2191
Y    1233
Name: Concom_Narcotics, dtype: int64
---
N    2451
Y     973
Name: Concom_Systemic_Corticosteroids_Plain, dtype: int64
---
N    2465
Y     959
Name: Concom_Anti_Depressants_And_Mood_Stabilisers, dtype: int64
---
N    2787
Y     637
Name: Concom_Fluoroquinolones, dtype: int64
---
N    2821
```

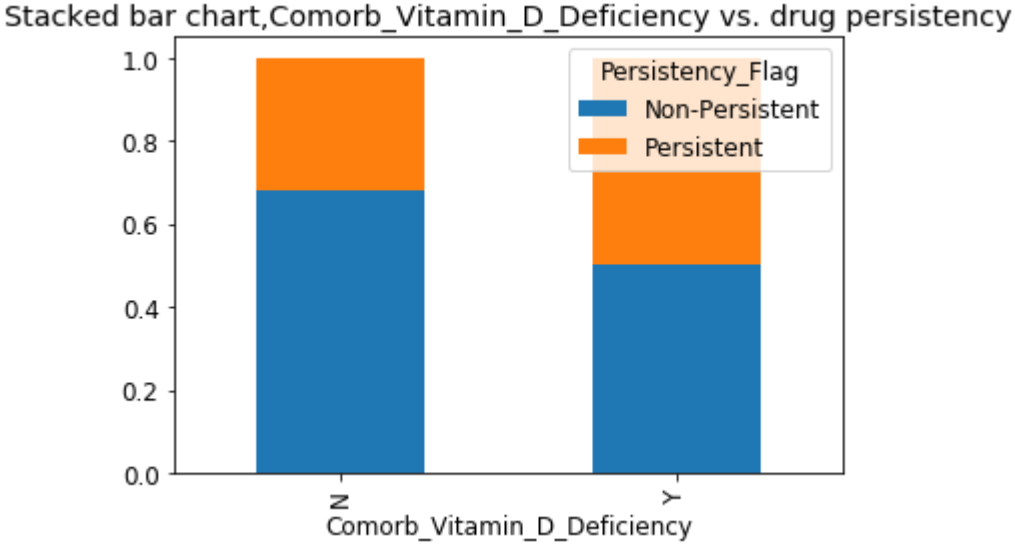
```
Y     603
Name: Concom_Cephalosporins, dtype: int64
---
N    2853
Y     571
Name: Concom_Macrolides_And_Similar_Types, dtype: int64
---
N    2985
Y     439
Name: Concom_Broad_Spectrum_Penicillins, dtype: int64
---
N    2927
Y     497
Name: Concom_Anaesthetics_General, dtype: int64
---
```



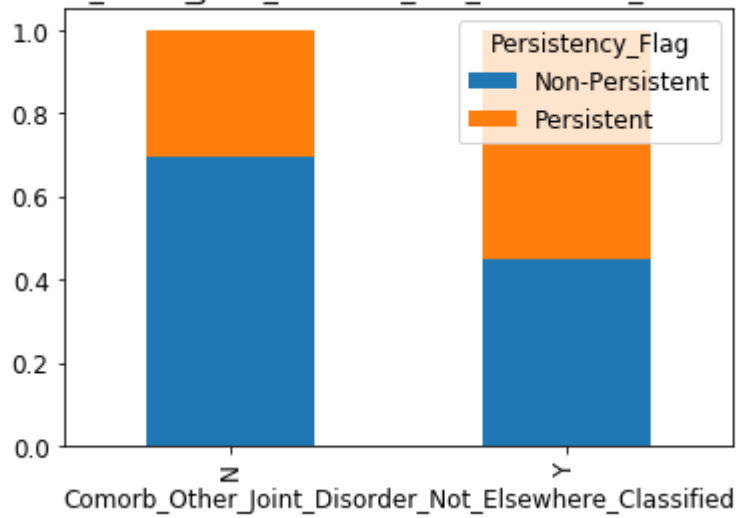

Stacked bar chart,Comorb_Encounter_For_Screening_For_Malignant_Neoplasms vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Encounter_For_Screening_For_Malignant_Neoplasms


Stacked bar chart,Comorb_Encounter_For_Immunization vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Encounter_For_Immunization

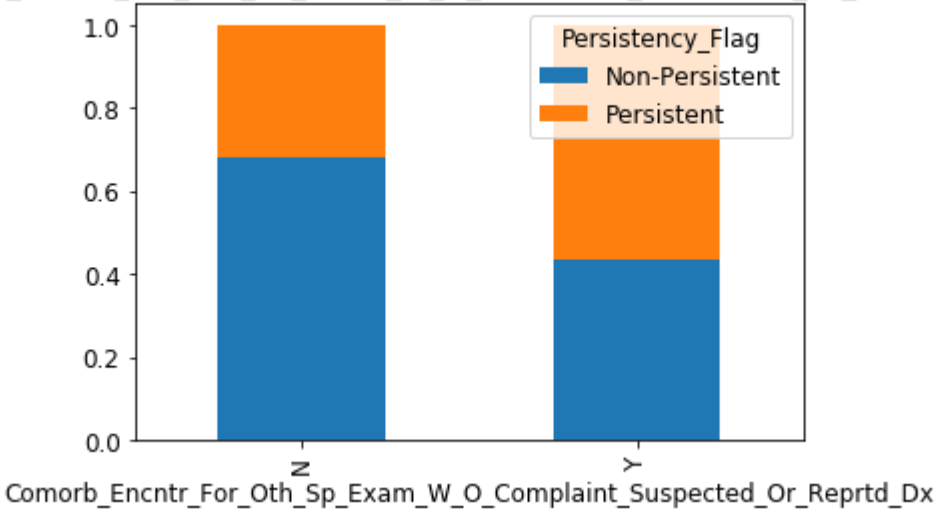

Stacked bar chart,Comorb_Encntr_For_General_Exam_W_O_Complaint_Susp_Or_Reprtd_Dx vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Encntr_For_General_Exam_W_O_Complaint_Susp_Or_Reprtd_Dx

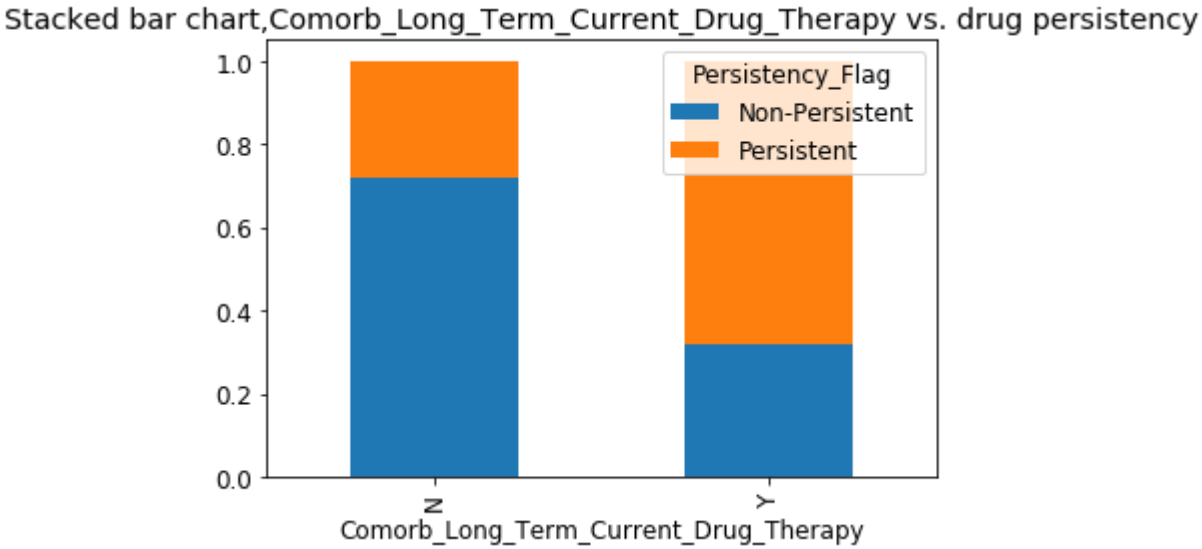

Stacked bar chart,Comorb_Vitamin_D_Deficiency vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Vitamin_D_Deficiency

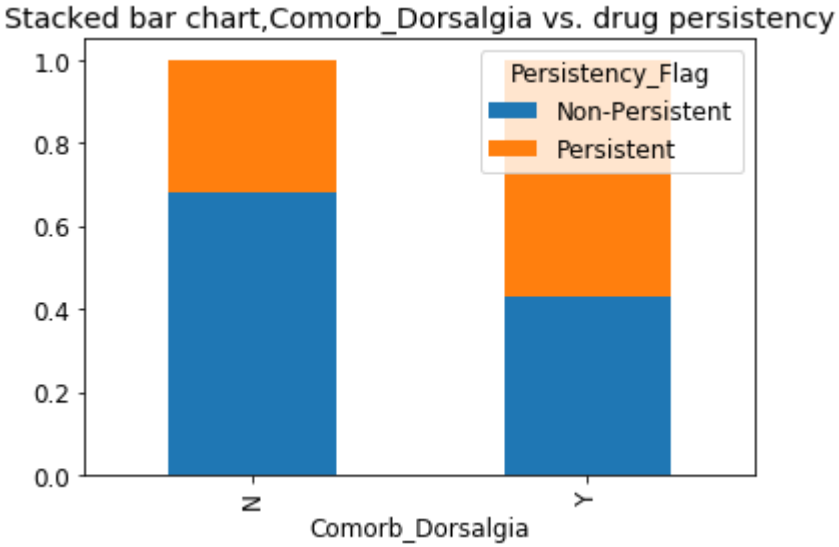
Stacked bar chart,Comorb_Other_Joint_Disorder_Not_Elsewhere_Classified vs. drug persistency

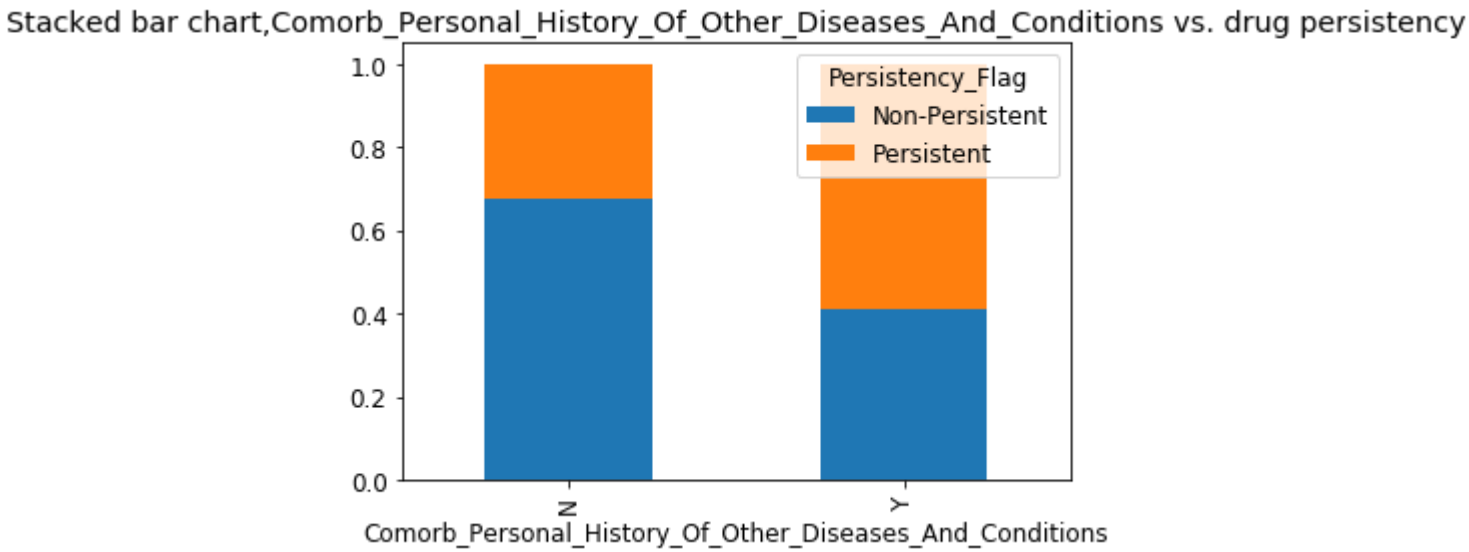


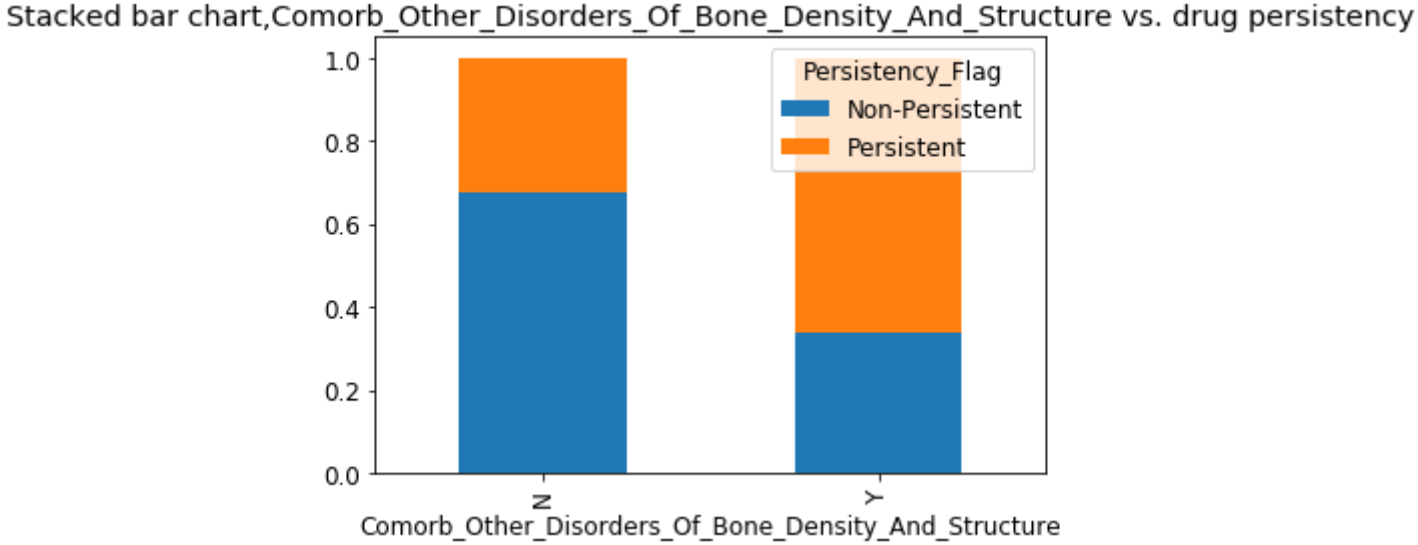
Stacked bar chart,Comorb_Encntr_For_Oth_Sp_Exam_W_O_Complaint_Suspected_Or_Reprtd_Dx vs. drug persistency



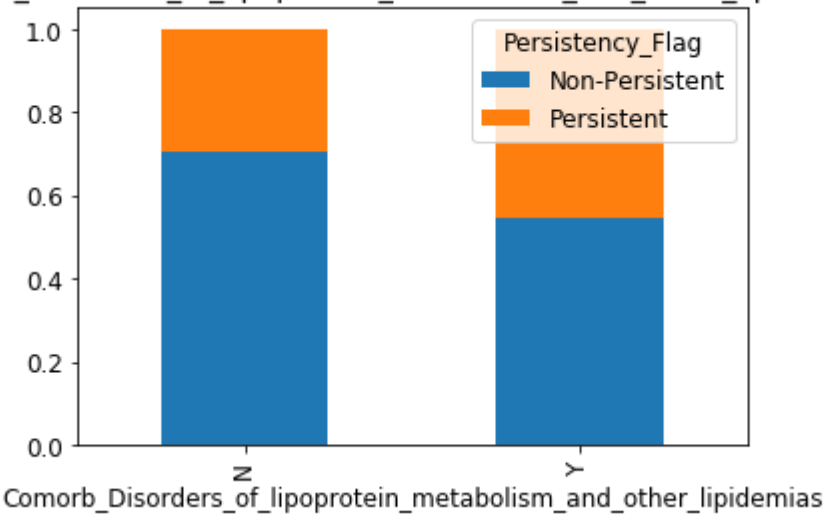

Stacked bar chart,Comorb_Long_Term_Current_Drug_Therapy vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Long_Term_Current_Drug_Therapy


Stacked bar chart,Comorb_Dorsalgia vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Dorsalgia

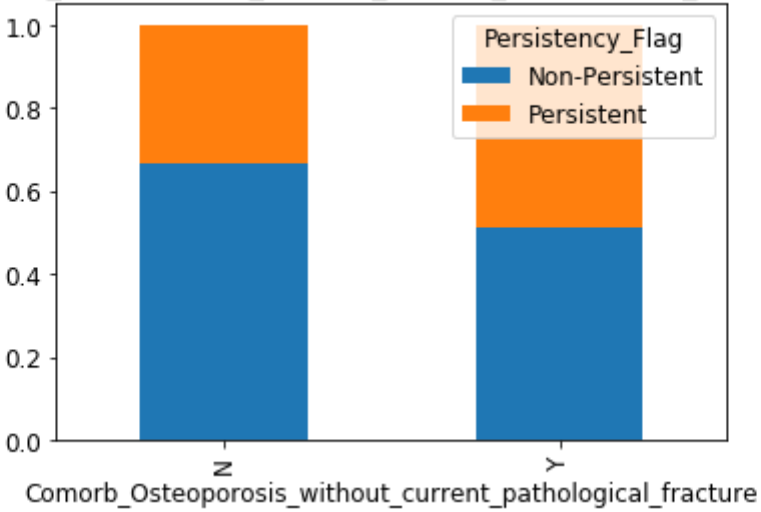

Stacked bar chart,Comorb_Personal_History_Of_Other_Diseases_And_Conditions vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Personal_History_Of_Other_Diseases_And_Conditions

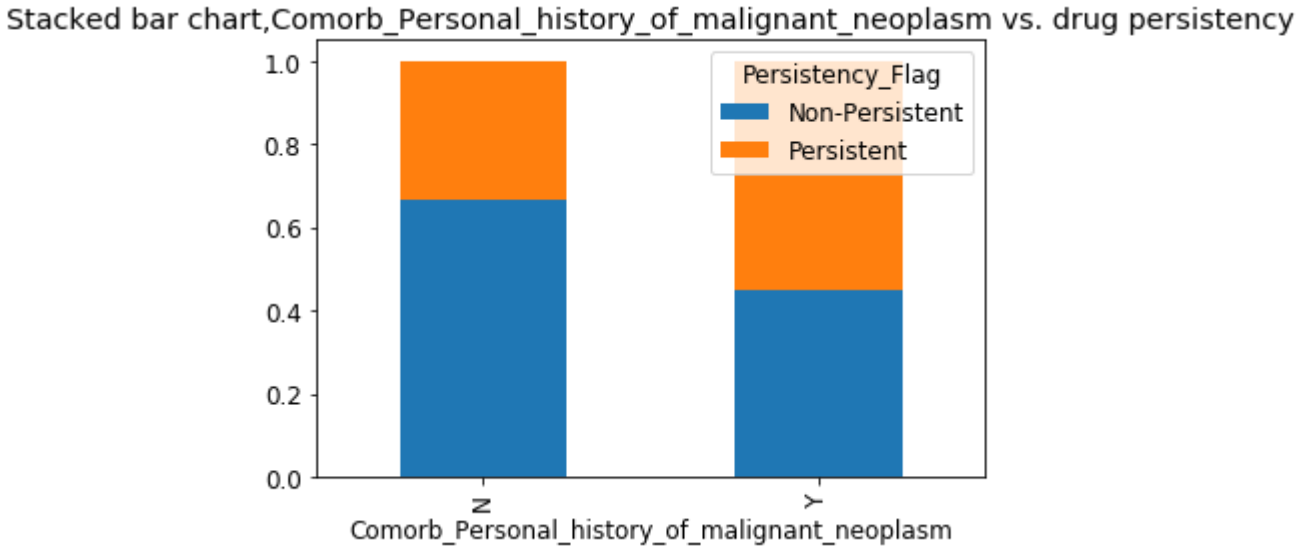

Stacked bar chart,Comorb_Other_Disorders_Of_Bone_Density_And_Structure vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Other_Disorders_Of_Bone_Density_And_Structure

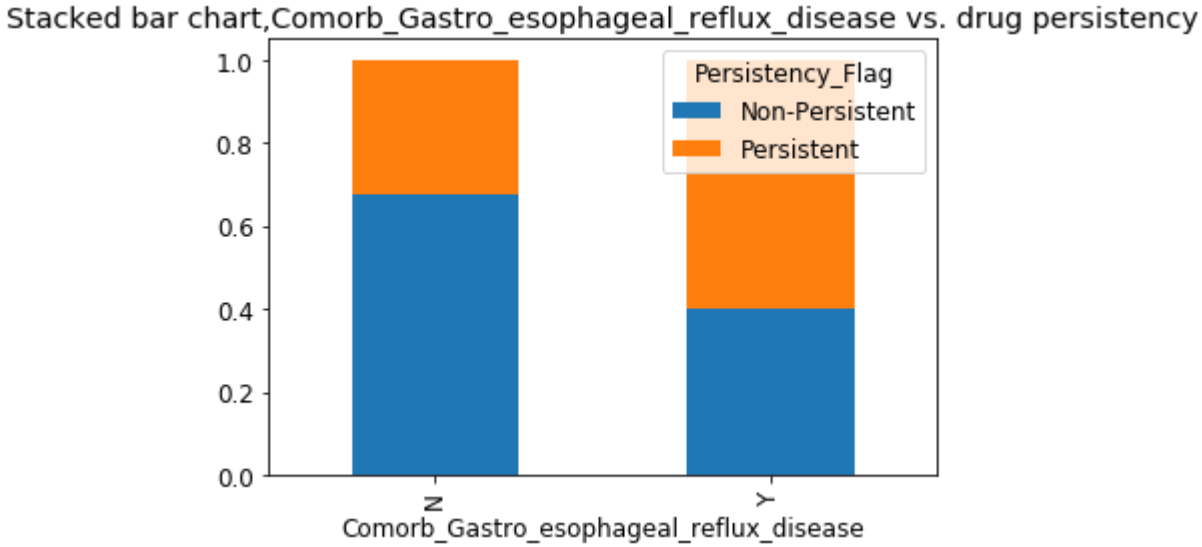
Stacked bar chart,Comorb_Disorders_of_lipoprotein_metabolism_and_other_lipidemias vs. drug persistency

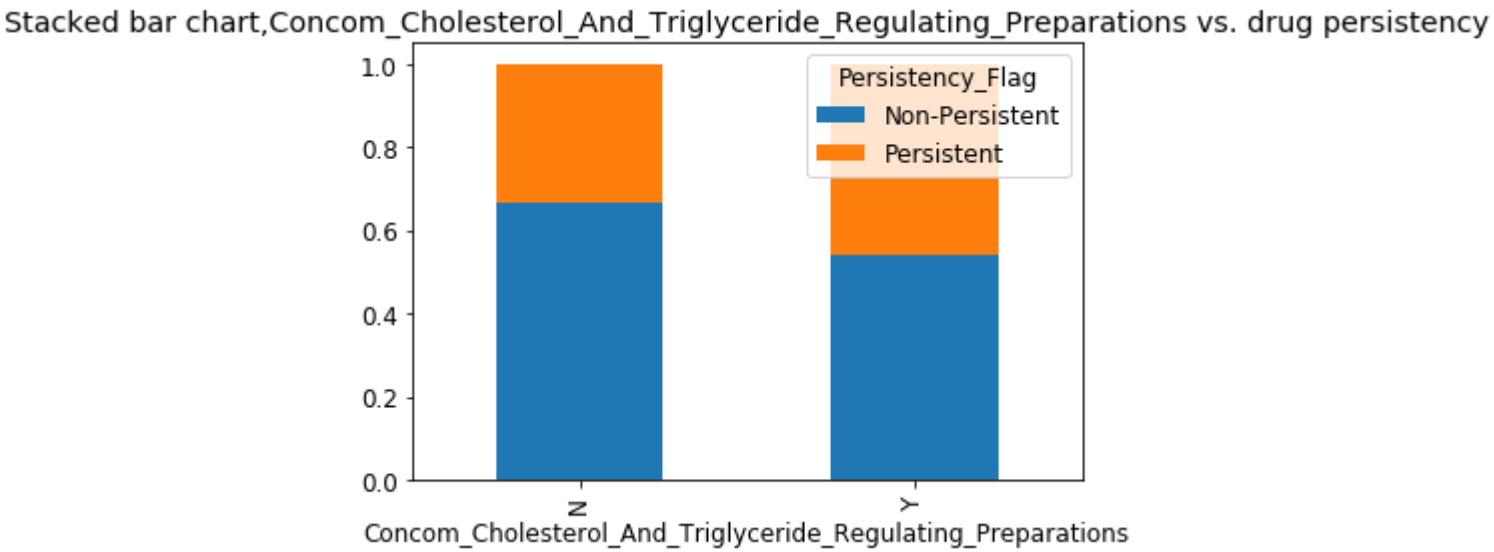


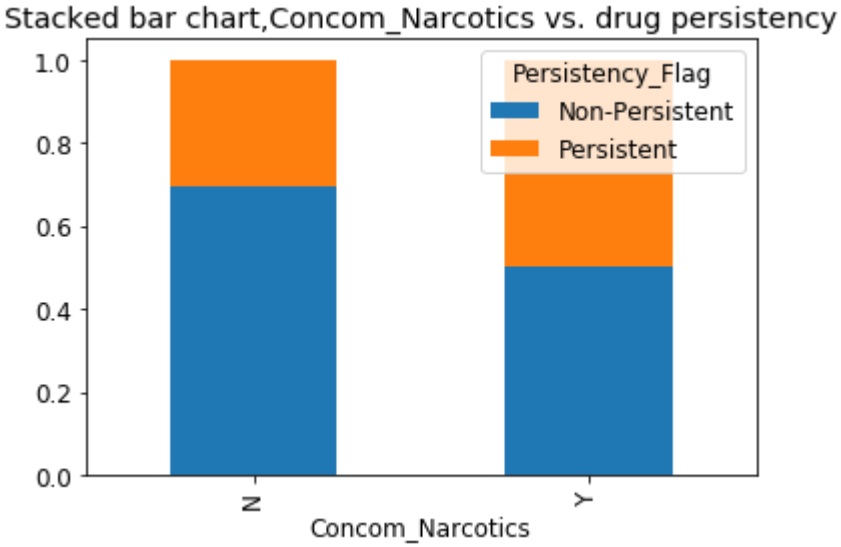
Stacked bar chart,Comorb_Osteoporosis_without_current_pathological_fracture vs. drug persistency

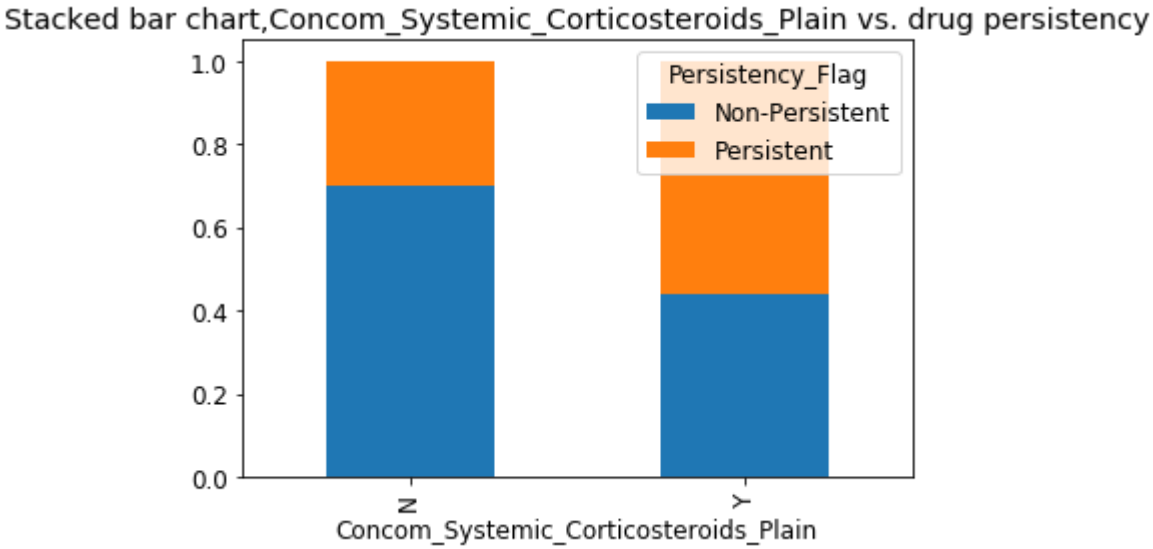


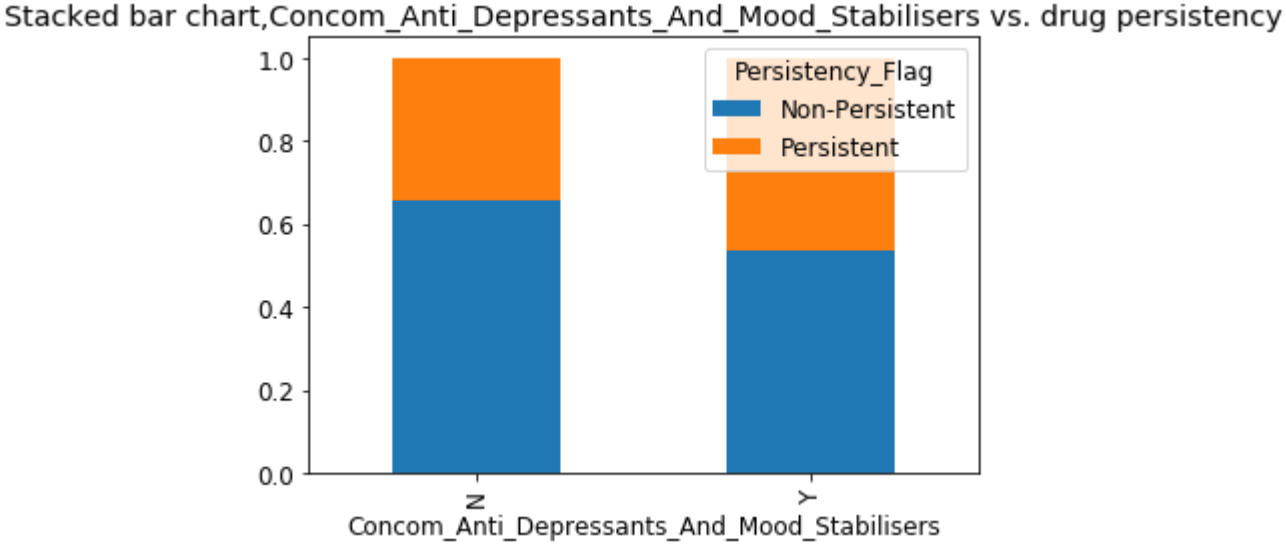

Stacked bar chart,Comorb_Personal_history_of_malignant_neoplasm vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Personal_history_of_malignant_neoplasm

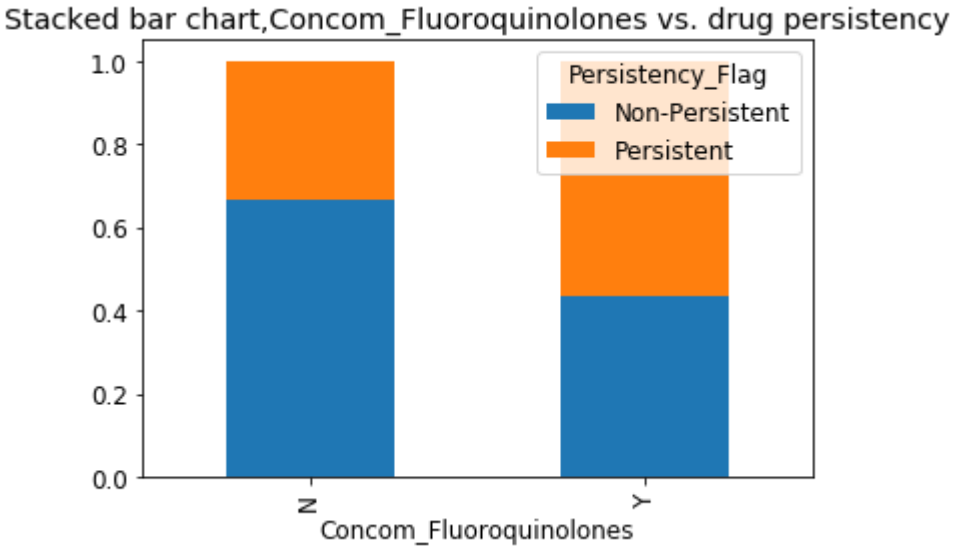

Stacked bar chart,Comorb_Gastro_esophageal_reflux_disease vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Comorb_Gastro_esophageal_reflux_disease

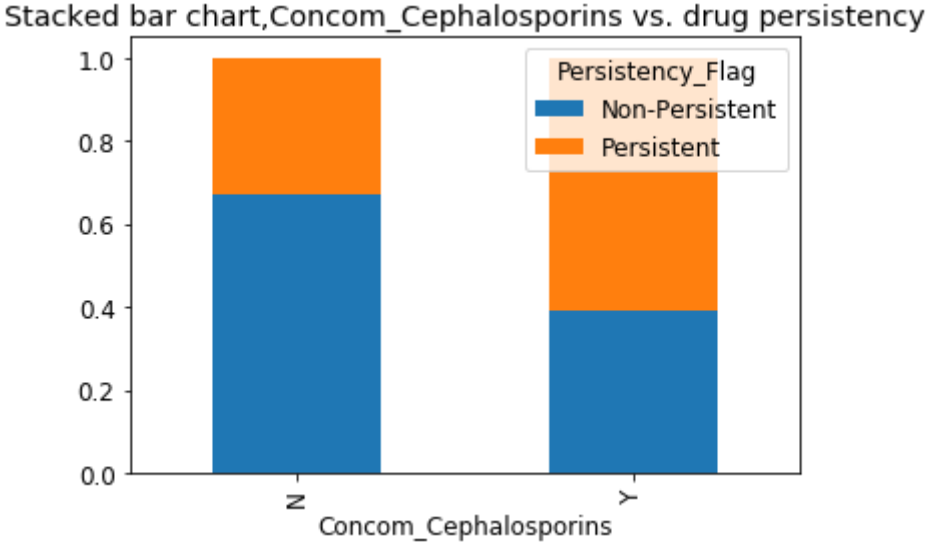

Stacked bar chart,Concom_Cholesterol_And_Triglyceride_Regulating_Preparations vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Cholesterol_And_Triglyceride_Regulating_Preparations

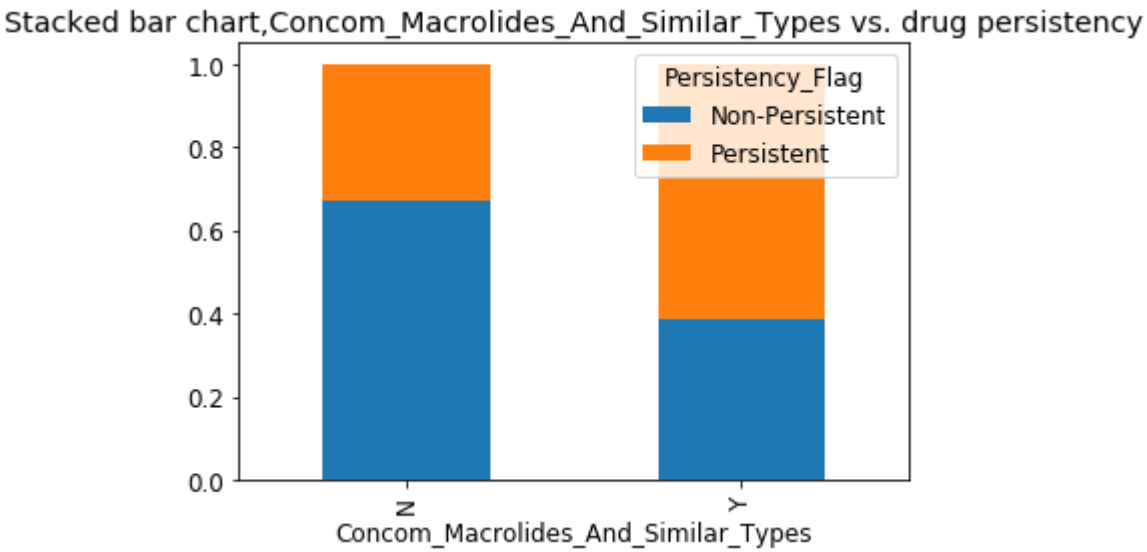

Stacked bar chart,Concom_Narcotics vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Narcotics

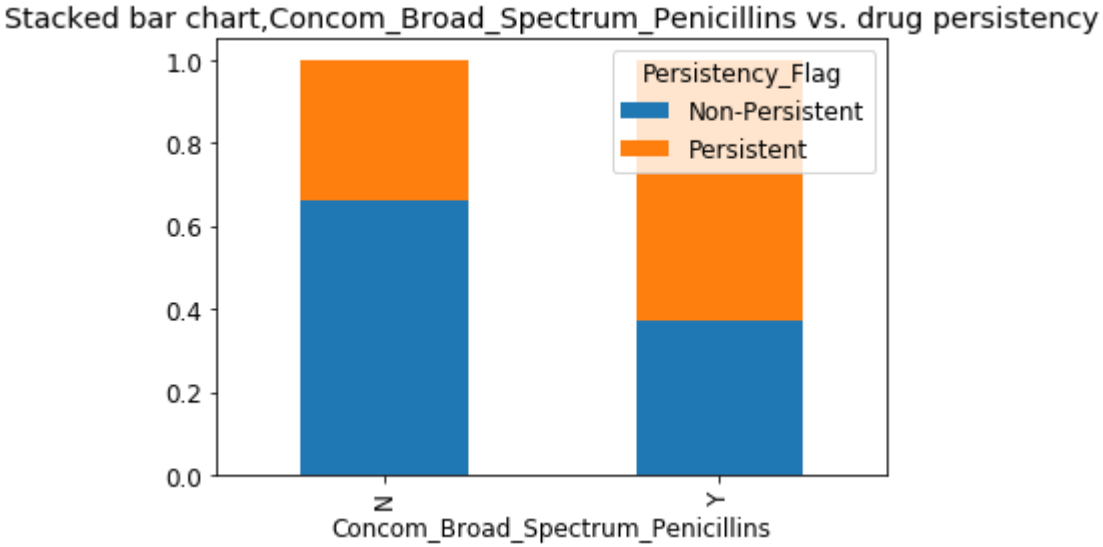

Stacked bar chart,Concom_Systemic_Corticosteroids_Plain vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Systemic_Corticosteroids_Plain

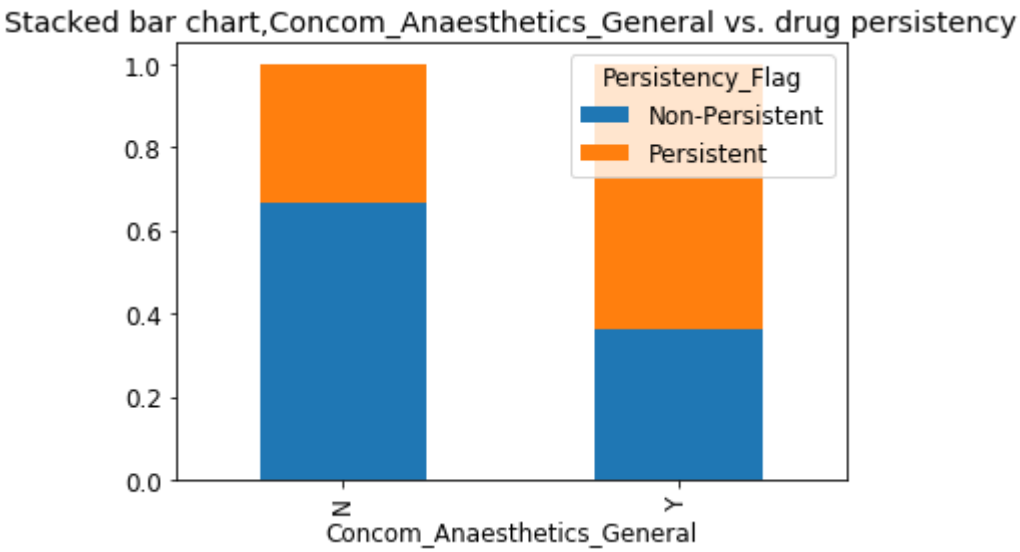

Stacked bar chart,Concom_Anti_Depressants_And_Mood_Stabilisers vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Anti_Depressants_And_Mood_Stabilisers


Stacked bar chart,Concom_Fluoroquinolones vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Concom_Fluoroquinolones


Stacked bar chart,Concom_Cephalosporins vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Concom_Cephalosporins


Stacked bar chart,Concom_Macrolides_And_Similar_Types vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Macrolides_And_Similar_Types


Stacked bar chart,Concom_Broad_Spectrum_Penicillins vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Broad_Spectrum_Penicillins

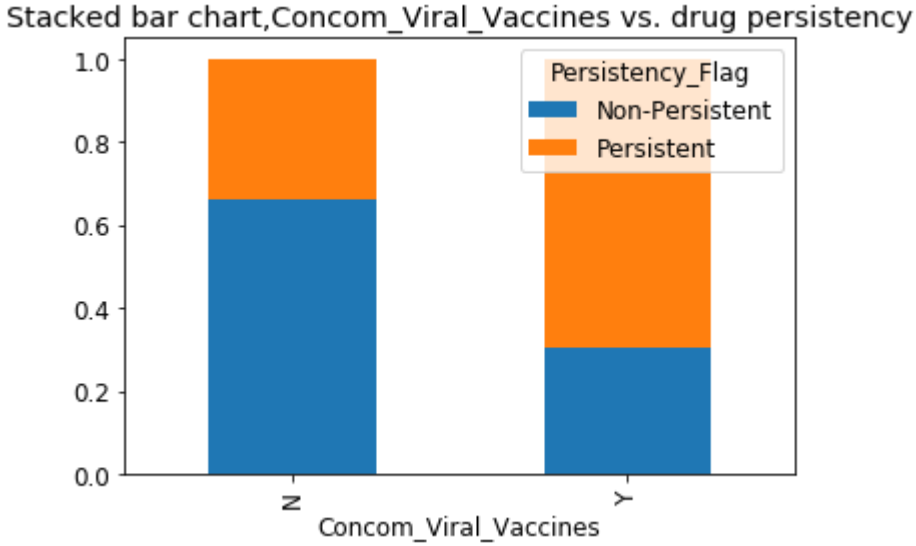

Stacked bar chart,Concom_Anaesthetics_General vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Concom_Anaesthetics_General

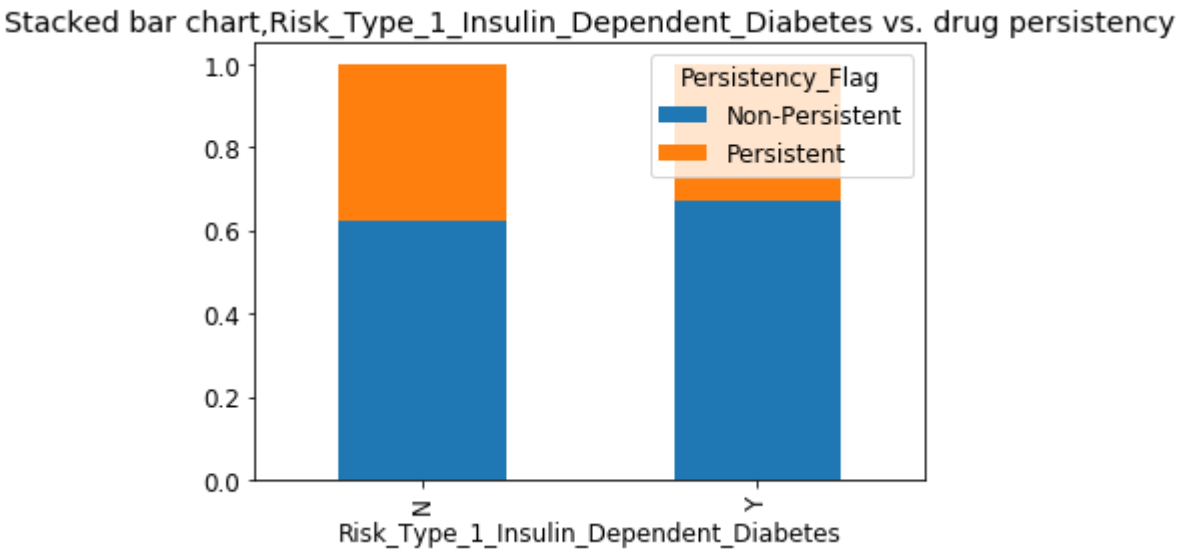
In [26]:
```
#do the same for the third clinical group

for item in clinical3:
    table = pd.crosstab(health[item], health.Persistency_Flag)
    table.div(table.sum(1).astype(float), axis = 0).plot(kind='bar', stacked = True)
    plt.title('Stacked bar chart,'+ item + ' vs. drug persistency')
    print(health[item].value_counts())
    print('---')
```

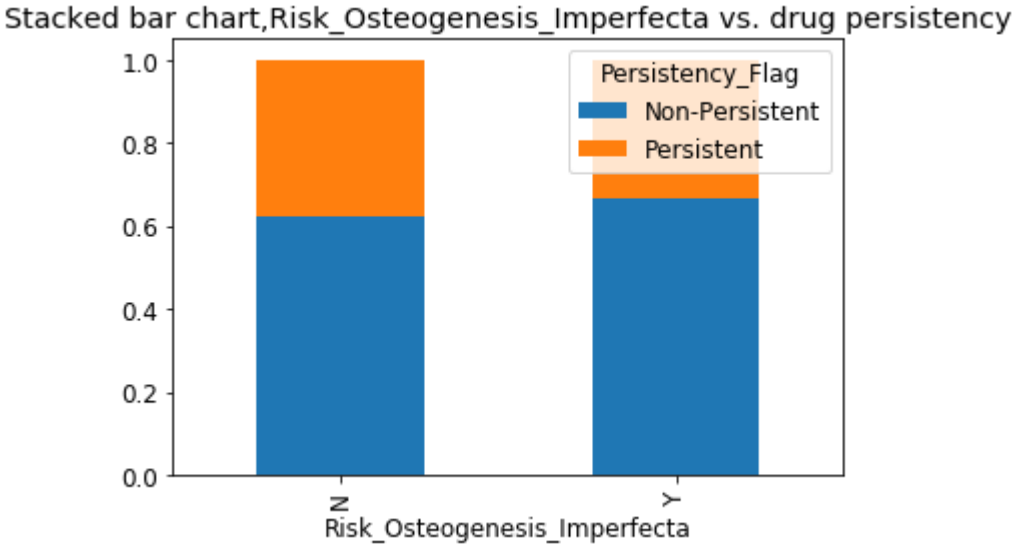
```
N    3071
Y     353
Name: Concom_Viral_Vaccines, dtype: int64
---
N    3285
Y     139
Name: Risk_Type_1_Insulin_Dependent_Diabetes, dtype: int64
---
N    3421
Y       3
Name: Risk_Osteogenesis_Imperfecta, dtype: int64
---
N    3294
Y     130
Name: Risk_Rheumatoid_Arthritis, dtype: int64
---
N    3422
Y       2
Name: Risk_Untreated_Chronic_Hyperthyroidism, dtype: int64
---
N    3297
Y     127
Name: Risk_Untreated_Chronic_Hypogonadism, dtype: int64
---
N    3412
Y      12
Name: Risk_Untreated_Early_Menopause, dtype: int64
---
N    3168
Y     256
Name: Risk_Patient_Parent_Fractured_Their_Hip, dtype: int64
---
N    2780
Y     644
Name: Risk_Smoking_Tobacco, dtype: int64
---
N    2954
Y     470
Name: Risk_Chronic_Malnutrition_Or_Malabsorption, dtype: int64
---
```

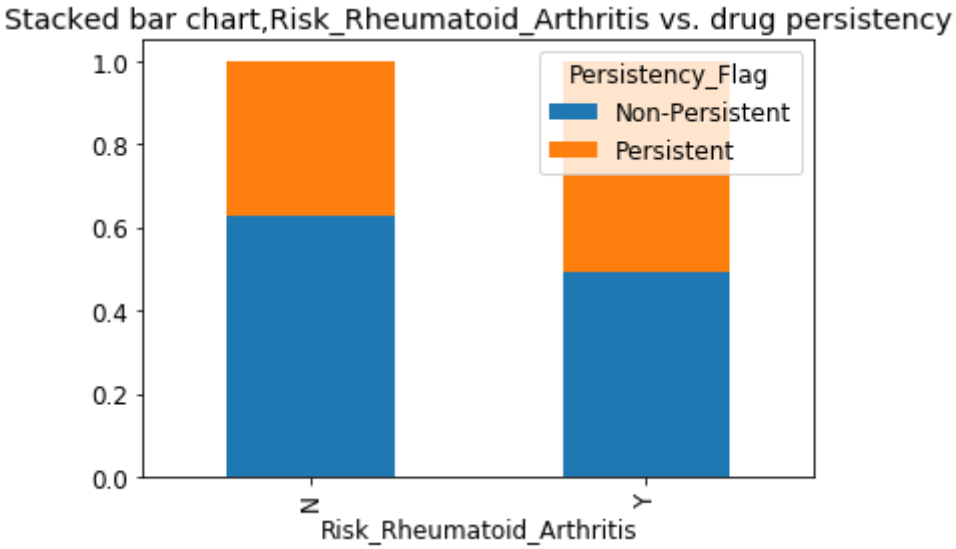
```
N    3406
Y      18
Name: Risk_Chronic_Liver_Disease, dtype: int64
---
N    3066
Y     358
Name: Risk_Family_History_Of_Osteoporosis, dtype: int64
---
N    3382
Y      42
Name: Risk_Low_Calcium_Intake, dtype: int64
---
N    1788
Y    1636
Name: Risk_Vitamin_D_Insufficiency, dtype: int64
---
N    3232
Y     192
Name: Risk_Poor_Health_Frailty, dtype: int64
---
N    3357
Y      67
Name: Risk_Excessive_Thinness, dtype: int64
---
N    3370
Y      54
Name: Risk_Hysterectomy_Oophorectomy, dtype: int64
---
N    3413
Y      11
Name: Risk_Estrogen_Deficiency, dtype: int64
---
N    3410
Y      14
Name: Risk_Immobilization, dtype: int64
---
N    3355
Y      69
Name: Risk_Recurring_Falls, dtype: int64
---
1    1242
```

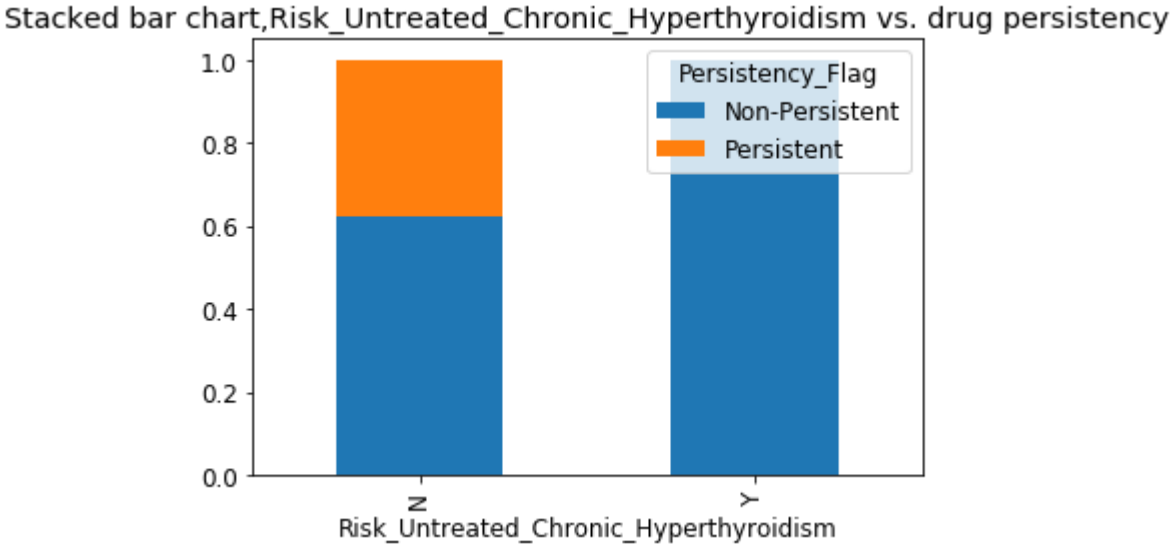
```
0     970
2     781
3     317
4      91
5      15
6       6
7       2
Name: Count_Of_Risks, dtype: int64
---
```

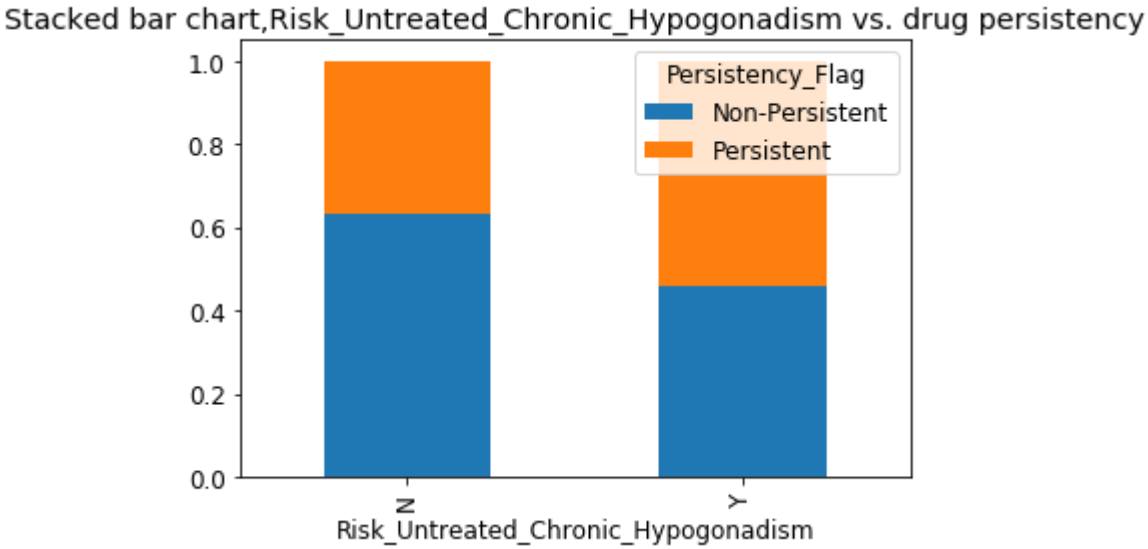


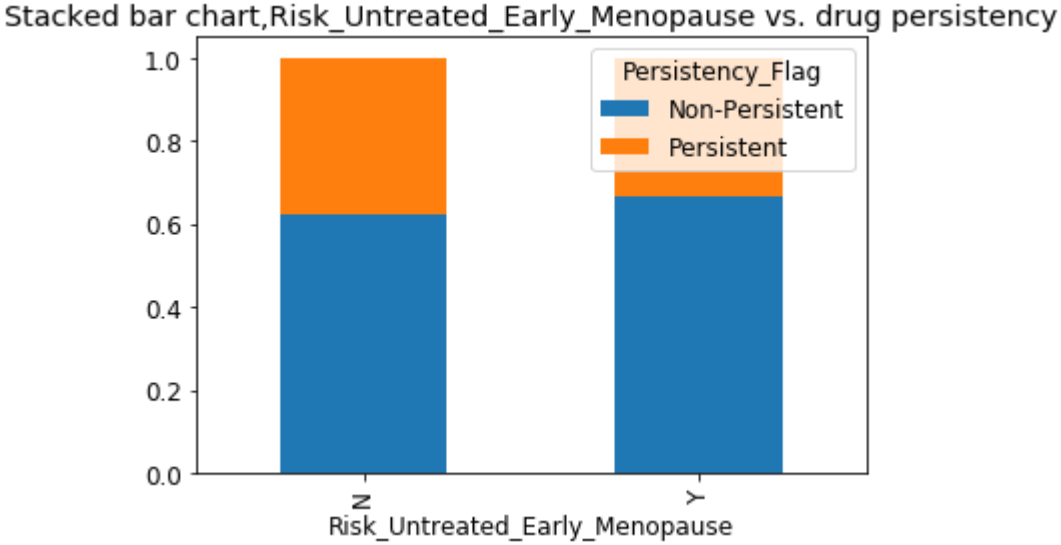

Stacked bar chart,Risk_Type_1_Insulin_Dependent_Diabetes vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Type_1_Insulin_Dependent_Diabetes

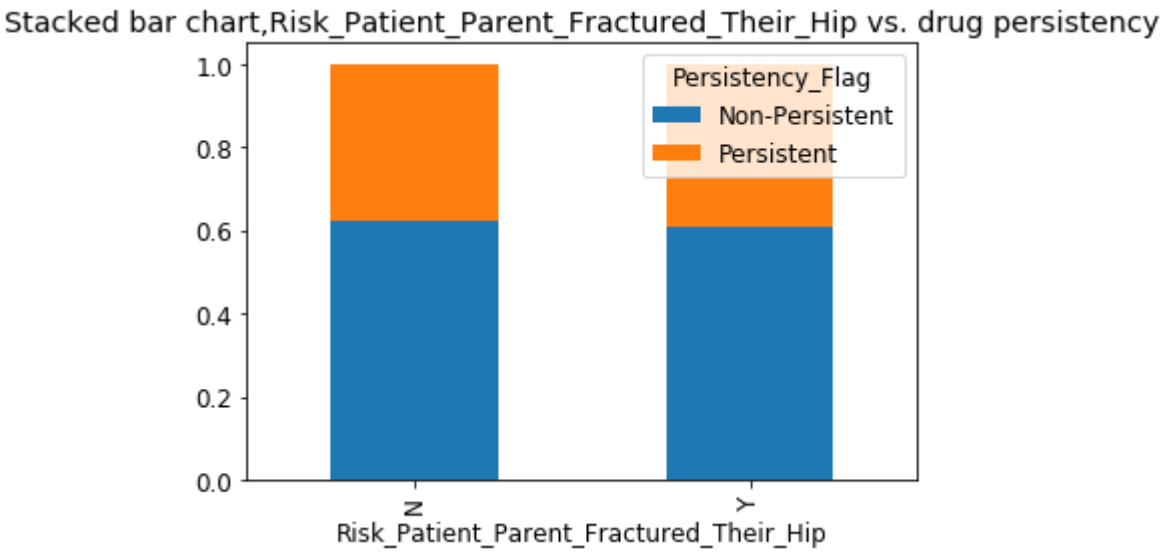

Stacked bar chart,Risk_Osteogenesis_Imperfecta vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Osteogenesis_Imperfecta

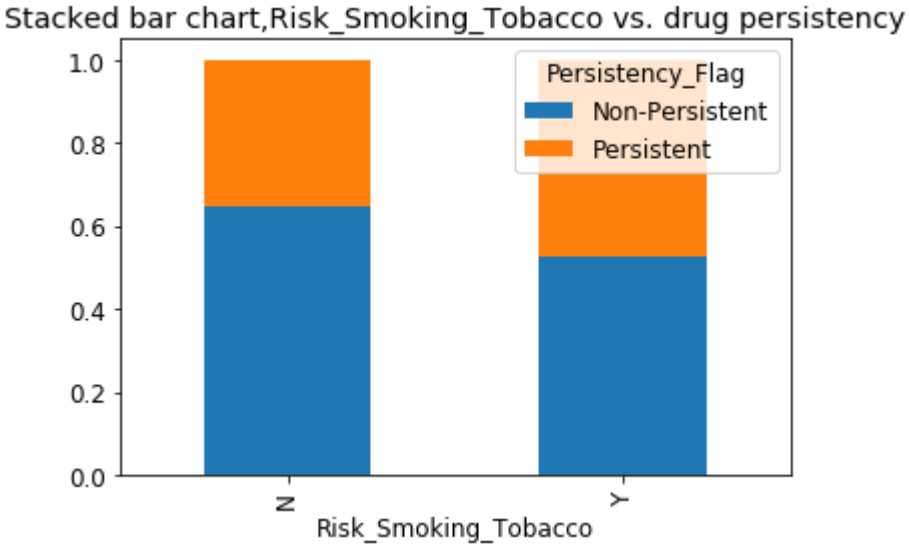

Stacked bar chart,Risk_Rheumatoid_Arthritis vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Rheumatoid_Arthritis

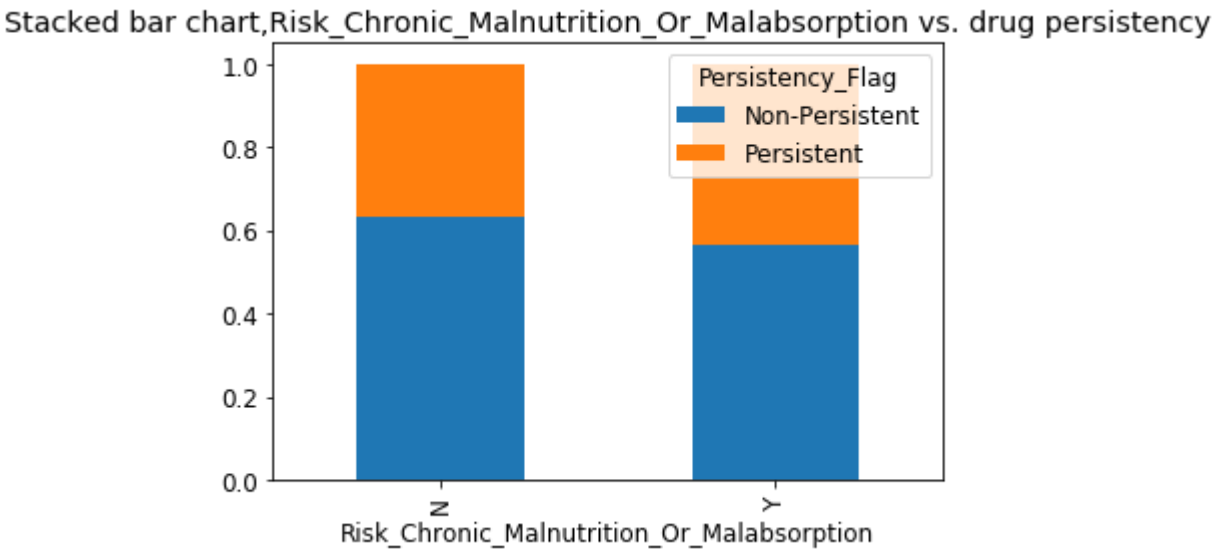

Stacked bar chart,Risk_Untreated_Chronic_Hyperthyroidism vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Untreated_Chronic_Hyperthyroidism

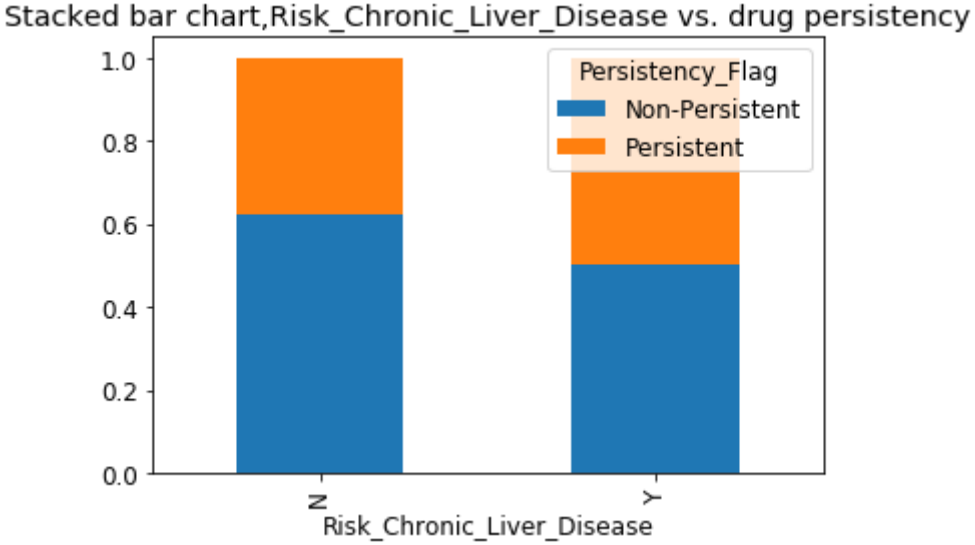

Stacked bar chart,Risk_Untreated_Chronic_Hypogonadism vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Untreated_Chronic_Hypogonadism

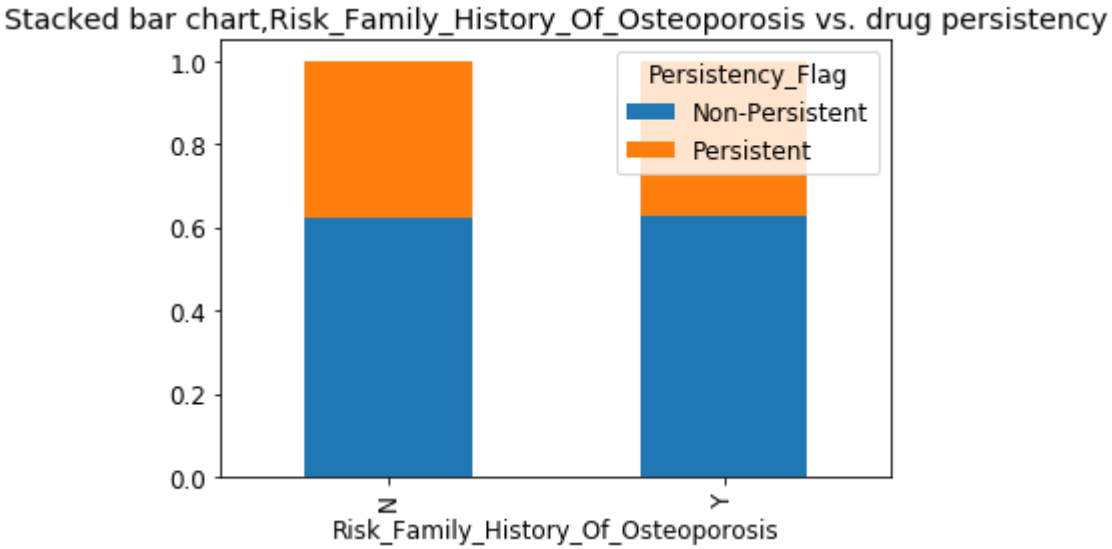

Stacked bar chart,Risk_Untreated_Early_Menopause vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Untreated_Early_Menopause

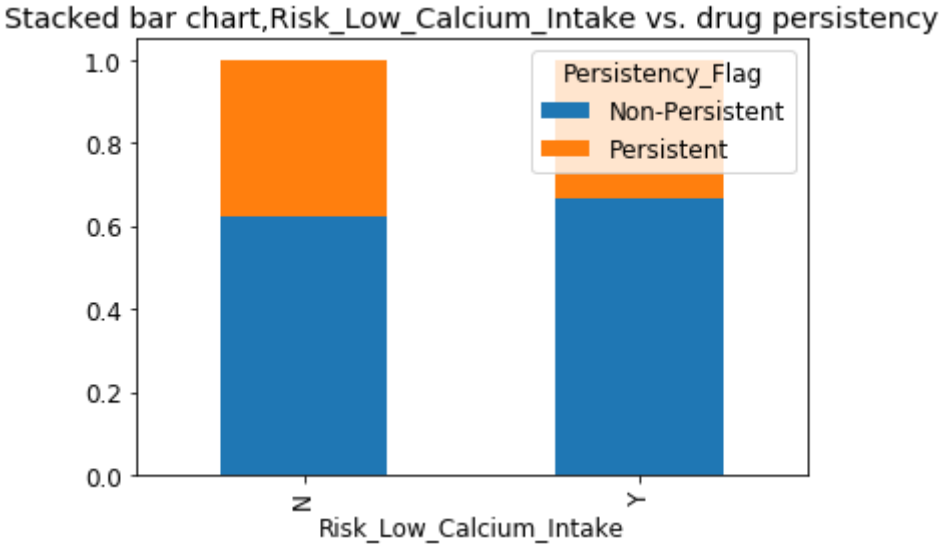

Stacked bar chart,Risk_Patient_Parent_Fractured_Their_Hip vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Patient_Parent_Fractured_Their_Hip

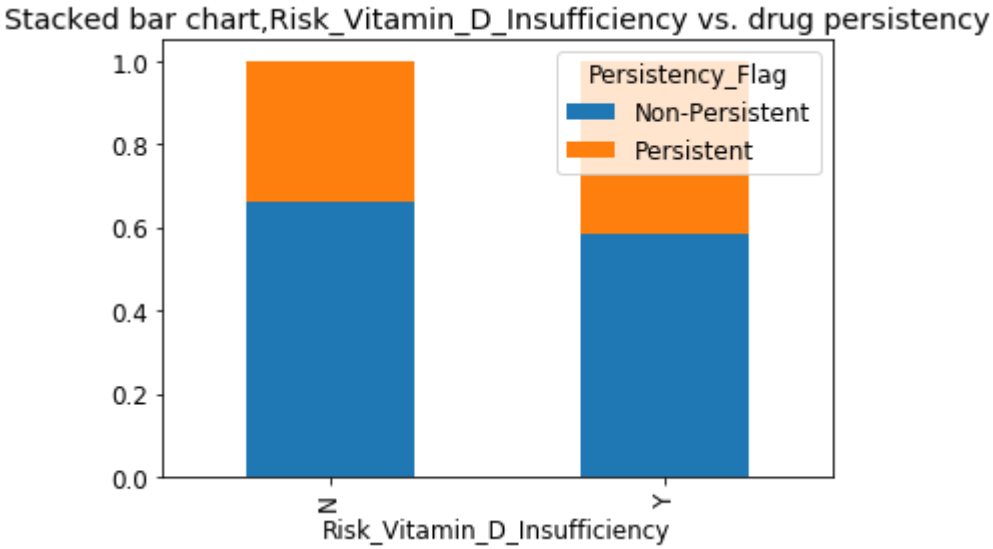

Stacked bar chart,Risk_Smoking_Tobacco vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Smoking_Tobacco

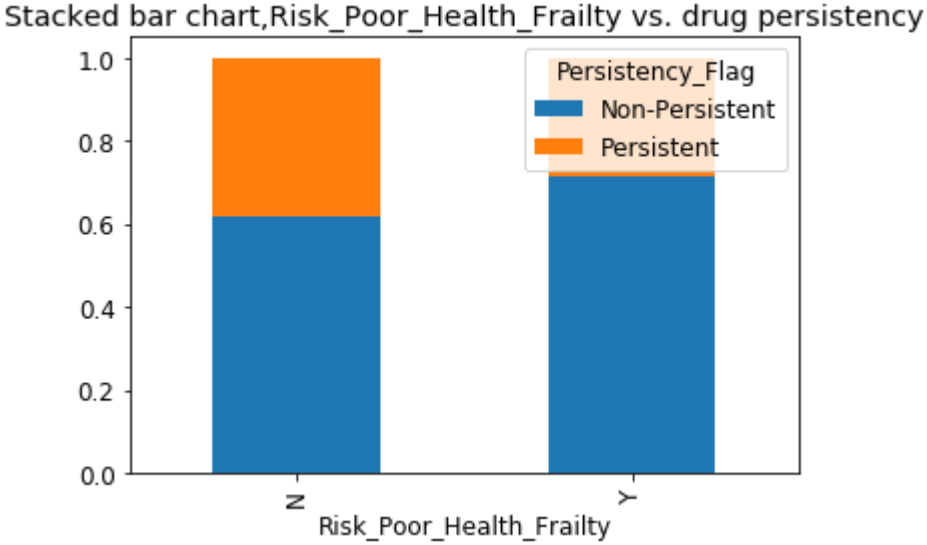

Stacked bar chart,Risk_Chronic_Malnutrition_Or_Malabsorption vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Chronic_Malnutrition_Or_Malabsorption

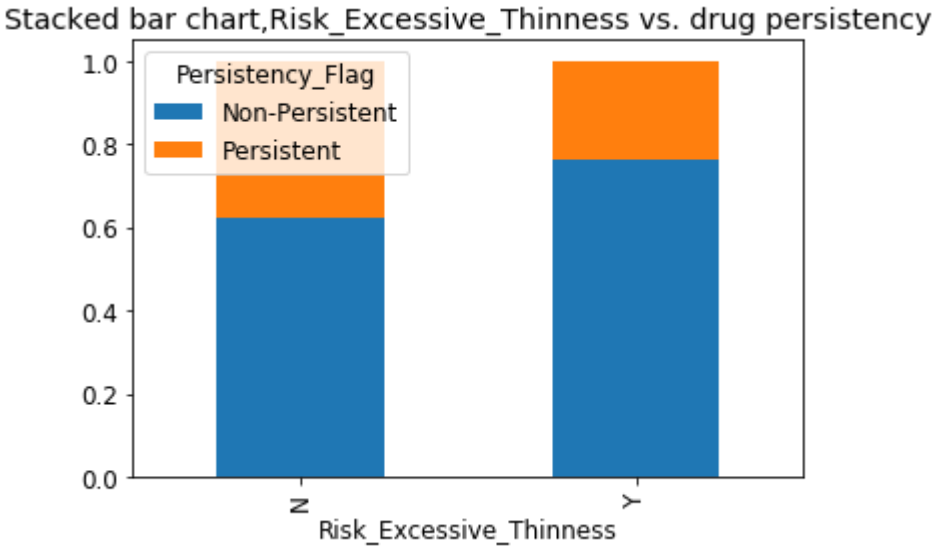

Stacked bar chart,Risk_Chronic_Liver_Disease vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Chronic_Liver_Disease

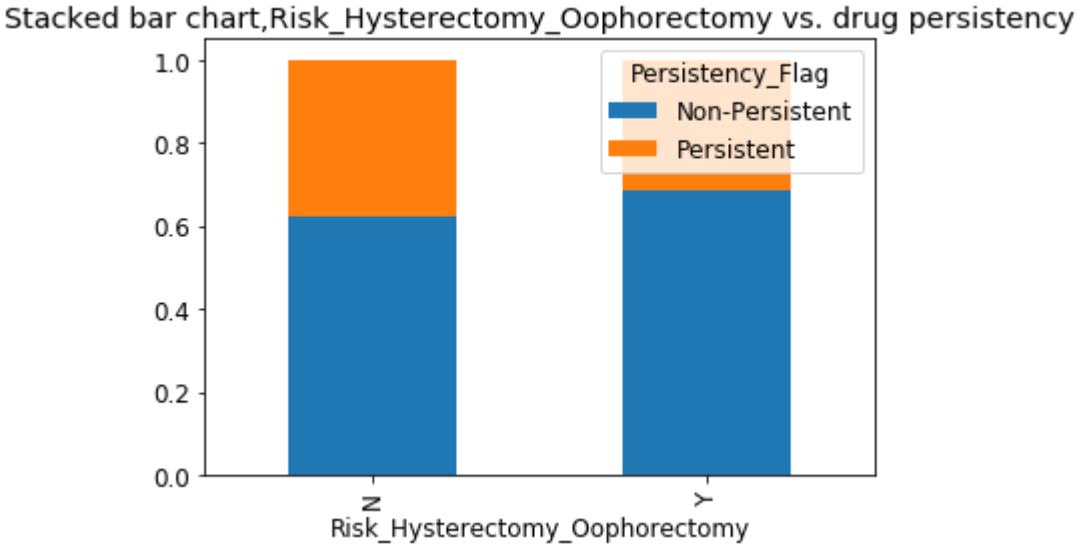

Stacked bar chart,Risk_Family_History_Of_Osteoporosis vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Family_History_Of_Osteoporosis

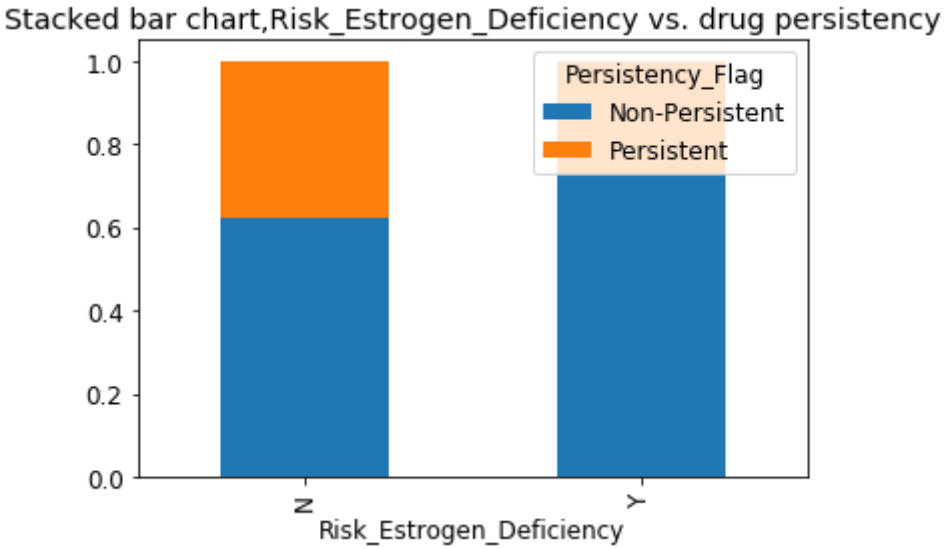

Stacked bar chart,Risk_Low_Calcium_Intake vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Low_Calcium_Intake

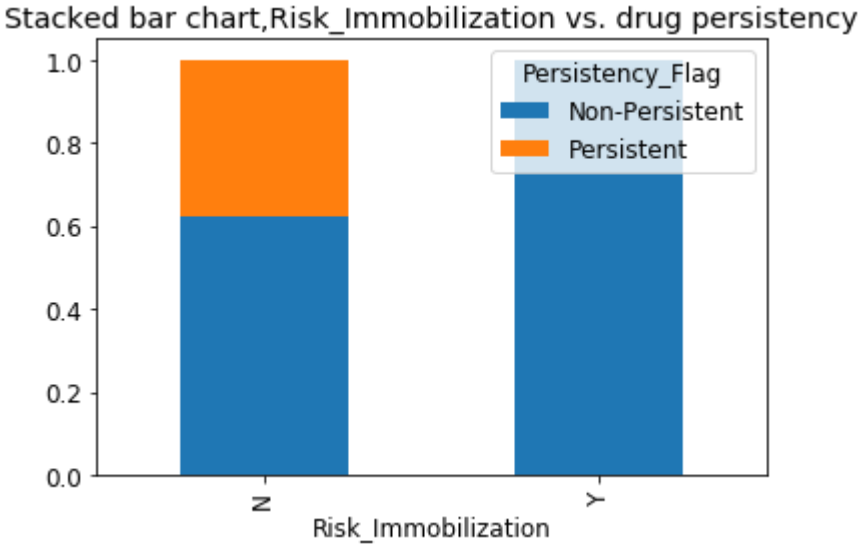

Stacked bar chart,Risk_Vitamin_D_Insufficiency vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Risk_Vitamin_D_Insufficiency

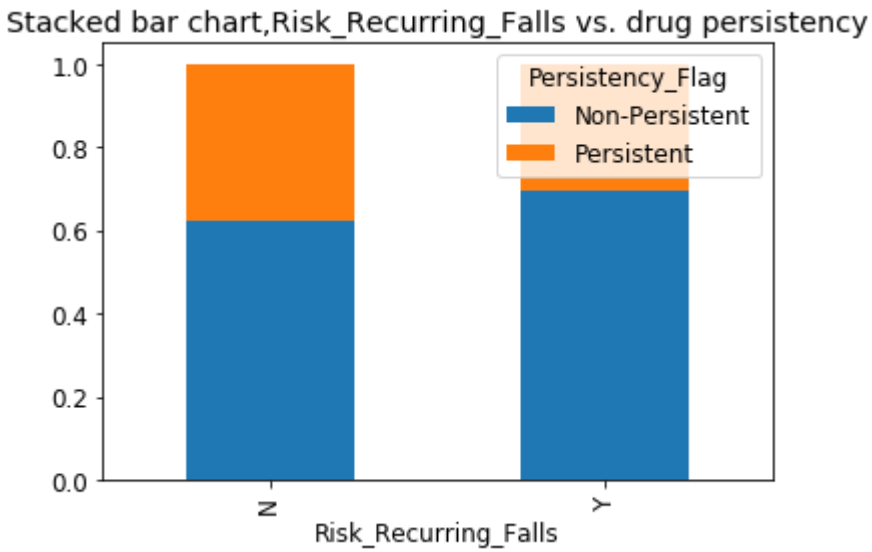

Stacked bar chart,Risk_Poor_Health_Frailty vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Risk_Poor_Health_Frailty

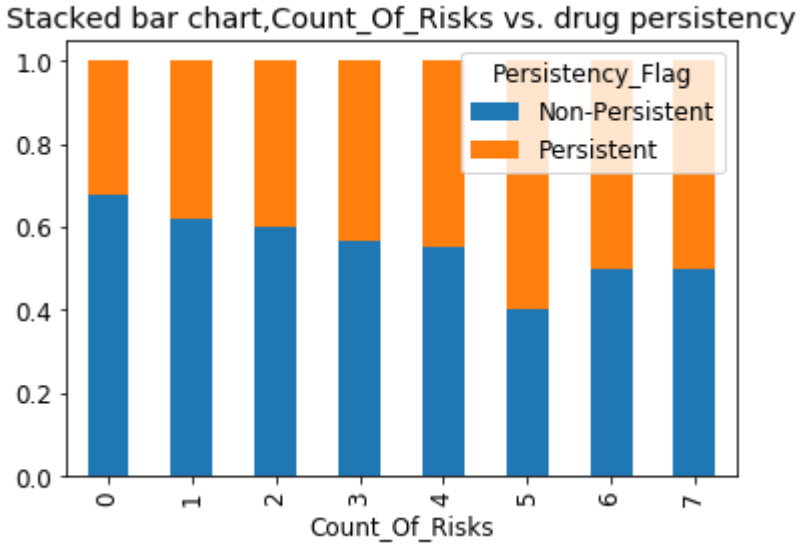

Stacked bar chart,Risk_Excessive_Thinness vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Risk_Excessive_Thinness


Stacked bar chart,Risk_Hysterectomy_Oophorectomy vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Risk_Hysterectomy_Oophorectomy


Stacked bar chart,Risk_Estrogen_Deficiency vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Estrogen_Deficiency


Stacked bar chart,Risk_Immobilization vs. drug persistency
1.0
0.8
0.6
0.4
0.2
0.0
Persistency_Flag
Non-Persistent
Persistent
N
Y
Risk_Immobilization


Stacked bar chart,Risk_Recurring_Falls vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
N
Y
Risk_Recurring_Falls


Stacked bar chart,Count_Of_Risks vs. drug persistency
Persistency_Flag
Non-Persistent
Persistent
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
Count_Of_Risks

In [ ]: