

# **Impact of HbA1c Measurement on Hospital Readmission Rates – Modelling and Pred**

# Summary of Data Preparations

To prepare the Data to be modelled, the following features were created:

- ✓ Create the age group variable
- ✓ Create the patient race variable
- ✓ Create the HbA1c variable using the A1Cresult and change features
- ✓ Creating the discharge disposition variable
- ✓ Creating the admission source variable
- ✓ Creating the admission source variable
- ✓ Creating the primary diagnosis variable
- ✓ Creating the physician speciality variable
- ✓ Creating the gender variable

```
prep_df <- cleaned_df |>
mutate(
  # Creating the AGE Variable
  'age_group' = case_when(
    (age %in% c('[0-10]', '[10-20]', '[20-30]')) ~ '30 years old or younger',
    (age %in% c('[30-40]', '[40-50]', '[50-60]')) ~ '30-60 years old',
    (age %in% c('[60-70]', '[70-80]', '[80-90]', '[90-100]')) ~ 'Older than 60'
  ),
  # Creating the Patient Race Variable
  'pat_race' = case_when(
    (race == 'AfricanAmerican') ~ 'African American',
    (race == 'Caucasian') ~ 'Caucasian',
    (race %in% c('Asian', 'Other', 'Hispanic')) ~ 'Other',
    (race == '?') ~ 'Missing',
  ),
  'HbA1c' = case_when(
    # Case 1: When No test is performed
    (A1Cresult == 'None') & (change %in% c('Ch', 'No')) ~ 'No Test was performed',
    # Case 2: When the Result is High and Med was changed
    (A1Cresult == '>8') & (change == 'Ch') ~ 'Result was high and the diabetic medication was changed',
    # Case 3: When result is high and Med was not changed
    (A1Cresult == '>8') & (change == 'No') ~ 'Result was high and the diabetic medication was not changed',
    # Case 4: Normal Test Result
    (A1Cresult %in% c('Norm', '>7')) & (change %in% c('Ch', 'No')) ~ 'Normal result of the test',
  ),
  'dscharge_disposition' = case_when(
    discharge_disposition_id == 1 ~ 'Discharged to home',
    .default = 'Otherwise'
  ),
  'admission_source' = case_when(
    admission_source_id == 7 ~ 'Admitted from Emergency Room',
    admission_source_id %in% c(1, 2) ~ 'Admitted because of physician/clinical referral',
    .default = 'Otherwise'
  ),
  'primary_diagnosis' = case_when(
    (diag_1 >= 390 & diag_1 <= 450) | (diag_1 == 785) ~ 'A disease of the circulatory system (icd9: 390-459, 785)',
    (diag_1 >= 250 & diag_1 <= 251) ~ 'Diabetes (icd9: 250.xx)',
    (diag_1 >= 460 & diag_1 <= 519) | (diag_1 == 786) ~ 'A disease of the respiratory system (icd9: 460-519, 786)',
    (diag_1 >= 520 & diag_1 <= 579) | (diag_1 == 787) ~ 'Diseases of the digestive system (icd9: 520-579, 787)',
    (diag_1 >= 800 & diag_1 <= 999) ~ 'Injury and poisoning (icd9: 800-999)',
    (diag_1 >= 710 & diag_1 <= 730) ~ 'Diseases of the musculoskeletal system and connective tissue (icd9: 710-730)',
    (diag_1 >= 580 & diag_1 <= 629) | (diag_1 == 788) ~ 'Diseases of the genitourinary system (icd9: 580-629, 788)',
    (diag_1 >= 140 & diag_1 <= 239) ~ 'Neoplasms (icd9: 140-239)',
    .default = 'Other'
  ),
  'physician_specialty' = case_when(
    medical_specialty == 'InternalMedicine' ~ 'Internal Medicine',
    medical_specialty == 'Cardiology' ~ 'Cardiology',
    str_starts(medical_specialty, 'Surg') ~ 'Surgery',
    medical_specialty == 'FamilyGeneralPractice' ~ 'Family/GeneralPractice',
    medical_specialty == 'missing' ~ 'Missing or Unknown',
    .default = 'Other'
  ),
  'gender' = case_when(
    gender == 'Unknown/Invalid' ~ NA,
    .default = gender
  )
) |>
# Select the relevant variables
select(c('gender', 'physician_specialty', 'primary_diagnosis',
  'admission_source', 'dscharge_disposition', 'HbA1c',
  'pat_race', 'age_group', 'time_in_hospital', 'readmitted')) |>
drop_na()
```

# Modelling Steps as Outlined in the Paper

- ✓ **Step 1:** Fit a Logistic regression model with all variables but `HbA1c`. This is referred to as the **\*\*core model\*\***.-
- ✓ **Step 2:** Add `HbA1c` to the **\*\*core model\*\***.-
- ✓ **Step 3:** Add pairwise interactions to the core model without `HbA1c` and keep only the significant ones.-
- ✓ **Step 4:** Add pairwise interactions with `HbA1c` leaving only the significant ones in the model.
- ✓ **Step 5:** Create the Final Model from Step 4.

## Step 1: Building the Core Model

```
# Specifying the model
log_reg_model <- logistic_reg() |>
  set_engine('glm')

# Building the recipe for the full model
full_model_rec <- recipe( readmitted ~ ., data = prep_df ) |>
  # Remove the HbA1c Variable
  step_rm(HbA1c) |>
  # Convert all the strings to factors
  step_string2factor(
    c('gender', 'physician_specialty', 'primary_diagnosis',
      'admission_source', 'dscharge_disposition', 'pat_race',
      'age_group'))

# Sanity Check: Preview the core-model-recipe
full_model_rec |>
  prep(prep_df) |>
  bake(new_data = NULL) |>
  glimpse()
```

```
Rows: 69,977
Columns: 9
$ gender           <fct> Female, Female, Female, Male, Male, Male, Male, M...
$ physician_specialty <fct> Other, Missing or Unknown, Missing or Unknown, Mi...
$ primary_diagnosis <fct> "Diabetes (icd9: 250.xx)", "Other", "Other", "Oth...
$ admission_source  <fct> Admitted because of physician/clinical referral,...
$ dscharge_disposition <fct> Otherwise, Discharged to home, Discharged to home...
$ pat_race          <fct> Caucasian, Caucasian, African American, Caucasian...
$ age_group         <fct> 30 years old or younger, 30 years old or younger,...
$ time_in_hospital  <dbl> 1, 3, 2, 2, 1, 3, 4, 5, 13, 12, 9, 7, 7, 10, 1, 1...
$ readmitted        <fct> No, No, No, No, No, No, No, No, No, No, No, Yes, ...
```

# Step 1: The Full Model

## The Full Model

```
# Fitting the Logistic regression model for the full model
full_model_fit <-
  # Build the workflow
  workflow() |>
  # Add the model
  add_model(log_reg_model) |>
  # Add the recipe
  add_recipe(full_model_rec) |>
  # Fit to data
  fit(prepare_df)

full_model <- full_model_fit |>
  extract_fit_engine()
```

term	statistic	df	p.value	significance
gender	1.111692	1	0.291714378	not significant
pat_race	7.543346	3	0.056455107	not significant
physician_specialty	31.67474	5	6.89E-06	significant
primary_diagnosis	55.62615	8	3.34E-09	significant
admission_source	9.494333	2	0.008676245	significant
dscharge_disposition	337.5848	1	2.14E-75	significant
age_group	46.26966	2	8.97E-11	significant
time_in_hospital	47.39944	1	5.79E-12	significant

Table 2: ANOVA Results from the Full Model

## The Results

term	estimate	std.error	statistic	p.value	significance
genderMale	0.028447	0.026974	1.054627	0.291596	not significant
physician_specialtyMissing or Unknown	0.164857	0.0653	2.524613	0.011583	not significant
physician_specialtyOther	0.09589	0.071243	1.345946	0.17832	not significant
physician_specialtySurgery	0.037369	0.087226	0.428419	0.668346	not significant
primary_diagnosisDiabetes (icd9: 250.xx)	-0.01698	0.05396	-0.31462	0.753054	not significant
primary_diagnosisDiseases of the genitourinary system (icd9: 580-629, 788)	-0.09869	0.064867	-1.52149	0.128138	not significant
primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.00452	0.053596	-0.08432	0.9328	not significant
primary_diagnosisNeoplasms (icd9: 140-239)	-0.07877	0.074842	-1.05251	0.292564	not significant
primary_diagnosisOther	-0.09471	0.040809	-2.32095	0.020289	not significant
admission_sourceAdmitted from Emergency Room	0.018203	0.03185	0.571518	0.567649	not significant
admission_sourceOtherwise	-0.1052	0.043956	-2.39341	0.016693	not significant
pat_raceCaucasian	0.017146	0.036143	0.474392	0.63522	not significant
pat_raceMissing	-0.18701	0.094015	-1.98919	0.04668	not significant
pat_raceOther	-0.0865	0.074609	-1.15936	0.24631	not significant
age_group30 years old or younger	-0.08972	0.103739	-0.86485	0.38712	not significant
(Intercept)	-2.91886	0.075571	-38.6242	0	significant
physician_specialtyFamily/GeneralPractice	0.274737	0.07874	3.489165	4.85E-04	significant
physician_specialtyInternal Medicine	0.279933	0.070778	3.95508	7.65E-05	significant
primary_diagnosisA disease of the respiratory system (icd9: 460-519, 786)	-0.29681	0.04651	-6.38173	1.75E-10	significant
primary_diagnosisDiseases of the digestive system (icd9: 520-579, 787)	-0.14336	0.051951	-2.75956	0.005788	significant
primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710-739)	-0.22995	0.065024	-3.5363	4.06E-04	significant
dscharge_dispositionOtherwise	0.53949	0.029302	18.41119	1.07E-75	significant
age_groupOlder than 60	0.205144	0.031923	6.426179	1.31E-10	significant
time_in_hospital	0.031357	0.004504	6.962152	3.35E-12	significant

Table 1: Logistic Regression Results for the Full Model

# Step 1: The Core Model

```
# The core model

# Building the for the full model
core_model_rec <- recipe( readmitted ~ ., data = prep_df ) |>
  # Remove the HbA1c Variable
  step_rm(HbA1c, gender) |>
  # Convert all the strings to factors
  step_string2factor(
    c('physician_specialty','primary_diagnosis',
      'admission_source','dscharge_disposition','pat_race',
      'age_group'))

# Fitting the logistic regression model for the full model
core_model_fit <-
  # Build the workflow
  workflow() |>
  # Add the model
  add_model(log_reg_model) |>
  # Add the recipe
  add_recipe(core_model_rec) |>
  # Fit to data
  fit(prep_df)

core_model <- core_model_fit |>
  extract_fit_engine()
```

	term	estimate	std.error	statistic	p.value
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
	(Intercept)	-2.903199527	0.074077856	-39.1911924	0.000000e+00
	physician_specialtyFamily/GeneralPractice	0.272753401	0.078714579	3.4650938	5.300466e-04
	physician_specialtyInternal Medicine	0.277786817	0.070746261	3.9265230	8.618267e-05
	physician_specialtyMissing or Unknown	0.163631379	0.065288691	2.5062745	1.220109e-02
	physician_specialtyOther	0.094430322	0.071230504	1.3257006	1.849389e-01
	physician_specialtySurgery	0.037023896	0.087226850	0.4244553	6.712338e-01
	primary_diagnosisA disease of the respiratory system (icd9: 460–519, 786)	-0.298758464	0.046472628	-6.4286975	1.287020e-10
	primary_diagnosisDiabetes (icd9: 250.xx)	-0.016212990	0.053950744	-0.3005147	7.637846e-01
	primary_diagnosisDiseases of the digestive system (icd9: 520–579, 787)	-0.145371418	0.051916234	-2.8001148	5.108443e-03
	primary_diagnosisDiseases of the genitourinary system (icd9: 580–629, 788)	-0.101363414	0.064817003	-1.5638399	1.178552e-01
	primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710–739)	-0.232447918	0.064982663	-3.5770759	3.474593e-04
	primary_diagnosisInjury and poisoning (icd9: 800–999)	-0.006976068	0.053548333	-0.1302761	8.963480e-01
	primary_diagnosisNeoplasms (icd9: 140–239)	-0.080094317	0.074829418	-1.0703587	2.844579e-01
	primary_diagnosisOther	-0.096901462	0.040756501	-2.3775707	1.742710e-02
	admission_sourceAdmitted from Emergency Room	0.017974658	0.031848909	0.5643728	5.725005e-01
	admission_sourceOtherwise	-0.105543197	0.043956645	-2.4010749	1.634699e-02
	dscharge_dispositionOtherwise	0.537838599	0.029263292	18.3792922	1.924635e-75
	pat_raceCaucasian	0.019970223	0.036042277	0.5540777	5.795257e-01
	pat_raceMissing	-0.183823327	0.093964869	-1.9562984	5.043001e-02
	pat_raceOther	-0.083780429	0.074562221	-1.1236311	2.611696e-01
	age_group30 years old or younger	-0.093465881	0.103669428	-0.9015761	3.672821e-01
	age_groupOlder than 60	0.203422499	0.031884044	6.3800721	1.770046e-10
	time_in_hospital	0.031366743	0.004504312	6.9637147	3.314154e-12

# Step 2: Fit a Logistic Regression model of all variables including HbA1c

```
# The core model

# Building the recipe for the full model
core_hb_model_rec <- recipe( readmitted |~ ., data = prep_df ) |>
  # Remove the Gender Variable
  step_rm(gender) |>
  # Convert all the strings to factors
  step_string2factor(
    c(physician_specialty,primary_diagnosis,
      admission_source,dscharge_disposition,pat_race,
      age_group, HbA1c))

# Sanity Check: Preview the core-model-recipe
# core_hb_model_rec |>
#   prep(prep_df) |>
#   bake(new_data = NULL) |>
#   glimpse()

# Fitting the Logistic regression model for the full model
core_hb_model_fit <-
  # Build the workflow
  workflow() |>
  # Add the model
  add_model(log_reg_model) |>
  # Add the recipe
  add_recipe(core_hb_model_rec) |>
  # Fit to data
  fit(prep_df)

core_hb_model <- core_hb_model_fit |>
  extract_fit_engine()
```

term	estimate	std.error	statistic	p.value	significance
physician_specialtyMissing or Unknown	0.163898	0.065299	2.509954	0.012075	not significant
physician_specialtyOther	0.095406	0.071232	1.339378	0.180448	not significant
physician_specialtySurgery	0.034055	0.087232	0.390397	0.696243	not significant
primary_diagnosisDiabetes (icd9: 250.xx)	-0.00756	0.054254	-0.13941	0.889123	not significant
primary_diagnosisDiseases of the genitourinary system (icd9: 580-629, 788)	-0.10608	0.06484	-1.63608	0.101823	not significant
primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.0141	0.053615	-0.26307	0.792494	not significant
primary_diagnosisNeoplasms (icd9: 140-239)	-0.0881	0.074882	-1.17649	0.2394	not significant
primary_diagnosisOther	-0.1002	0.040781	-2.45691	0.014014	not significant
admission_sourceAdmitted from Emergency Room	0.024371	0.03194	0.763015	0.445454	not significant
admission_sourceOtherwise	-0.1069	0.043981	-2.43058	0.015075	not significant
HbA1cNormal result of the test	-0.10293	0.046818	-2.19843	0.027919	not significant
HbA1cResult was high and the diabetic medication was changed	-0.06054	0.059543	-1.0167	0.309295	not significant
HbA1cResult was high and the diabetic medication was not changed	-0.16029	0.084606	-1.89453	0.058155	not significant
pat_raceCaucasian	0.020401	0.036048	0.565927	0.571443	not significant
pat_raceMissing	-0.18122	0.094001	-1.92785	0.053874	not significant
pat_raceOther	-0.07859	0.074592	-1.0536	0.292064	not significant
age_group30 years old or younger	-0.07893	0.103914	-0.75962	0.447484	not significant
(Intercept)	-2.88788	0.074271	-38.883	0	significant
physician_specialtyFamily/General Practice	0.274637	0.078725	3.488555	4.86E-04	significant
physician_specialtyInternal Medicine	0.282379	0.070771	3.990019	6.61E-05	significant
primary_diagnosisA disease of the respiratory system (icd9: 460-519, 786)	-0.30106	0.046491	-6.47552	9.45E-11	significant
primary_diagnosisDiseases of the digestive system (icd9: 520-579, 787)	-0.15356	0.052009	-2.95254	0.003152	significant
primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710-739)	-0.23854	0.06503	-3.66823	2.44E-04	significant
dscharge_dispositionOtherwise	0.536482	0.029278	18.32384	5.34E-75	significant
age_groupOlder than 60	0.197613	0.032035	6.16865	6.89E-10	significant
time_in_hospital	0.032236	0.004524	7.126127	1.03E-12	significant

Table 3: Modelling Results of Core Model with HbA1c

# Step 2: Fit a Logistic Regression model of all variables including HbA1c

---

```
# Preview the relevance of the variables with Anova
```

```
anova_core_model <- Anova(core_hb_model)
```

```
anova_core_model
```

```
A anova: 8 x 2
```

term	statistic	df	p.value	significance
HbA1c	8.630571	3	0.034628092	not significant
pat_race	7.247187	3	0.064422799	not significant
physician_specialty	32.41079	5	4.93E-06	significant
primary_diagnosis	58.14431	8	1.08E-09	significant
admission_source	10.59883	2	0.004994506	significant
dscharge_disposition	334.4275	1	1.04E-74	significant
age_group	42.14044	2	7.07E-10	significant
time_in_hospital	49.63988	1	1.85E-12	significant

Table 4: ANOVA Results of Core Model with HbA1c

# Step 3: Fit a Logistic Regression with Pairwise interactions to The Core Model – The New Core Model

```
# Removing Discharge Position and Testing for model Sensitivity with CM

# Building core_model wit interactions
core_model_with_inter_rec <- recipe( readmitted ~ ., data = prep_df ) |>
  # Remove the HbA1c Variable
  step_rm(HbA1c, gender)

# Sanity Check: Preview the core-model-recipe
df <- core_model_with_inter_rec |>
  prep(prep_df) |>
  bake(new_data = NULL)

df |> glimpse()

Rows: 69,977
Columns: 8
$ physician_specialty <fct> Other, Missing or Unknown, Missing or Unknown, Mi...
$ primary_diagnosis <fct> "Diabetes (icd9: 250.xx)", "Other", "Other", "Oth...
$ admission_source <fct> Admitted because of physician/clinical referral,...
$ dscharge_disposition <fct> Otherwise, Discharged to home, Discharged to home...
$ pat_race <fct> Caucasian, Caucasian, African American, Caucasian...
$ age_group <fct> 30 years old or younger, 30 years old or younger,...
$ time_in_hospital <dbl> 1, 3, 2, 2, 1, 3, 4, 5, 13, 12, 9, 7, 7, 10, 1, 1...
$ readmitted <fct> No, No, No, No, No, No, No, No, No, No, No, Yes, ...

core_model_with_inter_fit <- glm( readmitted ~ . + physician_specialty:primary_diagnosis + physician_specialty:admission_source +
  physician_specialty:dscharge_disposition +
  physician_specialty:pat_race + physician_specialty:age_group + physician_specialty:time_in_hospital +
  primary_diagnosis:admission_source + primary_diagnosis:dscharge_disposition + primary_diagnosis:pat_race +
  primary_diagnosis:age_group + primary_diagnosis:time_in_hospital + admission_source:dscharge_disposition +
  admission_source:pat_race + admission_source:age_group + admission_source:time_in_hospital +
  dscharge_disposition:pat_race + dscharge_disposition:age_group + dscharge_disposition:time_in_hospital +
  pat_race:age_group + pat_race:time_in_hospital + age_group:time_in_hospital,
  data = df, family = binomial)

tidy(core_model_with_inter_fit)
```

term	estimate	std.error	statistic	p.value	significanc e
(Intercept)	-3.26396	0.22779	-14.3288	1.45E-46	significant
physician_specialtyFamily/GeneralPractice	0.86886 5	0.28291 3	3.07113 5	0.00213 2	significant
physician_specialtyMissing or Unknown	0.62319 8	0.23436 1	2.65914 2	0.00783 4	significant
time_in_hospital	0.09915 9	0.02484 6	3.99088 6	6.58E-05	significant
physician_specialtyMissing or Unknown:primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.82802 7	0.31944	-2.59204	0.00954 1	significant
physician_specialtyOther:primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.91205 4	0.33325 4	-2.73681	0.00620 4	significant
physician_specialtySurgery:admission_sourceAdmitted from Emergency Room	-0.61316 4	0.22929 4	-2.67412	0.00749 2	significant
physician_specialtyOther:dscharge_dispositionOtherwise	0.47021 7	0.17097	2.75029 6	0.00595 4	significant
physician_specialtySurgery:dscharge_dispositionOtherwise	0.64190 1	0.20466 7	3.13631 4	0.00171 1	significant
physician_specialtyOther:age_group30 years old or younger	-2.51327 2	0.94062	-2.67192	0.00754 2	significant
physician_specialtySurgery:time_in_hospital	-0.09377 6	0.03021	-3.10323 4	0.00191 4	significant
primary_diagnosisOther:admission_sourceAdmitted from Emergency Room	-0.3047 9	0.09586	-3.1783	0.00148 1	significant
primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710-739):dscharge_dispositionOtherwise	0.48983 3	0.16117 9	3.03907 2	0.00237 3	significant
primary_diagnosisOther:dscharge_dispositionOtherwise	0.25639 9	0.08782 8	2.91934 1	0.00350 8	significant
primary_diagnosisDiseases of the genitourinary system (icd9: 580-629, 788):time_in_hospital	0.08188 8	0.02201 8	3.71904 4	0.0002	significant
primary_diagnosisOther:time_in_hospital	-0.03535 3	0.01365	-2.58897 6	0.00962 6	significant
admission_sourceOtherwise:age_groupOlder than 60	-0.30321	0.10381	-2.92081	0.00349 1	significant
dscharge_dispositionOtherwise:pat_raceOther	0.54436 3	0.16419 4	3.31536 4	0.00091 5	significant
dscharge_dispositionOtherwise:time_in_hospital	-0.02564 3	0.00951	-2.69559	0.00702 6	significant



# Step 3: Fit a Logistic Regression with Pairwise interactions to The Core Model – The New Core Model

```
# Estimating the Anova Values
anova_core_model_with_inter_fit <- Anova(core_model_with_inter_fit)

anova_core_model_with_inter_fit
```

Term	statistic	df	p.value	significance
admission_source	8.215	2	0.016	not significant
pat_race	7.191	3	0.066	not significant
physician_specialty:admission_source	14.021	10	0.172	not significant
physician_specialty:pat_race	20.254	15	0.162	not significant
physician_specialty:time_in_hospital*	12.757	5	0.026	not significant
primary_diagnosis:admission_source	29.696	16	0.020	not significant
primary_diagnosis:pat_race	32.704	24	0.110	not significant
primary_diagnosis:age_group	25.999	16	0.054	not significant
admission_source:dscharge_disposition	0.456	2	0.796	not significant
admission_source:pat_race	3.968	6	0.681	not significant
admission_source:age_group	10.702	4	0.030	not significant
admission_source:time_in_hospital	0.738	2	0.691	not significant
dscharge_disposition:age_group	1.207	2	0.547	not significant
pat_race:age_group	6.094	6	0.413	not significant
pat_race:time_in_hospital	0.350	3	0.950	not significant
age_group:time_in_hospital	0.296	2	0.862	not significant
physician_specialty	25.348	5	0.000	significant
primary_diagnosis	59.660	8	0.000	significant
dscharge_disposition	337.861	1	0.000	significant
age_group	38.688	2	0.000	significant
time_in_hospital	40.621	1	0.000	significant
physician_specialty:primary_diagnosis	69.864	40	0.002	significant
physician_specialty:dscharge_disposition*	16.191	5	0.006	significant
physician_specialty:age_group	28.323	10	0.002	significant
primary_diagnosis:dscharge_disposition*	25.144	8	0.001	significant
primary_diagnosis:time_in_hospital*	39.668	8	0.000	significant
dscharge_disposition:pat_race*	13.683	3	0.003	significant
dscharge_disposition:time_in_hospital*	7.226	1	0.007	significant

Table 6: ANOVA Results for the Core Model with Pairwise Interactions. The \* shows interactions that were significant according to the authors.

# Step 4: Fit a Logistic Regression with Pairwise interactions, including HbA1c, to the new Core Model

```
# Create the base model
core_model_with_hb_interactions <- recipe( readmitted ~ ., data = prep_df ) |>
  # Remove the HbA1c Variable
  step_rm(gender)

# The core_model-base df
cm_hb_df <- core_model_with_hb_interactions |>
  # Prep
  prep(prep_df) |>
  # tidy to show the interactions
  bake(new_data = NULL)

cm_hb_df |> glimpse()

Rows: 69,977
Columns: 9
$ physician_specialty <fct> Other, Missing or Unknown, Missing or Unknown, Mi...
$ primary_diagnosis <fct> "Diabetes (icd9: 250.xx)", "Other", "Other", "Oth...
$ admission_source <fct> Admitted because of physician/clinical referral,...
$ dscharge_disposition <fct> Otherwise, Discharged to home, Discharged to home...
$ HbA1c <fct> No Test was performed, No Test was performed, No ...
$ pat_race <fct> Caucasian, Caucasian, African American, Caucasian...
$ age_group <fct> 30 years old or younger, 30 years old or younger,...
$ time_in_hospital <dbl> 1, 3, 2, 2, 1, 3, 4, 5, 13, 12, 9, 7, 7, 10, 1, 1...
$ readmitted <fct> No, No, No, No, No, No, No, No, No, No, No, Yes, ...

core_model_with_hb_interactions_fit <- glm( readmitted ~ . + physician_specialty:primary_diagnosis + physician_specialty:admission_source +
  physician_specialty:dscharge_disposition + physician_specialty:HbA1c +
  physician_specialty:pat_race + physician_specialty:age_group +
  physician_specialty:time_in_hospital + primary_diagnosis:admission_source +
  primary_diagnosis:dscharge_disposition + primary_diagnosis:HbA1c +
  primary_diagnosis:pat_race + primary_diagnosis:age_group +
  primary_diagnosis:time_in_hospital + admission_source:dscharge_disposition +
  admission_source:HbA1c + admission_source:pat_race +
  admission_source:age_group + admission_source:time_in_hospital +
  dscharge_disposition:HbA1c + dscharge_disposition:pat_race +
  dscharge_disposition:age_group + dscharge_disposition:time_in_hospital +
  HbA1c:pat_race + HbA1c:age_group + HbA1c:time_in_hospital +
  pat_race:age_group + pat_race:time_in_hospital + age_group:time_in_hospital,
  # Insert the new df with the family
  data = cm_hb_df, family = binomial
)
```

term	estimate	std. error	statistic	p.value	significance
(Intercept)	-3.260	0.229	-14.253	0.000	significant
admission_sourceOtherwise:age_groupOlder than 60	-0.290	0.104	-2.775	0.006	significant
dscharge_dispositionOtherwise:pat_raceOther	0.549	0.165	3.334	0.001	significant
dscharge_dispositionOtherwise:time_in_hospital	-0.027	0.010	-2.783	0.005	significant
physician_specialtyFamily/GeneralPractice	0.911	0.284	3.206	0.001	significant
physician_specialtyMissing or Unknown	0.609	0.235	2.593	0.010	significant
physician_specialtyMissing or Unknown:primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.831	0.321	-2.589	0.010	significant
physician_specialtyMissing or Unknown:time_in_hospital	-0.060	0.023	-2.666	0.008	significant
physician_specialtyOther:dscharge_dispositionOtherwise	0.482	0.171	2.811	0.005	significant
physician_specialtyOther:primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.900	0.335	-2.689	0.007	significant
physician_specialtySurgery:admission_sourceAdmitted from Emergency Room	-0.622	0.231	-2.690	0.007	significant
physician_specialtySurgery:dscharge_dispositionOtherwise	0.668	0.205	3.251	0.001	significant
physician_specialtySurgery:time_in_hospital	-0.100	0.031	-3.242	0.001	significant
primary_diagnosisDiabetes (icd9: 250.xx):HbA1cResult was high and the diabetic medication was changed	-0.482	0.183	-2.638	0.008	significant
primary_diagnosisDiseases of the genitourinary system (icd9: 580-629, 788):time_in_hospital	0.080	0.022	3.599	0.000	significant
primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710-739):dscharge_dispositionOtherwise	0.500	0.162	3.095	0.002	significant
primary_diagnosisOther:admission_sourceAdmitted from Emergency Room	-0.299	0.096	-3.099	0.002	significant
primary_diagnosisOther:dscharge_dispositionOtherwise	0.261	0.088	2.962	0.003	significant
time_in_hospital	0.103	0.025	4.046	0.000	significant

Table 7: Logistic Regression Results showing only significant results of the pairwise interactions between all predictors in the

# Step 4: Fit a Logistic Regression with Pairwise interactions, including HbA1c, to the new Core Model

```
# Determine the anova of the core_model_with_hb_interactions
anova_core_model_with_hb_interactions_fit = Anova(core_model_with_hb_interactions_fit)

anova_core_model_with_hb_interactions_fit
```

term	statistic	df	p.value	significance
physician_specialty	25.9755	5.0000	0.0001	significant
primary_diagnosis	62.4786	8.0000	0.0000	significant
dscharge_disposition	334.8902	1.0000	0.0000	significant
age_group	35.0384	2.0000	0.0000	significant
time_in_hospital	40.5677	1.0000	0.0000	significant
physician_specialty:primary_diagnosis	69.5790	40.0000	0.0026	significant
physician_specialty:dscharge_disposition*	16.8070	5.0000	0.0049	significant
primary_diagnosis:dscharge_disposition*	25.9445	8.0000	0.0011	significant
primary_diagnosis:time_in_hospital*	37.3886	8.0000	0.0000	significant
dscharge_disposition:pat_race*	13.7221	3.0000	0.0033	significant
dscharge_disposition:time_in_hospital*	7.6991	1.0000	0.0055	significant
admission_source	8.7861	2.0000	0.0124	not significant
HbA1c	8.0978	3.0000	0.0440	not significant
pat_race	7.0695	3.0000	0.0697	not significant
physician_specialty:admission_source	14.0736	10.0000	0.1697	not significant
physician_specialty:HbA1c	23.4456	15.0000	0.0751	not significant
physician_specialty:pat_race	20.9497	15.0000	0.1384	not significant
physician_specialty:age_group*	22.7567	10.0000	0.0117	not significant
physician_specialty:time_in_hospital*	13.7533	5.0000	0.0173	not significant
primary_diagnosis:admission_source	30.6234	16.0000	0.0150	not significant
primary_diagnosis:HbA1c*	42.8032	24.0000	0.0105	not significant

primary_diagnosis:pat_race	33.9238	24.0000	0.0860	not significant
primary_diagnosis:age_group	27.2036	16.0000	0.0393	not significant
admission_source:dscharge_disposition	0.4152	2.0000	0.8125	not significant
admission_source:HbA1c	4.0181	6.0000	0.6742	not significant
admission_source:pat_race	3.8539	6.0000	0.6964	not significant
admission_source:age_group	10.2432	4.0000	0.0365	not significant
admission_source:time_in_hospital	0.6437	2.0000	0.7248	not significant
dscharge_disposition:HbA1c	0.4919	3.0000	0.9207	not significant
dscharge_disposition:age_group	1.1539	2.0000	0.5616	not significant
HbA1c:pat_race	11.8548	9.0000	0.2216	not significant
HbA1c:age_group	4.8427	6.0000	0.5641	not significant
HbA1c:time_in_hospital	2.7139	3.0000	0.4379	not significant
pat_race:age_group	7.2798	6.0000	0.2957	not significant
pat_race:time_in_hospital	0.3657	3.0000	0.9472	not significant
age_group:time_in_hospital	0.2739	2.0000	0.8720	not significant

# Step 5: The Final Model – Non-Interaction Terms

```
final_model_fit <- glm( readmitted ~ . + physician_specialty:dscharge_disposition + primary_diagnosis:dscharge_disposition + primary_diagnosis:time_in_hospital +
  dscharge_disposition:pat_race + dscharge_disposition:time_in_hospital + physician_specialty:age_group +
  physician_specialty:time_in_hospital + primary_diagnosis:HbA1c,
  # Insert the new df with the family
  data = cm_hb_df, family = binomial
)

tidy(final_model_fit)
```

	term	estimate	std.error	statistic	p.value	significance
Admission	Emergency	Reference				
	admission_sourceAdmitted Emergency Room	0.01861	0.03220	0.57791	0.56333	not significant
	admission_sourceOtherwise	-0.10353	0.04420	-2.34258	0.01915	not significant
Age	b/n 30-60	Reference				
	age_group30 years old or younger	1.84186	0.84815	2.17162	0.02988	not significant
	age_groupOlder than 60	0.26613	0.14099	1.88760	0.05908	not significant
Discharge Disposition	Home	Reference				
	dscharge_dispositionOtherwise	0.26317	0.16700	1.57586	0.11506	not significant
	Not Measured	Reference				
HbA1c	HbA1cNormal result of the test	-0.04217	0.07715	-0.54658	0.58467	not significant
	HbA1cResult was high and the diabetic medication was changed	0.14951	0.09571	1.56212	0.11826	not significant
	HbA1cResult was high and the diabetic medication was not changed	-0.00818	0.13372	-0.06120	0.95120	not significant
Patient Race	African American	Reference				
	pat_raceCaucasian	0.01624	0.04844	0.33519	0.73749	not significant
	pat_raceMissing	-0.31640	0.13406	-2.36023	0.01826	not significant
	pat_raceOther	-0.29484	0.10485	-2.81208	0.00492	significant

Physician Specialty	Cardiology	Reference				
	physician_specialtyFamily/General Practice	0.40688	0.18433	2.20732	0.02729	not significant
	physician_specialtyInternal Medicine	0.39938	0.16491	2.42177	0.01545	not significant
	physician_specialtyMissing or Unknown	0.41987	0.15006	2.79809	0.00514	significant
	physician_specialtyOther	0.29200	0.16403	1.78012	0.07506	not significant
Primary Diagnosis	Diabetes	Reference				
	primary_diagnosisA disease of the respiratory system (icd9: 460-519, 786)	-0.43067	0.08888	-4.84557	0.00000	significant
	primary_diagnosisDiabetes (icd9: 250.xx)	-0.09439	0.10488	-0.89992	0.36816	not significant
	primary_diagnosisDiseases of the digestive system (icd9: 520-579, 787)	-0.15945	0.09926	-1.60637	0.10819	not significant
	primary_diagnosisDiseases of the genitourinary system (icd9: 580-629, 788)	-0.38115	0.12736	-2.99274	0.00276	significant
Time	primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710-739)	-0.82423	0.15812	-5.21274	0.00000	significant
	primary_diagnosisInjury and poisoning (icd9: 800-999)	-0.10323	0.11953	-0.86365	0.38778	not significant
	primary_diagnosisNeoplasms (icd9: 140-239)	0.05402	0.14402	0.37508	0.70760	not significant
	primary_diagnosisOther	-0.07169	0.07971	-0.89932	0.36848	not significant
	time_in_hospital	0.09456	0.02038	4.63870	0.00000	significant

Table 9: Coefficient for Non-Interaction Terms in the Final Model

# Step 5: The Final Model – Interaction Terms

	term	estimate	std.error	statistic	p.value	significance
Physician Specialty: Discharge Disposition	physician_specialtyFamily/GeneralPractice:dscharge_dispositionOtherwise	0.3173	0.1802	1.7606	0.0783	not significant
	physician_specialtyInternal Medicine:dscharge_dispositionOtherwise	0.1888	0.1648	1.1452	0.2521	not significant
	physician_specialtyMissing or Unknown:dscharge_dispositionOtherwise	0.2336	0.1541	1.5159	0.1295	not significant
	physician_specialtyOther:dscharge_dispositionOtherwise	0.3794	0.1660	2.2854	0.0223	not significant
Primary Diagnosis: Discharge Disposition	primary_diagnosisA disease of the respiratory system (icd9: 460â€³519, 786):dscharge_dispositionOtherwise	0.1306	0.0976	1.3384	0.1808	not significant
	primary_diagnosisDiabetes (icd9: 250.xx):dscharge_dispositionOtherwise	0.0246	0.1112	0.2214	0.8248	not significant
	primary_diagnosisDiseases of the digestive system (icd9: 520â€³579, 787):dscharge_dispositionOtherwise	0.0540	0.1123	0.4808	0.6306	not significant
	primary_diagnosisDiseases of the genitourinary system (icd9: 580â€³629, 788):dscharge_dispositionOtherwise	-0.1534	0.1353	-1.1341	0.2567	not significant
	primary_diagnosisInjury and poisoning (icd9: 800â€³999):dscharge_dispositionOtherwise	0.3007	0.1212	2.4806	0.0131	not significant
	primary_diagnosisNeoplasms (icd9: 140â€³239):dscharge_dispositionOtherwise	-0.1324	0.1572	-0.8423	0.3996	not significant
Primary Diagnosis Time	primary_diagnosisA disease of the respiratory system (icd9: 460â€³519, 786):time_in_hospital	0.0269	0.0154	1.7504	0.0801	not significant
	primary_diagnosisDiabetes (icd9: 250.xx):time_in_hospital	0.0324	0.0169	1.9120	0.0559	not significant
	primary_diagnosisDiseases of the digestive system (icd9: 520â€³579, 787):time_in_hospital	-0.0011	0.0181	-0.0584	0.9534	not significant
	primary_diagnosisDiseases of the musculoskeletal system and connective tissue (icd9: 710â€³739):time_in_hospital	0.0553	0.0251	2.2079	0.0273	not significant
	primary_diagnosisInjury and poisoning (icd9: 800â€³999):time_in_hospital	-0.0091	0.0189	-0.4828	0.6292	not significant
	primary_diagnosisNeoplasms (icd9: 140â€³239):time_in_hospital	-0.0142	0.0235	-0.6038	0.5460	not significant
	primary_diagnosisOther:time_in_hospital	-0.0274	0.0134	-2.0361	0.0417	not significant
Discharge Disposition : Race	dscharge_dispositionOtherwise:pat_raceCaucasian	0.0158	0.0719	0.2201	0.8258	not significant
	dscharge_dispositionOtherwise:pat_raceMissing	0.2874	0.1884	1.5255	0.1271	not significant

Physician Specialty: Age	physician_specialtyFamily/GeneralPractice:age_group30 years old or younger	-2.4875	0.9949	-2.5002	0.0124	not significant
	physician_specialtyInternal Medicine:age_group30 years old or younger	-2.0117	0.9022	-2.2296	0.0258	not significant
	physician_specialtyMissing or Unknown:age_group30 years old or younger	-1.4577	0.8588	-1.6974	0.0896	not significant
	physician_specialtySurgery:age_group30 years old or younger	-3.1779	1.3247	-2.3990	0.0164	not significant
	physician_specialtyFamily/GeneralPractice:age_groupOlder than 60	0.0616	0.1805	0.3415	0.7328	not significant
	physician_specialtyInternal Medicine:age_groupOlder than 60	-0.0160	0.1621	-0.0987	0.9214	not significant
	physician_specialtyMissing or Unknown:age_groupOlder than 60	-0.0972	0.1479	-0.6570	0.5111	not significant
	physician_specialtyOther:age_groupOlder than 60	-0.1194	0.1598	-0.7471	0.4550	not significant
	physician_specialtySurgery:age_groupOlder than 60	-0.1680	0.1960	-0.8572	0.3913	not significant
Physician Specialty: Time	physician_specialtyFamily/GeneralPractice:time_in_hospital	-0.0606	0.0261	-2.3194	0.0204	not significant
	physician_specialtyInternal Medicine:time_in_hospital	-0.0355	0.0231	-1.5366	0.1244	not significant
	physician_specialtyOther:time_in_hospital	-0.0496	0.0236	-2.1012	0.0356	not significant

# Step 6: ANOVA Results for Final Model

- The ANOVA results reveal that not all the variables included have a p-value less than 0.01.
- Some interaction terms do have significance levels of 0.01.

	LR Chisq	Df	Pr(>Chisq)
physician_specialty	31.37	5.00	0.00
primary_diagnosis	61.55	8.00	0.00
admission_source	9.17	2.00	0.01
dscharge_disposition	334.23	1.00	0.00
HbA1c	8.05	3.00	0.04
pat_race	6.96	3.00	0.07
age_group	36.61	2.00	0.00
time_in_hospital	48.55	1.00	0.00
physician_specialty:dscharge_disposition	17.61	5.00	0.00
primary_diagnosis:dscharge_disposition	25.74	8.00	0.00
primary_diagnosis:time_in_hospital	33.80	8.00	0.00
dscharge_disposition:pat_race	14.14	3.00	0.00
dscharge_disposition:time_in_hospital	8.72	1.00	0.00
physician_specialty:age_group	27.30	10.00	0.00
physician_specialty:time_in_hospital	16.68	5.00	0.01
primary_diagnosis:HbA1c	44.12	24.00	0.01

Table 3: ANOVA Results for the final model

# Step 7: Performing Predictions

- Performing predictions with the final model using the entire dataset

```
prep_df$predicted_prob <- fitted(final_model_fit)
head(prepare_df)
```

A tibble: 6 × 11

gender	physician_specialty	primary_diagnosis	admission_source	dscharge_disposition	HbA1c	pat_race	age_group	time_in_hospital	readmitted	predicted_prob
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<chr>	<dbl>
Female	Other	Diabetes (icd9: 250.xx)	Admitted because of physician/clinical referral	Otherwise	No Test was performed	Caucasian	30 years old or younger	1	N	0.06414370
Female	Missing or Unknown	Other	Admitted from Emergency Room	Discharged to home	No Test was performed	Caucasian	30 years old or younger	3	N	0.09221401
Female	Missing or Unknown	Other	Admitted from Emergency Room	Discharged to home	No Test was performed	African American	30 years old or younger	2	N	0.08994223
Male	Missing or Unknown	Other	Admitted from Emergency Room	Discharged to home	No Test was performed	Caucasian	30–60 years old	2	N	0.06402737
Male	Missing or Unknown	Neoplasms (icd9: 140–239)	Admitted from Emergency Room	Discharged to home	No Test was performed	Caucasian	30–60 years old	1	N	0.07211571
Male	Missing or Unknown	A disease of the circulatory system (icd9: 390–459, 785)	Admitted because of physician/clinical referral	Discharged to home	No Test was performed	Caucasian	30–60 years old	3	N	0.07337874

```
# Checking the Distribution of Probs
summary(prepare_df$predicted_prob)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.007919	0.063585	0.082480	0.089701	0.116344	0.355551

# Step 5: The Final Model –Interaction Terms

I tested the impact of several thresholds on Accuracy, Sensitivity and Specificity.

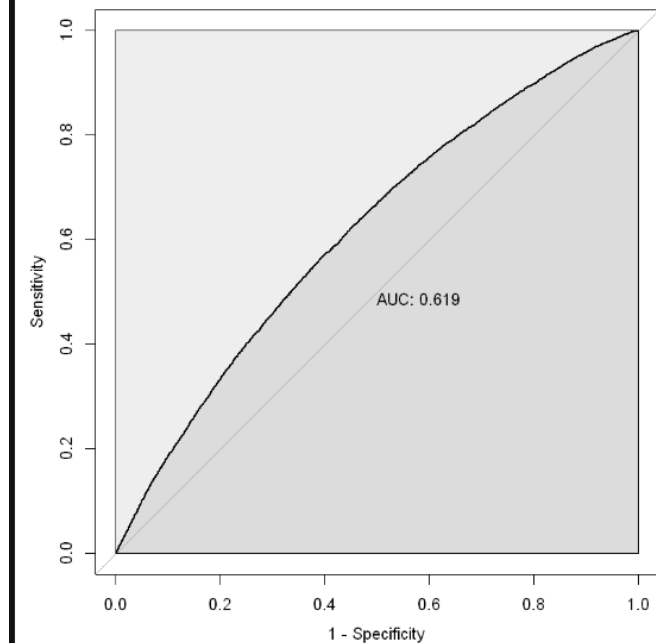
Threshold	Accuracy	Sensitivity	Specificity
0.1	0.645	0.9316	0.1281
0.15	0.8812	0.9137	0.1655
0.2	0.9092	0.9105	0.2101

## ROC Plots

```
rocplot <- roc(readmitted ~ fitted(final_model_fit),  
              smoothed = TRUE,  
              plot=T, auc.polygon=T, max.auc.polygon=T, grid=F,  
              print.auc=T, legacy.axes=T, data=cm_hb_df)
```

Setting levels: control = No, case = Yes

Setting direction: controls < cases





# Thank you