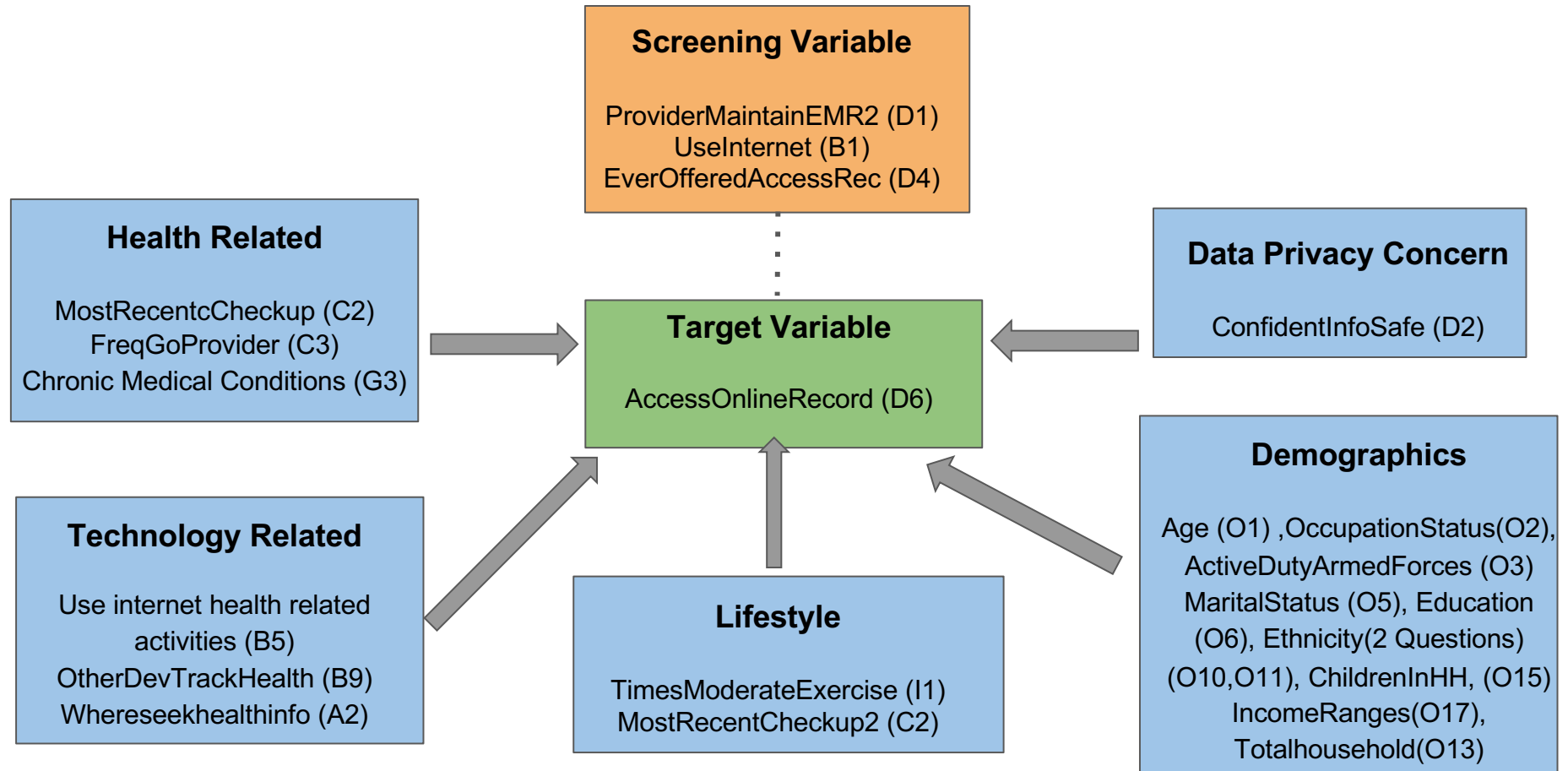# NIH Survey Trend Analysis:
<u>Increasing EMR presence</u>

Jigyasa Sachdeva
Liautaud Business School, University of Illinois at Chicago

Technology Assessment Model

Screening Variable

ProviderMaintainEMR2 (D1)
UseInternet (B1)
EverOfferedAccessRec (D4)

Health Related

MostRecentcCheckup (C2)
FreqGoProvider (C3)
Chronic Medical Conditions (G3)

Target Variable

AccessOnlineRecord (D6)

Data Privacy Concern

ConfidentInfoSafe (D2)

Technology Related

Use internet health related activities (B5)
OtherDevTrackHealth (B9)
Whereseekhealthinfo (A2)

Lifestyle

TimesModerateExercise (I1)
MostRecentCheckup2 (C2)

Demographics

Age (O1) ,OccupationStatus(O2),
ActiveDutyArmedForces (O3)
MaritalStatus (O5), Education
(O6), Ethnicity(2 Questions)
(O10,O11), ChildrenInHH, (O15)
IncomeRanges(O17),
Totalhousehold(O13)

Screening Variables

Variable name	Survey Question	Data Type	Treatment
ProviderMaintainEMR2	**D1**	Categorical	Remove people who said "No" (91), Put -9 and -5 (missing and errored values) into 3 (Don't know)
UseInternet	**B1**	Binary	(yes = 1, no =0) only 35 people use EMR that said no Remove 2
EverOfferedAccessRec	**D4**	Categorical	No Access to EMR will not use it Remove 2 and 3

Target Variable - Cleaning

AccessOnlineRecord	D6	Categorical	Remove -9, -2, -1

3 approaches:

Binary Target

2 levels:
0 (Never accessed EMR) as '0'
1,2,3,4 (Accessed EMR once or more) as '1'

Ordinal Target

3 categories:
0 (Never accessed) as '0'
1,2 (Accessed 1-5 times) as '1-5 times'
3,4 (Accessed more than 6 times) as '6+ times'

Multinomial Target

5 categories:
0,1,2,3,4 (as given)

Demographic Variables - Cleaning

Age	O1	Numerical: Converted into Categorical Base: 60+	Replace -9 with median: 55 Age category (new column): <30, 30-45, 45-60, 60+
OccupationStatus	O2	Categorical Base:1 (Employed)	1,2,3,4,5,6, 91 (Other: Clubbed -5 and -9)
ActiveDutyArmedForces	O3	Categorical Base: 0 (Never active)	1,2,3: 1 (active some point) 4+5: 0 (never active) -9: missing group
MaritalStatus	O5	Categorical Base: 6 (Single)	-5 and -9 = 0 1,2,3,4,5,6
Education	O6	Categorical Base: 1 (Less than HS)	Clump 1 and 2 into 1 (less than high school) 2-6 for rest of categories -9=0
Totalhousehold	O13	Categorical Base: 1 (1 person)	1,2,3,4+, unknown

Demographic Variables - Cleaning

Hisp_Cat	**O10**	Categorical Use only Hisp_Cat (No dependency) Base: Not Hispanic	Unknown (-9) Not Hispanic (10) Hispanic (21,22,23,24,25) Data was unbalanced (very few Hispanic) 10 = not hispanic , 21= Mexican, 22= puerto Rican, 23=Cuban, 24= other hispanic, 25= multiple hispanic, -9 = another bucket
Race_Cat2	**O11**	Categorical Use only Race_Cat Base: White	Unknown (-9) White (11) Black (12) Others (14, 16,31, 32, 33, 34, 35, 36, 37, 52, 54, -9) Data was unbalanced (very few other) 11 = white, 12 = black, 14 = american indian, 16 = multiple races, 31 = asian indian, 32= chinese, 33= Filipino, 34= Japanese, 35= Korean, 36= Vietnamese, 37 = Other Asian, 52= Guamanian or Chamorro, 54= Other pacific Islander -9 = no answer
ChildrenInHH	**O15**	Categorical Base: 0	0 (0), 1+ (1,2,3,4..8), Unknown(9)
IncomeRanges	**O17**	Categorical Base: 1	0 (unknown), 1-4 (low), 5-6 (moderate), 7-9 = (high)

Health Related Variables - Cleaning

FreqGoProvider	**C3**	Categorical Base: 0	0,1,2,3,4,5,6, -9
MedConditions_Diabetes, MedConditions_HighBP, MedConditions_HeartCondition, MedConditions_LungDiseases, MedConditions_Arthritis, MedConditions_Depression	**G3**	All binary variables clubbed into one category Base: 0	Variable IsChronic: If all answers are 'No disease', then= 0 If any 1, then= 1 -9: Unknown
MostRecentCheckUp2	**C2**	Categorical Base: 1	1, 2,3+ (3,4,5), 6 = Don't know

Technology Related Variables - Cleaning

Electronic_SelfHealthInfo, Electronic_HealthInfoSE, Electronic_buyMedicine, Electronic_LookedAssistance, Electronic_TalkDoctor, Electronic_TrackedHealthCosts, Electronic_TestResults	**B5**	Ordinal Base is 0	TechSavvy: Ordinal additive variable If all unknown: -9 For any 0: add 0, for any 1: add 1 Final categories: 0,1,2,3,4,5,6,7,-9
Whereseekhealthinfo	**A2**	Categorical Base: 'Not uses'	Unknown, uses internet, does not use internet
OtherDevTrackHealth	**B9**	Binary Base is 0	2 = 0(no) 1 = 1(yes) -9 = unknown

Security Related Variables - Cleaning

ConfidentInfoSafe	**D2**	Ordinal Reverse Code Base: 1	3 → High Confident 2 → Somewhat Confident 1 → Not Confident -9 and -5 = unknown = 0

Lifestyle Related Variables - Cleaning

TimesModerateExercise	**I1**	Categorical Base: None	None = 0 1-2 (low exercise level) = 1 3-4 (moderate) = 2 5-7 (high) = 3 Unknown = -9

Logistic Regression- Binomial

Tech Savvy: (B5)

#Tech_savvy2	1.12709	0.34964	3.224	0.00127	**
#Tech_savvy3	1.97556	0.35257	5.603	2.10e-08	***
#Tech_savvy4	3.20274	0.36461	8.784	< 2e-16	***
#Tech_savvy5	3.83389	0.38542	9.947	< 2e-16	***
#Tech_savvy6	3.72819	0.39461	9.448	< 2e-16	***
#Tech_savvy7	4.69926	0.58780	7.995	1.30e-15	***

Incomeranges: (O17)

#incomerangesHigh	0.60251	0.22407	2.689	0.00717	**
#incomerangesUnknown	0.83013	0.30182	2.750	0.00595	**

Confident into safe: (D2)

#confidentinfosafe2	0.39986	0.19307	2.071	0.03835	*
#confidentinfosafe3	0.63149	0.20319	3.108	0.00188	**

Marital status: (O5)

maritalstatus5	-1.53089	0.52561	-2.913	0.00358	**

Frequgoprovider: (C3)

#freqgoprovider4	0.71934	0.30779	2.337	0.01943	*
#freqgoprovider5	0.67197	0.30168	2.227	0.02592	*

Interpretations:

Variable	Level of the variable	Access EMR	With respect to
Tech savvy	Nature increases	More likely	Non- tech Savvy
Income Ranges	Above 75k	More likely	Below 35k
Confident about privacy	Yes	More likely	Not confident
Marital Status	Separated	Less likely	Single
Frequency to go to the provider	4-9 times	More likely	Who did not visit at all

Logistic Regression- Multinomial and Ordinal

Multinomial Attributes	Coefficient				p-Values				Ordinal Coefficients	p-value
	1	2	3	4	1	2	3	4		
(Intercept)	-3.067	-6.168	-15.273	-7.78	0	0	0.031	0		
occupationstatus5	-0.154	0.626	0.424	0.596	0.505	0.021	0.314	0.153		
activedutyarmedforces1	0.058	0.548	0.602	0.297	0.817	0.045	0.162	0.527		
maritalstatus3	-0.232	-0.326	-1.064	-0.42	0.382	0.275	0.033	0.37		
maritalstatus5	-1.27	-1.502	-8.065	-8.185	0.031	0.037	0.527	0.742	Negative	***
race_cat2Unknown	0.477	0.176	1.738	0.061	0.322	0.759	0.039	0.952		
incomerangesHigh	0.61	0.577	0.692	0.71	0.016	0.046	0.122	0.13	Positive	**
incomerangesUnknown	0.78	0.718	1.546	1.025	0.021	0.072	0.007	0.112	Positive	***
freqgoprovider4	0.376	1.142	5.588	0.78	0.266	0.006	0.3	0.241	Positive	**
freqgoprovider5	-0.008	1.046	6.243	1.652	0.981	0.012	0.246	0.01	Positive	***
freqgoprovider6	-0.325	0.5	5.269	1.715	0.387	0.262	0.329	0.008	Positive	***
otherdevtrackhealth1	0.065	0.231	0.432	0.936	0.658	0.169	0.096	0.001	Positive	**
confidentinfosafe2	0.523	0.426	-0.258	0.19	0.018	0.086	0.468	0.644		
confidentinfosafe3	0.716	0.556	0.12	1.102	0.002	0.032	0.741	0.007	Positive	**
mostrecentcheckup22	-0.496	-0.024	0.172	-0.133	0.021	0.921	0.646	0.74		
totalhousehold2	-0.05	-0.568	0.251	-0.593	0.828	0.034	0.546	0.169		
totalhousehold4+	-0.358	-0.882	-0.961	-0.459	0.294	0.03	0.139	0.456		
agecat30-45	-0.034	0.906	0.772	0.786	0.924	0.032	0.226	0.23		
agecat45-60	0.231	0.893	0.883	0.907	0.441	0.013	0.096	0.099	Positive	*
agecatless than 30	-0.131	1.065	0.939	0.936	0.732	0.021	0.179	0.189		
ischronic1	-0.317	0.265	0.792	0.044	0.058	0.195	0.026	0.889		
Tech_savvy2	1.152	1.19	3.187	1.006	0.005	0.049	0.466	0.411	Positive	**
Tech_savvy3	2.031	1.881	4.785	1.487	0	0.002	0.271	0.21	Positive	***
Tech_savvy4	2.874	3.654	6.418	3.595	0	0	0.139	0.002	Positive	***
Tech_savvy5	3.511	4.066	7.61	4.328	0	0	0.08	0	Positive	***
Tech_savvy6	3.056	4.278	7.464	4.736	0	0	0.086	0	Positive	***
Tech_savvy7	4.138	5	8.421	5.966	0	0	0.054	0	Positive	***

Yellow: Significant for only multinomial

Light blue: Significant for only multinomial and ordinal

Dark blue: Significant in all 3 models

Extra Findings From Models

Multinomial and Ordinal

Variable	Level of the variable	Access EMR	With respect to
Age category	45-60	More likely	>60
Having a device to track health	Yes	More likely	No
Frequency to go to the provider	More than 9 times	More likely	Who do not visit at all

Multinomial Only

Variable	Level of the variable	Access EMR	With respect to	Target Variable
Is chronic	Yes	Less likely	Not chronic	Access it 1-2 times than never
	Yes	More likely	Not chronic	Access it 6-9 times than never
Age category	Less than 30	More likely	More than 60	Access it 3-5 times than never
	30-45	More likely	More than 60	Access it 3-5 times than never
Total household	4+	Less likely	Than 1	Access it 3-5 times than never
	2	Less likely	Than 1	Access it 3-5 times than never

Most recent checkup	1-2 years ago	Less likely	Within the last year	Access it 1-2 times than never
Marital Status	Divorced	Less likely	Single	Access it 6-9 times than never
Active Duty Armed Forces	Yes	More likely	No	Access it 3-5 times than never
Occupational Status	Retired	Less likely	Employed	Access it 6-9 times than never

**In this case, multinomial model is better for capturing differences in levels of categorical variables as it covers higher variance from the given data due to granular level distribution of target variable.

Audience to be targeted

- Have high income and access to HMO insurance plans
- Are single as compared to separated or divorced
- Go to their provider frequently: 4+ times in a year
- Are confident about data privacy
- Are Tech Savvy ($\geq 2/7$)
- Have regular check-ups (most recent within last year)
- Have a chronic disease
- Use Other Smart Devices to track health
- Are young: Age between 30-45
- Never have been active in military

Key

Blue = All 3 models supported
Orange = 1/2 models supported