

Gout: Prevalence, Risk Factors and Statistical Analysis

Using Electronic Medical Records to
Predict Gout

Mengying Sun Nov. 5th 2015



Outline

- **Facts of Gout**
- **ARIC**
- **Statistical Analysis**
- **Discussion**



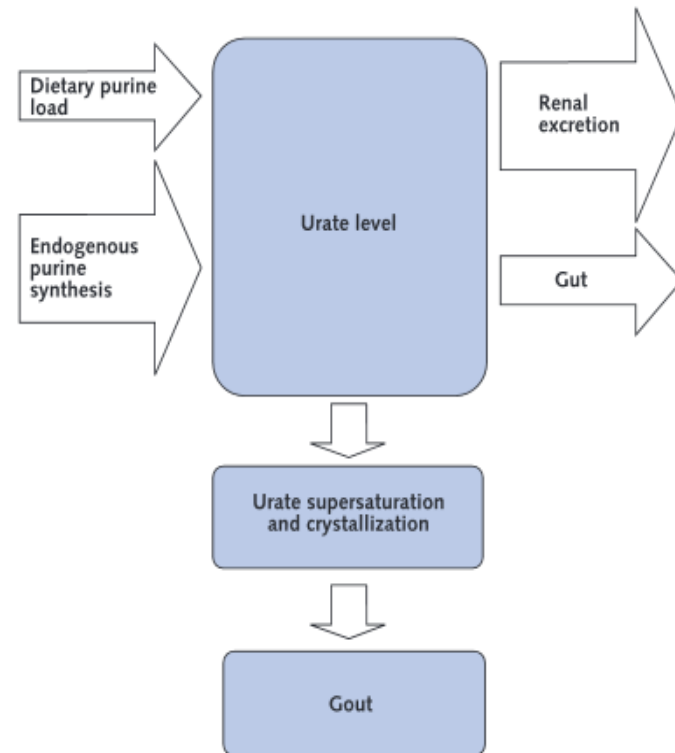
Gout: Pathogenesis

Symptom

- *A recurrent attack of acute inflammatory arthritis*
- *a red, tender, hot, swollen joint*

Cause

- *crystallization of uric acid within the joints*
- *hyperuricemia*



Choi et al., 2005 "Pathogenesis of Gout"

Gout: Prevalence and Temporal Trend

Prevalence

- *More than doubled between 1960s-1990s. (Lawrence et al., 2008)*
- *Increased by 80% from 1990 to 1999. (Wallace et al., 2004)*

2007 – 2008 [NHANES]

- *3.9% (8.3 million individual)*
- *Men: 5.9% (6.1 million)*
- *Women: 2.0% (2.2 million)*

“Remains substantial and may have increased over past 2 decades.” (Zhu et al., 2011)



Gout: Research history and findings

Before 1995	Men, White	Campion et al., 1987
	Small number of patients	Roubenoff R et al., 1991
	Single risk factor association	Hochberg et al., 1995
1995 – 2005	Overall (Sex, Race)	Arromdee et al., 2002
	Multi risk factors	Hyon K. Choi, 2005
	Diet, Comorbidities	Choi et al., 2004
	Baseline history	Ford ES et al., 2002
2005 – Current	Bigger dataset, Reviews	Zhu et al., 2012
	Longitudinal studies	McAdams-DeMarco et al., 2012
	Adjustment for risk factors	Janet W. Maynard., 2013
		Kuo et al., 2015

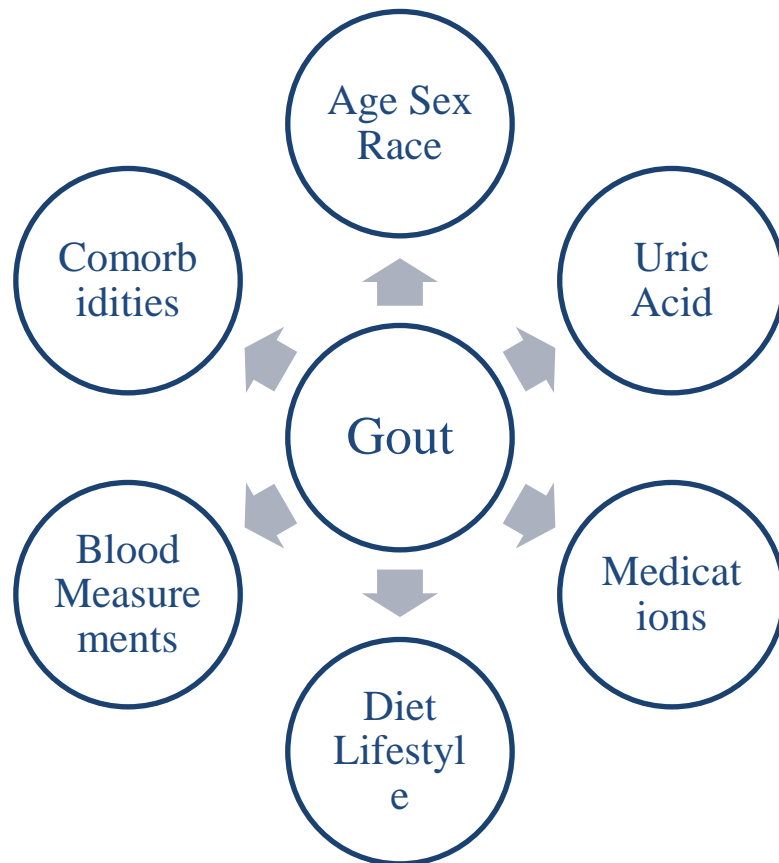


Gout: Risk Factors

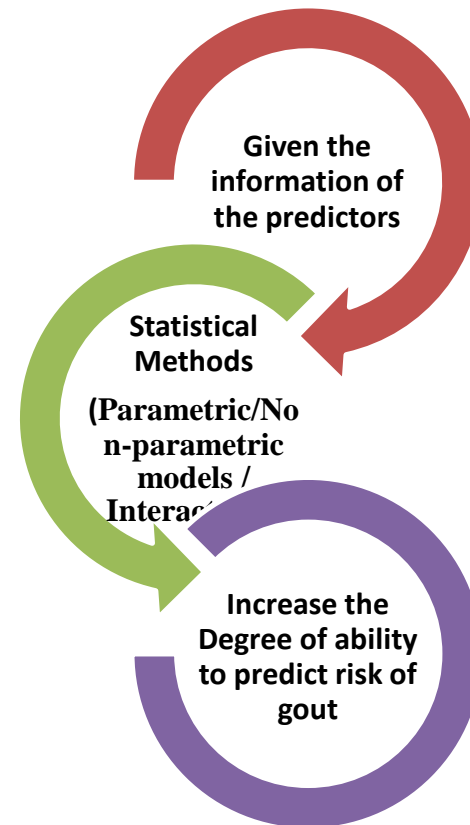
Covariates	Age
	Sex
	Race
Primary Risk Factor	Uric Acid (Serum Urate)
Secondary Risk Factor	Hypertension (Blood Pressure)
	Obesity (BMI)
	Medication (Diuretics)
	Renal Function (eGFR, Creatinine)
	Lipoprotein (HDL, LDL)
	Diet (Alcohol intake)
	Diabetes (Glucose)



Gout: Our Target?



Act as a System



ARIC

- 4 Exams, 4 Communities, every 3 years, 1987 - 1998
- 12771 obs, 363 variables
- Abundant Measurements
- Longitudinal data
- Can be used widely within area.
- Gout: Exam 4 self-report
- 'Y', 'N'

Measurements	Variable
Lipid Measurements	Cholesterol, triglycerides Lipoprotein
Chemistry Measurements	Glucose, insulin, sodium, calcium, magnesium, creatinine, urea nitrogen, uric acid, albumin, protein
Hematologic measurements (blood)	White blood cell, red blood cell, platelet counts, hematocrit, hemoglobin, blood pressure, heart rate,
Lifestyle	Diet, physical activity, Cigarette smoking, Hypertension
Diseases	Coronary heart disease, Diabetes, Intermittent claudication
Other variables	Sex, race, place, time, medication

Statistical Analysis

Outline

- Descriptive Statistics
- Generalized Linear Models
- Mixed Models



Descriptive Analysis

TABLE 1 Number of Subjects by Sex and Race

	B	W	Total
F	1266	4084	5350
	0.237	0.763	0.559
	0.640	0.538	
M	713	3511	4224
	0.169	0.831	0.441
	0.360	0.462	
Total	1979	7595	9574
	0.207	0.793	

Remarks:

- $M \sim F$
- $W > B$



Descriptive Analysis

TABLE 2 Incidence of Gout by Sex

	F	M	Total
N	5158	3855	9013
	0.572	0.428	0.941
	0.964	0.913	
Y	192	369	561
	0.342	0.658	0.059
	0.036	0.087	
Total	5350	4224	9574
	0.559	0.41	

TABLE 3 Incidence of Gout by Race

	B	W	Total
N	1819	7194	9013
	0.202	0.798	0.941
	0.919	0.947	
Y	160	401	561
	0.285	0.715	0.059
	0.081	0.053	
Total	1979	7595	9574
	0.207	0.793	

Remarks: $p(M) > p(F)$, $p(B) > p(W)$



Descriptive Analysis

TABLE 4 Mean and SD under different Gout condition

Variable	Mean(SD)		
	<i>Overall</i>	<i>With Gout</i>	<i>No Gout</i>
Uric Acid	5.94(1.49)	7.59(1.91)	5.83(1.4)
Creatinine	1.09(0.18)	1.18(0.21)	1.08(0.18)
BMI	27.48(5.14)	29.85(5.46)	27.34(5.09)
SBP	119.32(17.19)	125.45(19.15)	118.94(16.99)
Glucose	5.83(1.73)	6.28(2.09)	5.8(1.7)
HDL	1.35(0.43)	1.21(0.37)	1.36(0.44)
LDL	3.55(0.99)	3.63(1.03)	3.54(0.99)
Triglycerides	1.4(0.72)	1.75(0.87)	1.38(0.71)

Sequence of Models

Univariate Regression (Single Predictor)



Multiple Linear Regression Additive Model



**Multiple Linear Regression Additive Model
(Selected Predictors)**



**Multiple Linear Regression with Interactions
(selected predictors)**



Mixed Models (Selected Predictors)



Univariate Model

$\text{logit}(p) \sim x$

	<i>Intercept</i>	<i>Coefficient</i>	<i>P</i>
Uric Acid	-7.51	0.71	<2e-16
Creatinine	-5.75	2.63	<2e-16
BMI	-4.98	0.08	<2e-16
SBP	-5.17	0.02	<2e-16
Glucose	-3.39	0.10	1.56e-09
HDL	-1.54	-0.97	8.36e-16
LDL	-3.07	0.08	0.06
Triglycerides	-3.64	0.56	<2e-16
RaceW	-2.43	-0.46	2.64e-06
SexM	-3.29	0.94	<2e-16
Age	-4.66	0.03	5.73e-06

Remarks:

- All Significant

Multiple Linear Regression Model

logit(p)~.

	<i>Coefficient</i>	<i>Odds Ratio</i>	<i>P</i>
Intercept	-1.056e+01		<2e-16
Uric Acid	6.268e-01	1.87	<2e-16
Creatinine	3.345e-04	1.00	0.99909
BMI	2.813e-02	1.03	0.00284
SBP	7.195e-03	1.01	0.00949
Glucose	6.662e-02	1.07	0.00400
HDL	1.779e-01	1.19	0.20882
LDL	-6.909e-02	0.93	0.13840
Triglycerides	1.910e-01	1.21	<2e-16
RaceW	-2.215e-01	0.80	0.05474
SexM	3.637e-01	1.44	0.00344
Age	2.214e-02	1.02	0.00873

Remarks:

- Not significant
- Creatinine
- HDL
- LDL

Multiple Linear Regression Model (Selected Predictors)

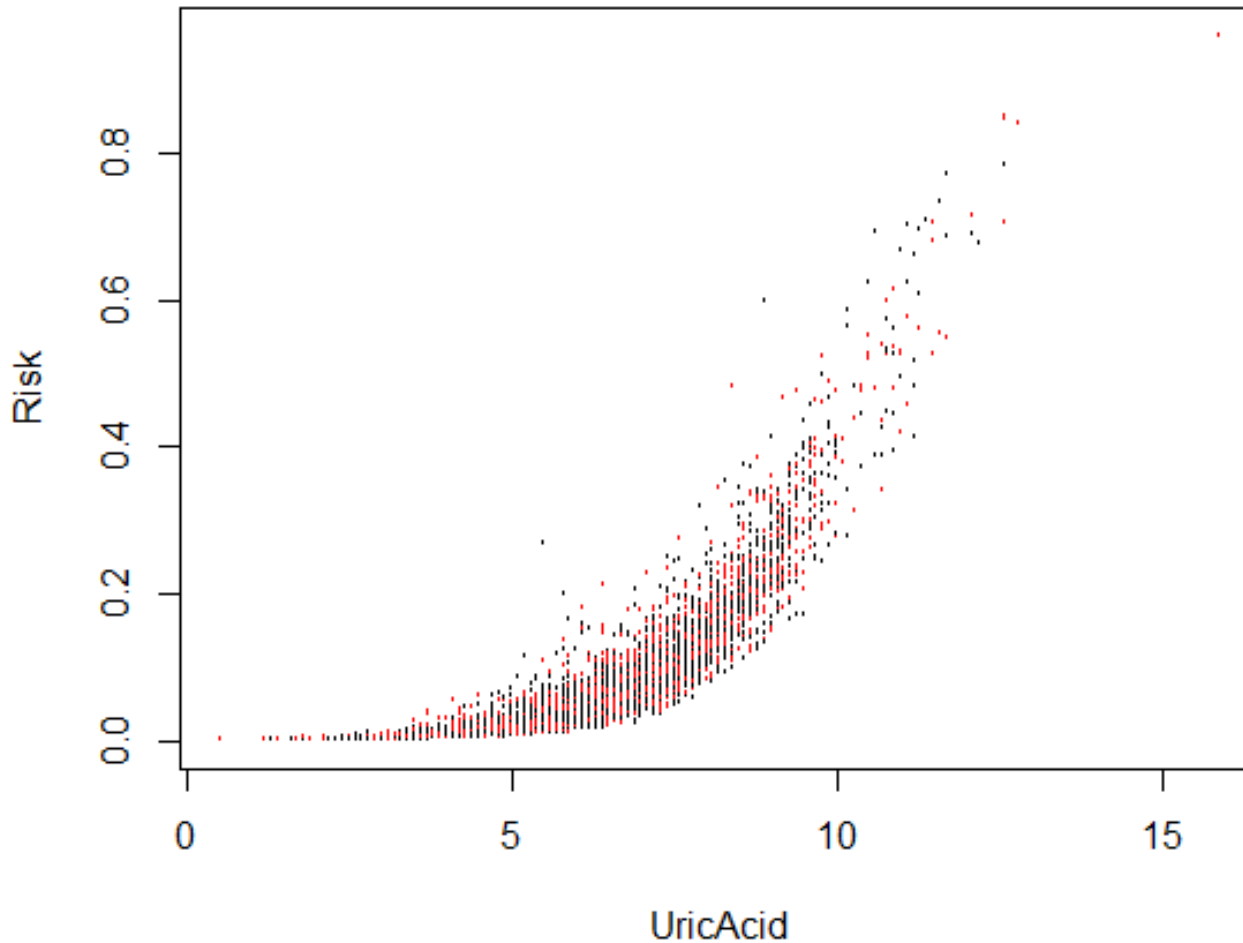
	<i>Coefficient</i>	<i>Odds Ratio</i>	<i>P</i>
Intercept	-10.3812	3.1e-05	<2e-16
Uric Acid	0.6212	1.86	<2e-16
BMI	0.0261	1.03	0.00489
SBP	0.0074	1.01	0.00732
Glucose	0.0659	1.07	0.00453
Triglycerides	0.1583	1.17	0.00867
RaceW	-0.2346	0.79	0.03798
SexM	0.3189	1.38	0.00229
Age	0.0215	1.02	0.01043

Remark:

1. Every predictor is significant
2. Numbers are boring
3. What effect may every predictor have wrt risk of developing gout
4. Plot (risk~pred)



Risk of Gout by Sex

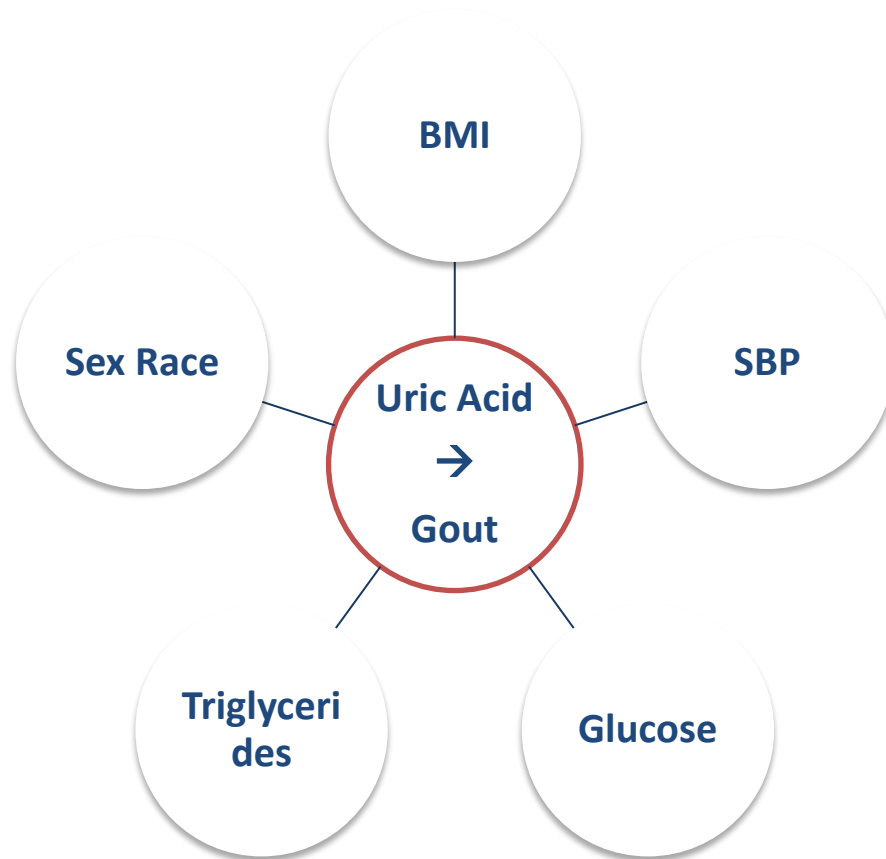


Remark:

1. Uric Acid = 7
2. Come up with Richard's question

“Why people have high level uric acid still don't get gout?”

3. Even though other factors singly don't have certain pattern affecting gout, will they affect the effect that uric acid have on gout?

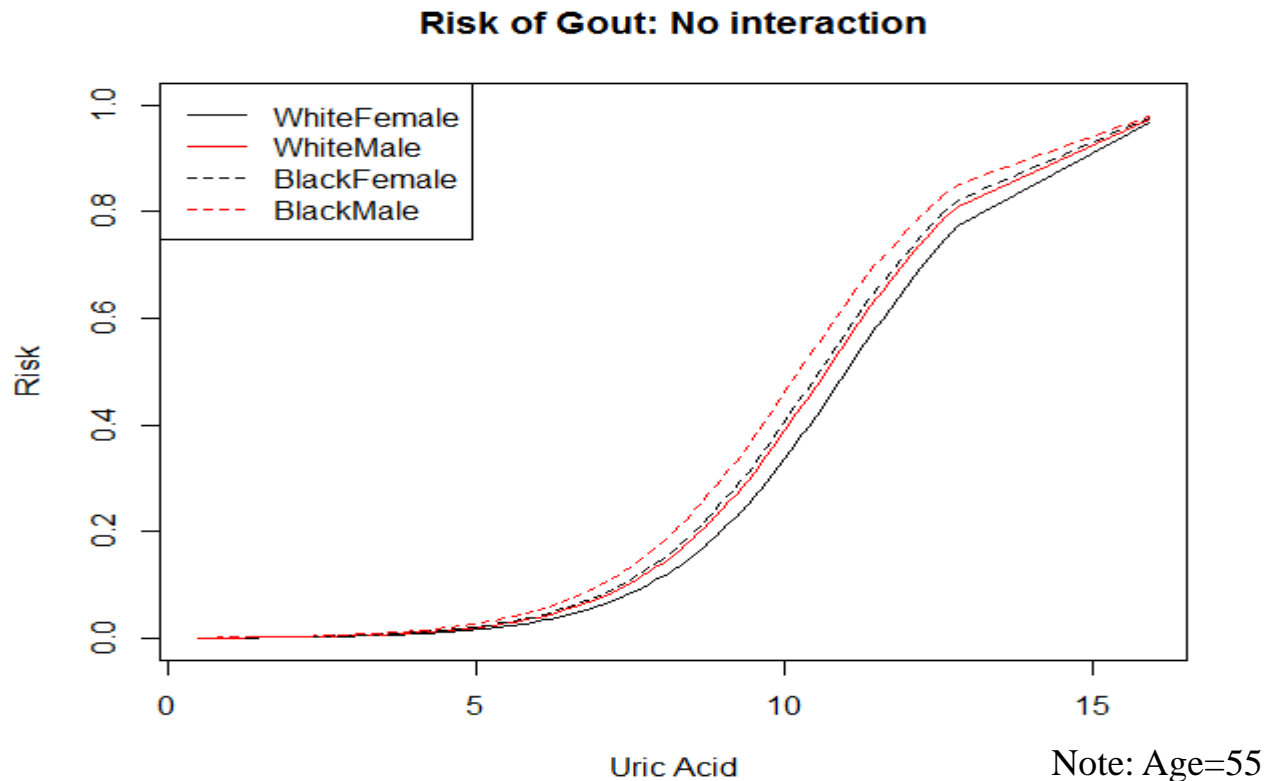


Interactions!



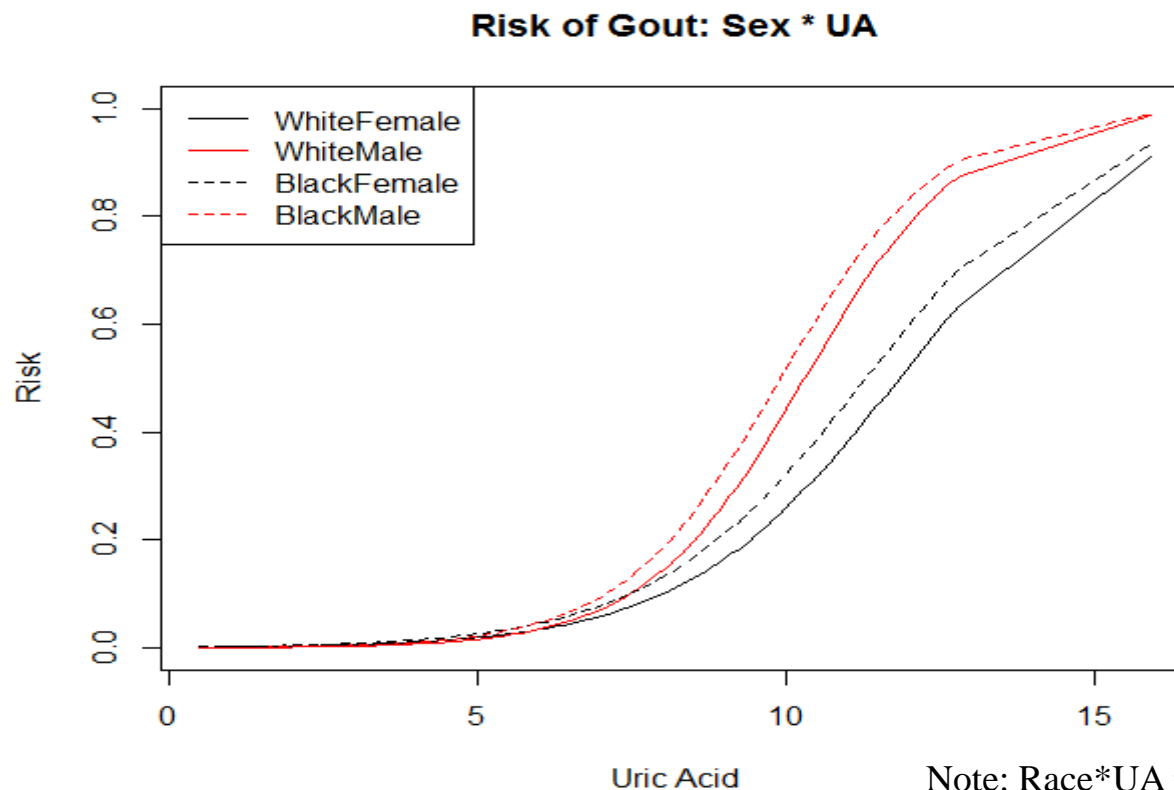
MLR with Interactions

Base Model: Gout ~ Sex+Race+Age+UricAcid



MLR with Interactions

Base Model: Gout ~ Sex+Race+Age+UricAcid+Sex*UA



Note: Race*UA is not that significant

MLR with Interactions

*Gout ~ Sex + Race + Age + UricAcid + Sex * UA*
*+ Uric Acid * CAT (Predictor)*



Thresholds used to categorize risk factors

	<i>Cut Point (N/H)</i>	<i>Overall Mean</i>
BMI	30	27.5
SBP	150	120
Glucose	7	5.8
Triglycerides	2	1.4

Remark:

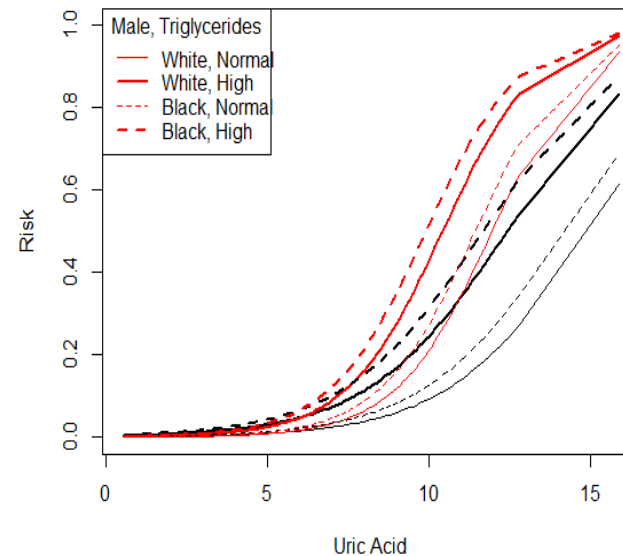
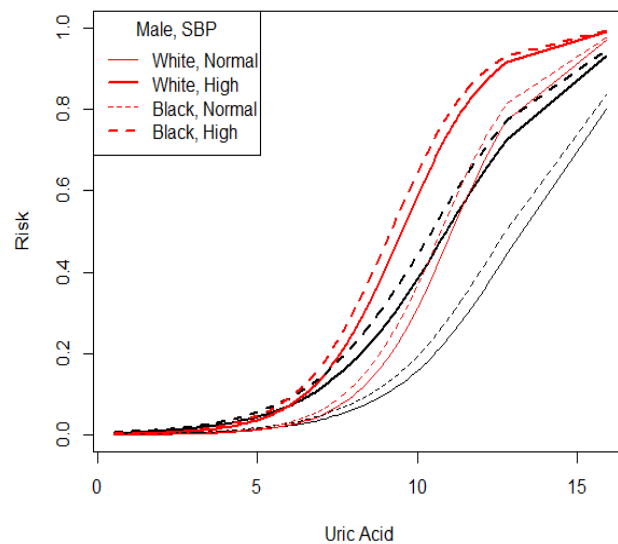
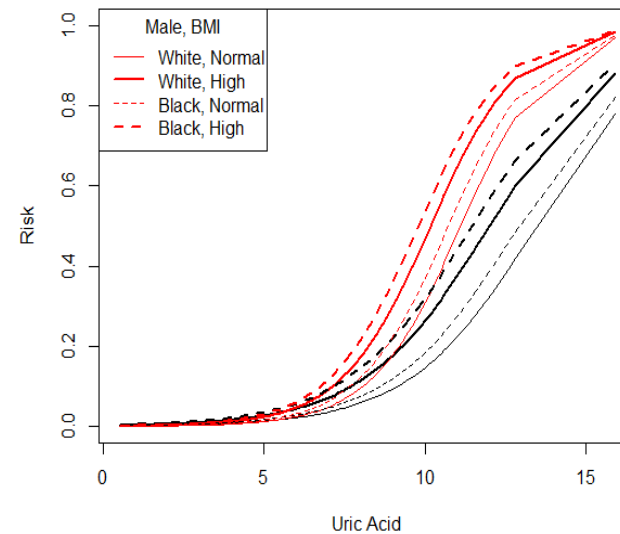
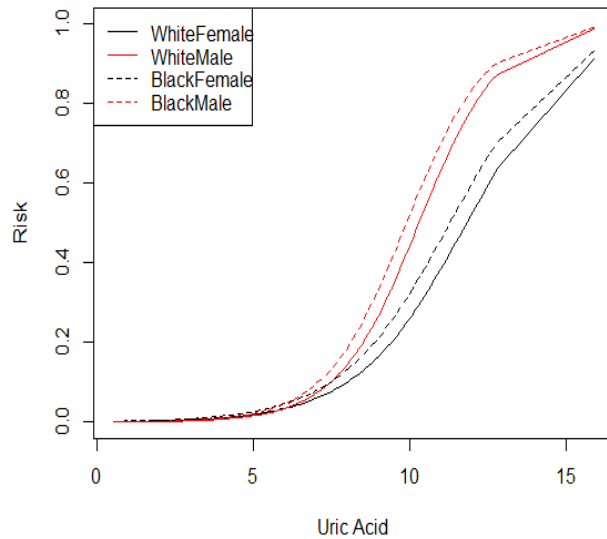
- Some using normal value
- Some higher than normal value

*Gout~Sex + Race + Age + UricAcid + Sex * UA + Uric Acid * CAT (BMI)*

*Gout~Sex + Race + Age + UricAcid + Sex * UA + Uric Acid * CAT (SBP)*

*Gout~Sex + Race + Age + UricAcid + Sex * UA + Uric Acid * CAT (Glucose)*

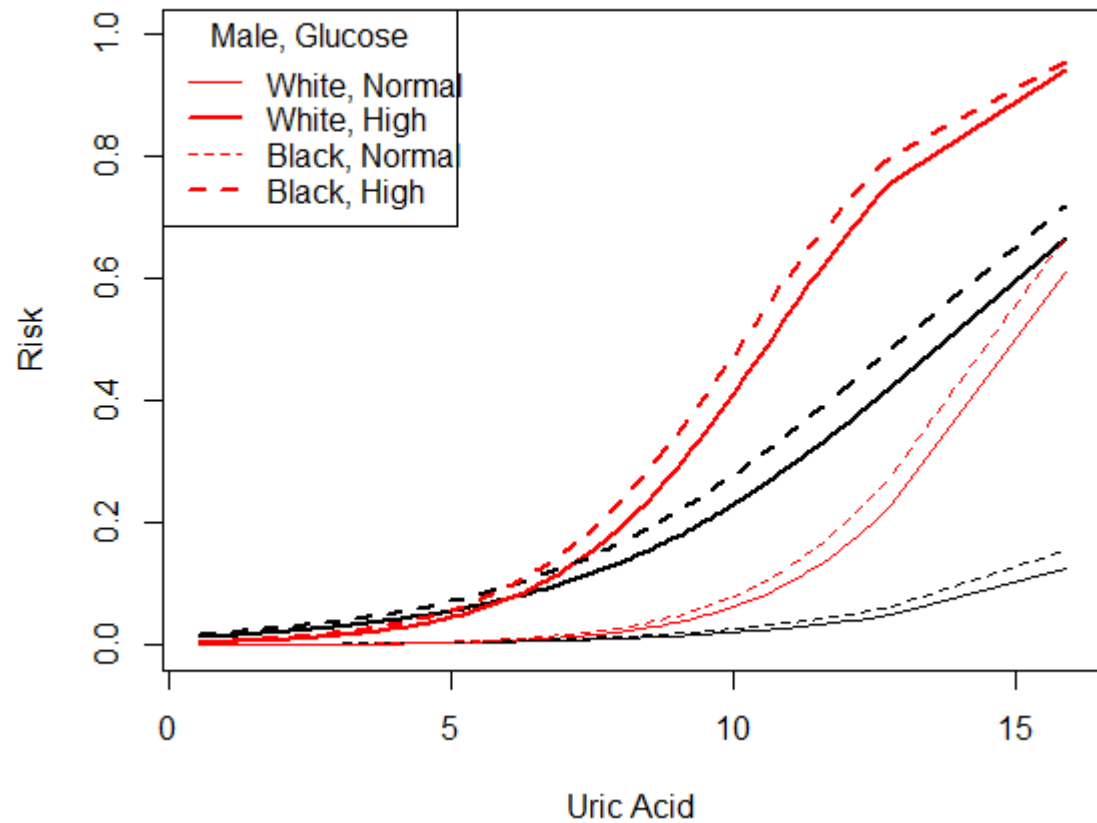
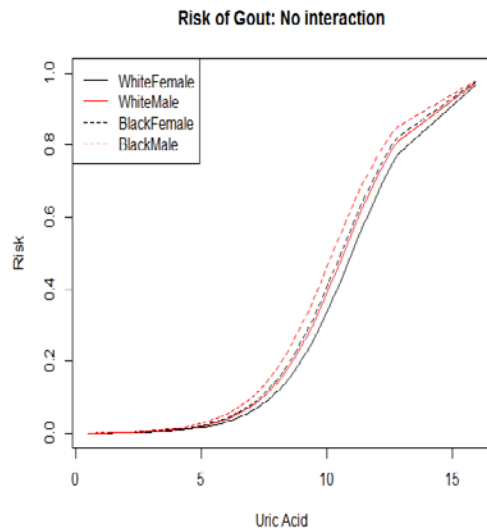
*Gout~Sex + Race + Age + UricAcid + Sex * UA + Uric Acid * CAT (Triglycerides)*



Remark:

- Color-Sex
- LineType-Race
- Lwd-Level

Risk of Gout: UA * Glucose



MLR with Interactions

Aspects we need To be careful with:

- Female Curve

Why Random effects?

- Doctors care more about Normal/High levels instead of specific numbers
- If treat interactions as fixed effects
- Explain fixed effects interactions, eg 3 variables ab,ac,bc,abc, # of parameters will increase dramatically
- Some categories will have little/0 observations -- MLE will not work



Mixed Models (Selected Predictors)

	<i>Cut Point</i>	<i>Levels</i>
BMI	30	N,H
SBP	150	N,H
Glucose	7	N,H
Triglycerides	2	N,H
Uric Acid	7	N,H
Sex	/	M,F
Race	/	B,W

- Combine all the category as one index.
- People in the same group share some dependence

Random Effect	Cat_Index
Obs 1	HHHHHMB
Obs 2	NNNNHFW
Obs 3	HNNHNFB
...	...

Mixed Models (Selected Predictors)

$$p(Y_{ij} = 1) = \theta_j$$

$$\log\left(\frac{\theta_j}{1 - \theta_j}\right) = \mu + u_j$$

$$\mathbf{u} \sim N(\mathbf{0}, \mathbf{I}\sigma_u^2)$$



Mixed Models (Selected Predictors)

Fixed effects	Coefficient	P
Intercept	-0.2341	0.288246
cat_GlucoseN	-0.4613	0.001558
cat_uaN	-1.6442	< 2e-16
cat_sbpN	-0.6863	2.23e-05
cat_TriN	-0.3440	0.003172
cat_bmiN	-0.3915	0.000383
SexM	0.5084	4.87e-06
RaceW	-0.3036	0.008335
Random Effect	Cat_Index	

	Risk	Index
1	0.02701296	N N N N N M W
2	0.21923477	N H N N H M B
3	0.03355356	N N N N H F B
4	0.03807269	N N N H N M W
5	0.05929691	H N N N N M B
6	0.16705677	N H N H N M W
7	0.02701296	N N N N N M W
8	0.11893686	N H N H N F W
9	0.01515288	N N N N N F W
10	0.02701296	N N N N N M W
11	0.01515288	N N N N N F W
12	0.14263139	N H N N N M W

Mixed Models (Selected Predictors)

Output



What else did we do?

- Principal Component Analysis
- Splines
- What else can we do? Suggestions



Thank you!

