Trans Health in Cancer Genetic Counseling

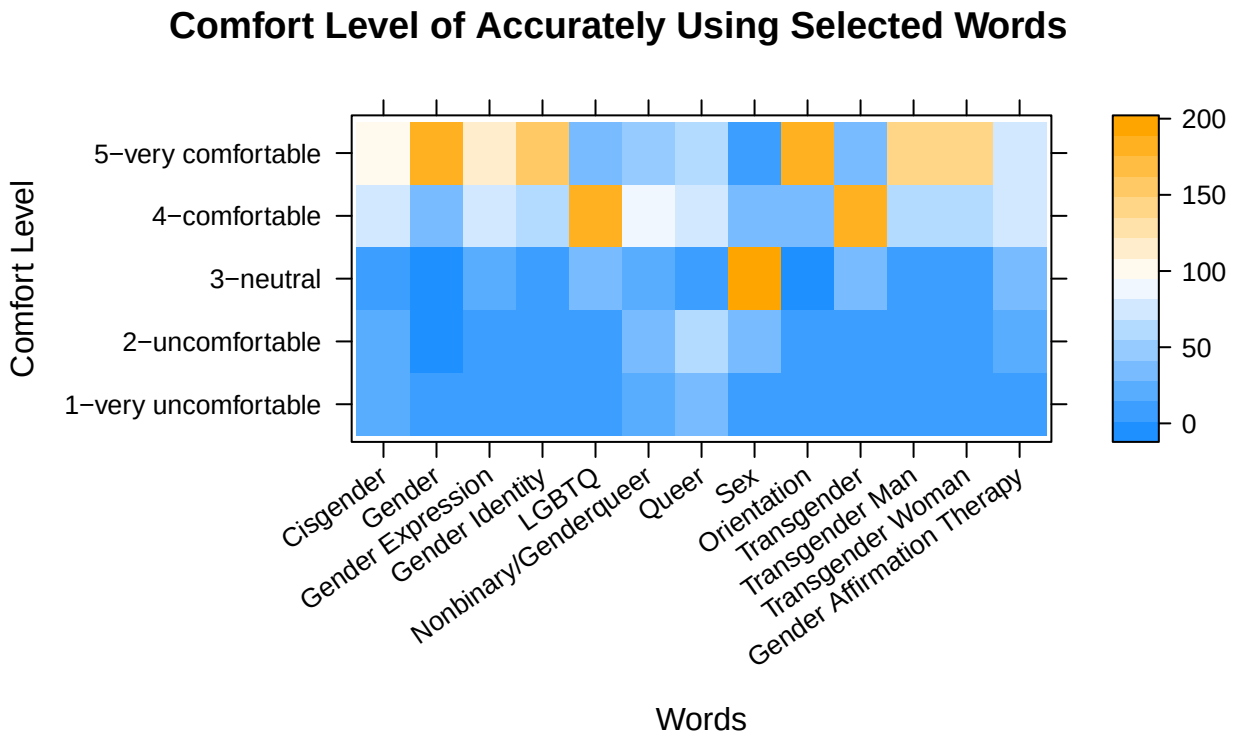
Xuechun Liu, Liwen Zhang, Kailun Huang
PhD mentor: David Reynolds
2/10/2018

Introduction and Summary Findings

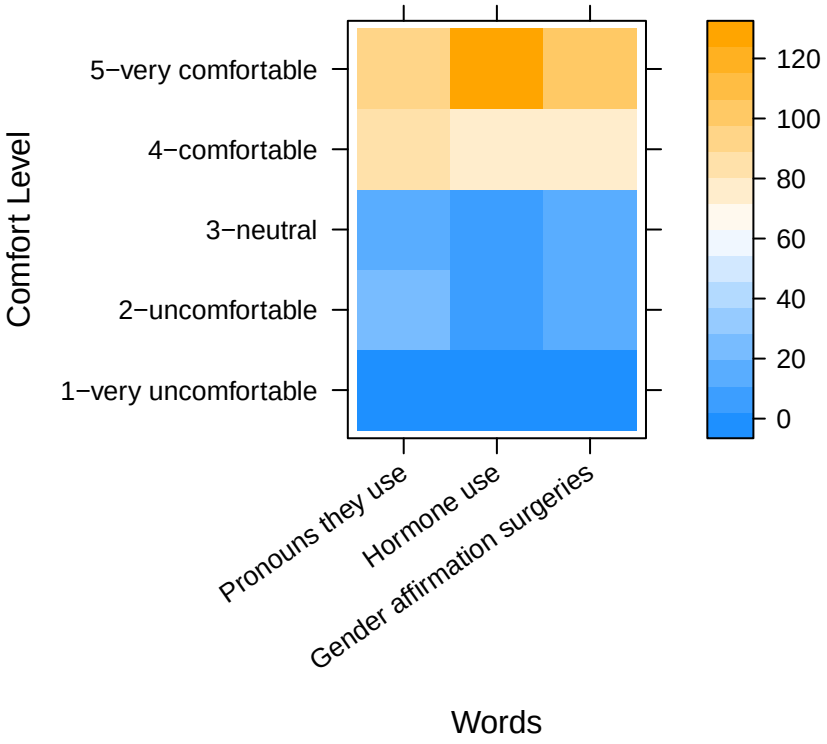
The client, Tala Berro, is a second-year genetic counseling MA student at Boston University. This project is her MA thesis and concerns cancer genetic counselors' anticipated behaviors when counseling transgender patients. Study participation involves a brief web-based survey and the participants are generated from genetic counselors in the Cancer Special Interest Group. Surveys were sent by email to all genetic counselors in this group and 257 responses were received. The client's interest lies primarily in the responses received on four case study questions that pertain to clinical situations involving transgender cancer patients. Each of these cases has a set of objectively correct answers. The client is interested in understanding how the performance of the respondents may depend on factors related to their experience or demographic profile. We measure performance by the Jaccard Index, described below. We fit linear regression models for each case study and in aggregate (performance across all case studies) and find that the factors measured do not exert significant effects on performance. That is, the performance of the cancer genetic counselors on the four cases provided in the survey do not appear to depend on their experience or demographics.

Exploratory Data Analysis

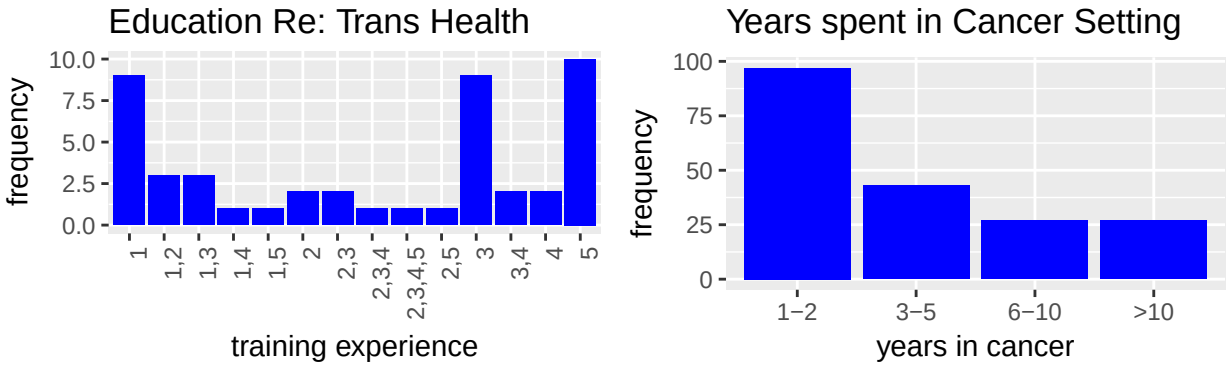
We begin our exploratory analysis below by visualizing heat maps of the response frequencies associated with comfort levels pertaining to transgender terminology. We see that the majority of respondents feel comfortable using terminology related to transgender individuals.



Comfort Level of Asking a Trans/Nonbinary Patient Selected Words

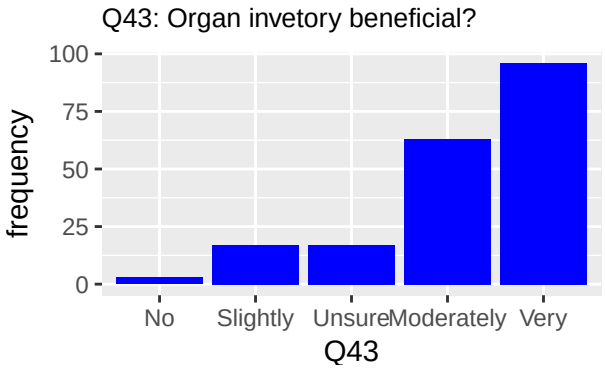


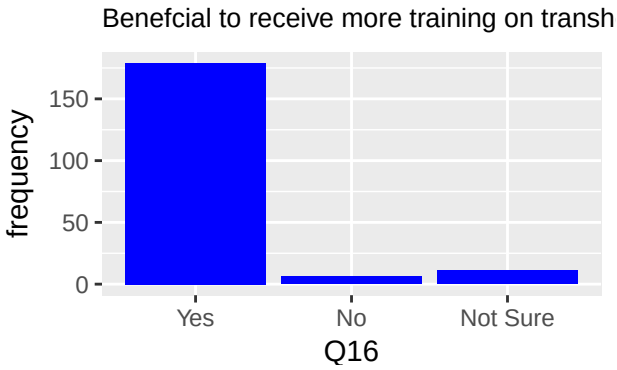
Next, we explore the distributions of responses pertaining broadly to the experience of respondents. In particular, in the plots below, we visualize the number of respondents who have education related to transgender patients and also the number of years of experience working in a cancer setting. Note that these responses do not include empty replies.



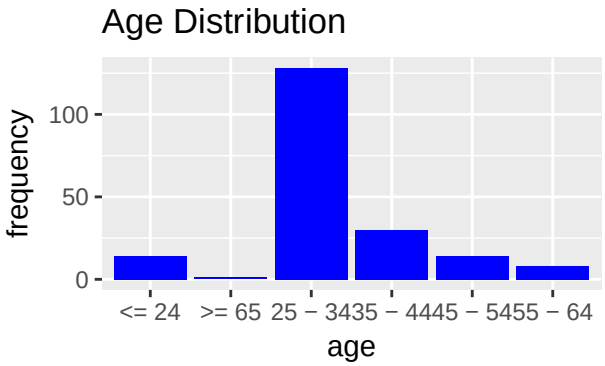
Note that on the left hand figure, response 1 corresponds to a genetic counseling program; response 2 corresponds to a conference; response 3 corresponds to workshops; response 4 corresponds with education specific to trans health in cancer; response 5 is other. This plot shows that most participants only go to a Genetic Counseling Graduate Program or workshops. Seldom do the respondents have multiple training experiences. Additionally, the plot on the years spent in a cancer setting indicate that the most frequent experience level among the respondents is 1-2 years.

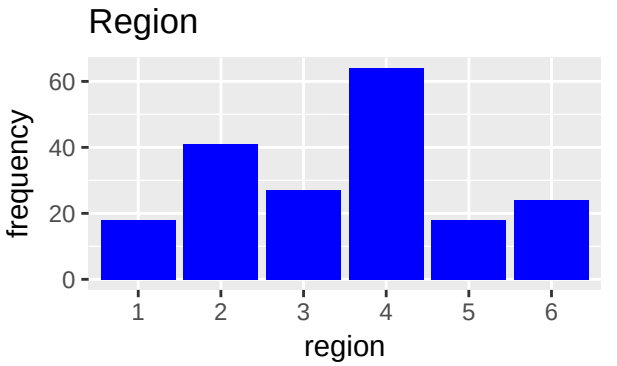
Next, we visualize the desire of the respondents for more information and training related to transgender patients. Question 43 indicates whether the counselors would find an inventory of each patients organs to be beneficial. We see that a majority of respondents to indeed desire more training related to transgender patients.





Finally, we explore the broad demographic profiles of the respondents with respect to their age and region.





Note that the responses for region are:

CT,MA,ME,NH,RI,VT,CN Maritime Provinces (1)
DC,DE,MD,NJ,NY,PA,VA,WV,PR,VI,Quebec (2)
AL,FL,GA,KY,LA,MS,NC,SC,TN (3)
AR,IA,IL,IN,KS,MI,MN,MO,ND,NE,OH,OK,SD,WI,Ontario (4)
AZ,CO,MT,NM,TX,UT,WY,Alberta, Manitoba, Sask. (5)
AK,CA,HI,ID,NV,OR,WA,British Columbia (6)

We see that the most frequent region is the midwest and the most frequent age range is 25-34.

Case Study Analysis

In the analysis of the case studies, we use the Jaccard index, which measures similarity between finite sample sets, as our dependent variable. We calculate the Jaccard index using the size of the intersection of participants' responses and correct answers, divided by the cardinality of their union. Other variables, such as age, gender, number of years spent working in a cancer setting, serve as covariates to fit models.

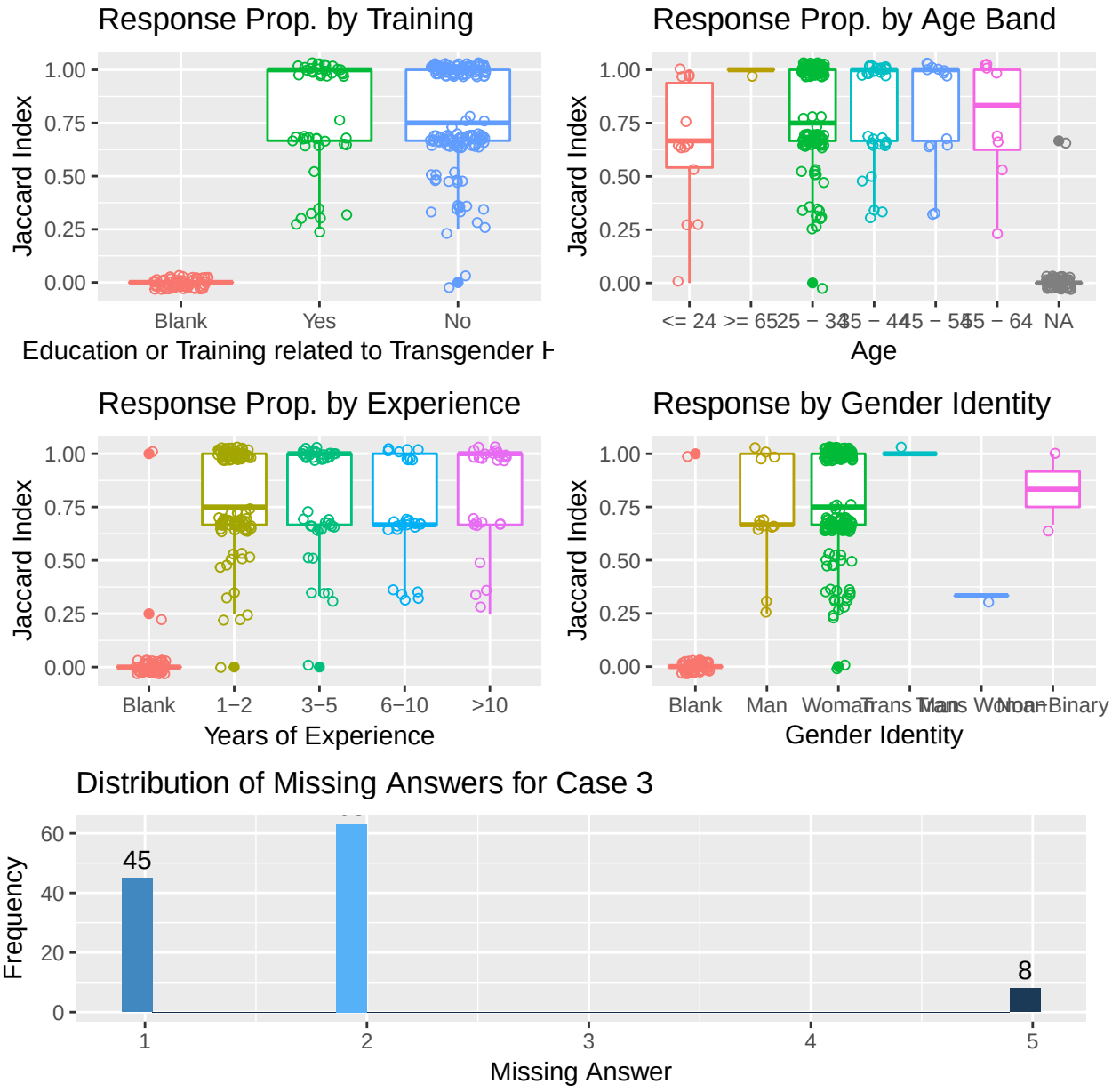
First, we see mean performance by case study for all respondents, with and without blank responses.

Case	Avg. JI (excl. blank)	Avg. JI (incl. blanks)
Case 3	0.78	0.6
Case 4	0.73	0.56
Case 5	0.85	0.66
Case 6a	0.83	0.63
Case 6b	0.78	0.59

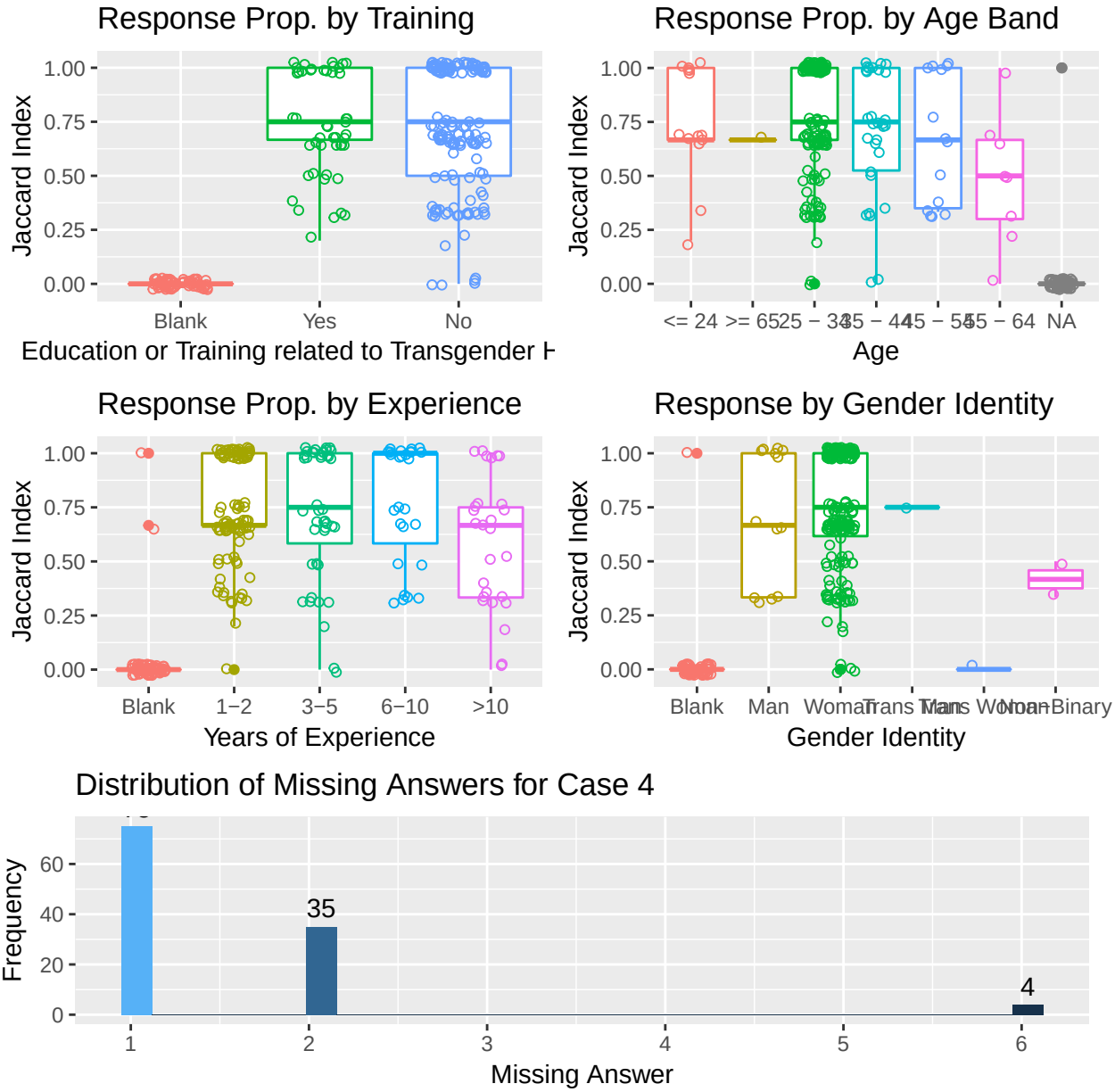
Overall, it appears that performance on the four cases is strong. Note that 6a refers to Case 6 when we do not include response 5 (which is subjectively correct) and 6b refers to Case 6 when we do include response 5.

Below, we visualize performance for each of the case studies broken out by a select group of experience and demographic covariates. We also examine the distribution of the missed answers.

Visualizations for Case Study 3

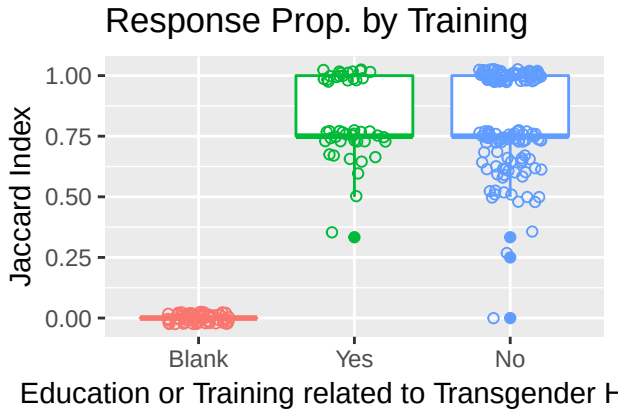


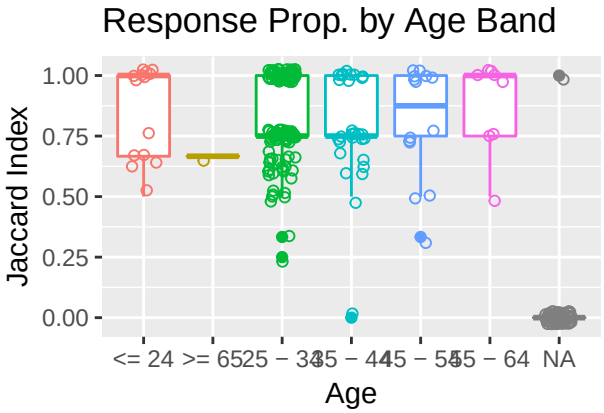
Visualizations for Case Study 4

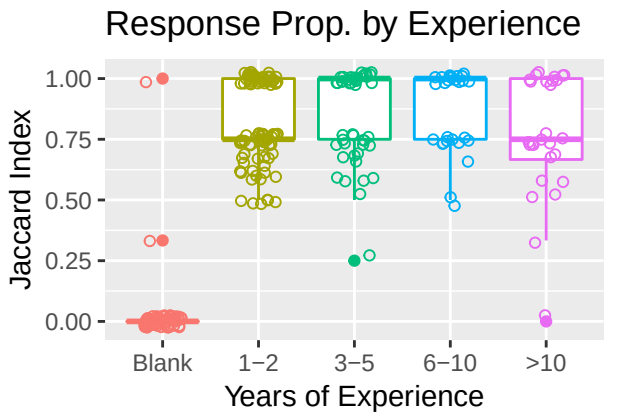


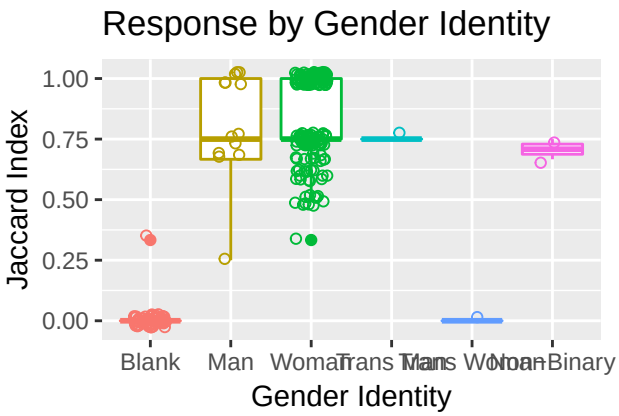
Visualizations for Case Study 5

Case 6 Visualizations (without Q5)

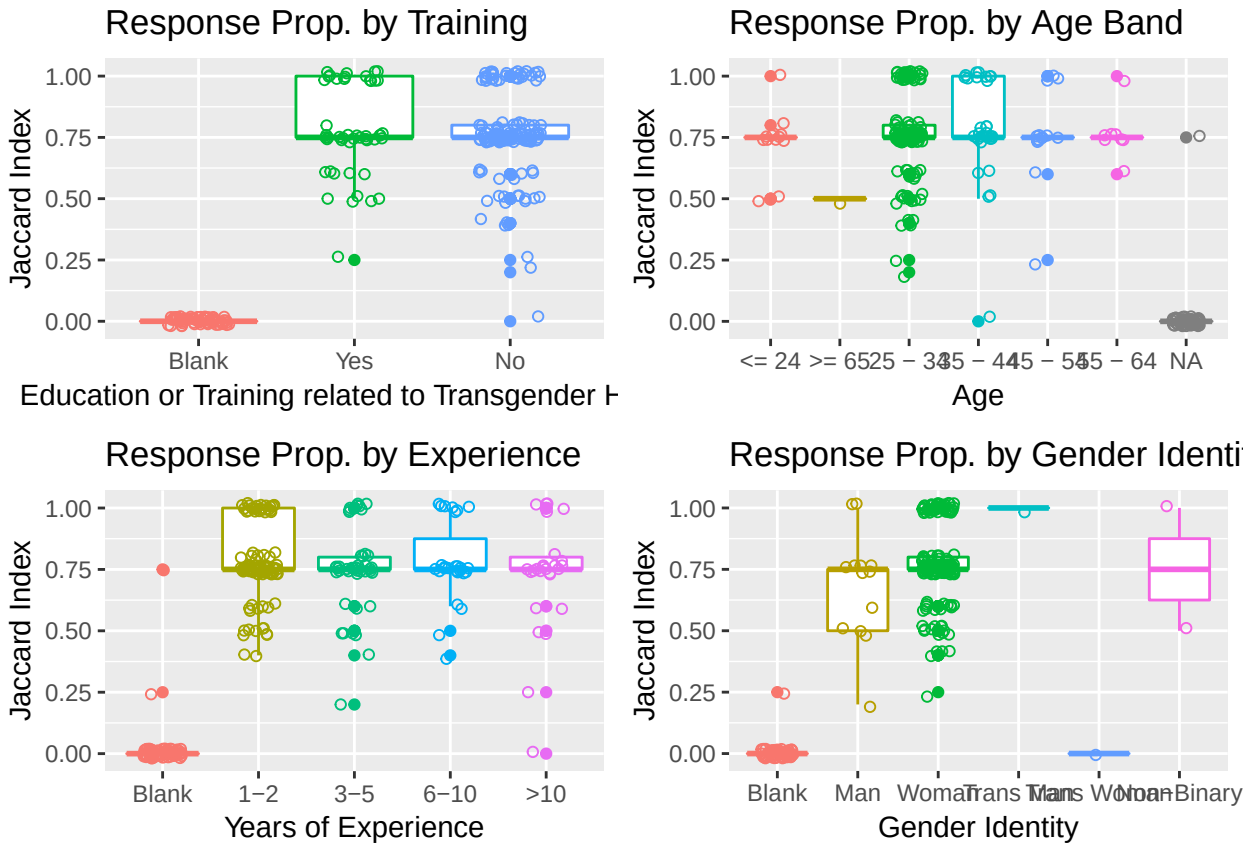






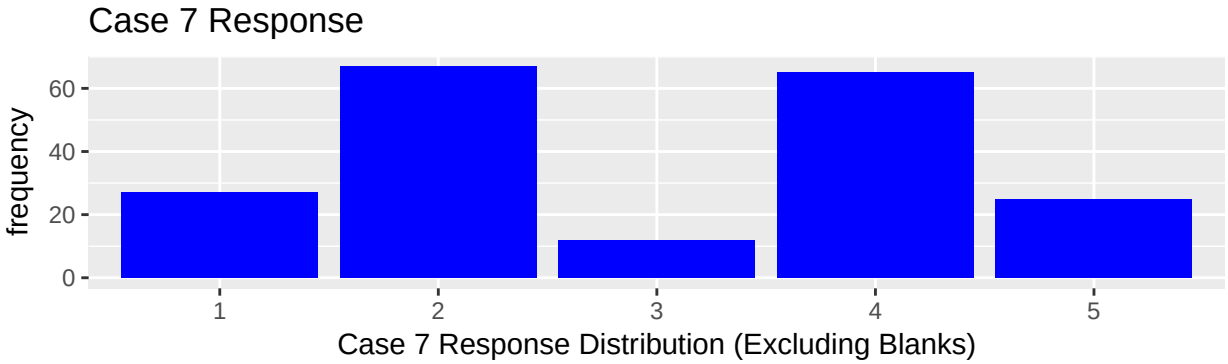


Case 6 Visualizations (with Q5)



Case 7 Visualization

Finally, we visualize the distribution of responses for Case 7.



3. Modeling

The potential covariates broadly fall into two main categories:

Clinician Experience

1. Number of patients seen each month (<40, ~40, >40)
2. Number of years spent working in a cancer setting

3. Seen any transgender patients?
4. Any transgender specific education/ training?

Clinician Demographics
1. Age
2. Gender identity
3. Sexual orientation
4. Race
5. Region of practice

Based on these, we fit a multiple linear regression model for each cases. In the model, jaccard Index was used as response variables and all the potential covariates mentioned above were used as predictors. In addition, since the survey was designed to stop when an answer is missing, we exclude all blank responses before we fit the model.

Here is our model:

$$Jaccard\ Index \sim \beta_0+\beta_1 specific\_training+\beta_2 Age\_Band+\beta_3 years\_experience+\beta_4 gender+\beta_5 n\_patients+\beta_6 seen\_trans+$$

Case Sutdy 3 Linear Model

Table 2: Fitting linear model: Q7_p[complete] ~ training + factor(Age_Band) + years + gender + n_patients + seen_trans + race + region

	Estimate	Std. Error	t value	Pr(>|t|)
training2	-0.072	0.0494	-1.457	0.1474
factor(Age_Band)25 - 34	-0.132	0.08762	-1.506	0.1344
factor(Age_Band)35 - 44	-0.1582	0.1099	-1.44	0.1524
factor(Age_Band)45 - 54	0.001312	0.1473	0.008907	0.9929
factor(Age_Band)55 - 64	0.04957	0.1477	0.3356	0.7377
years1	0.1906	0.3038	0.6275	0.5315
years2	0.2278	0.3085	0.7385	0.4615
years3	0.2216	0.315	0.7033	0.4832
years4	0.1732	0.3286	0.5272	0.5989
gender2	0.1499	0.08012	1.87	0.06369
gender5	0.02986	0.2698	0.1107	0.912
gender6	-0.0751	0.2689	-0.2793	0.7805
gender7	0.1876	0.2072	0.9057	0.3668
n_patients2	-0.02744	0.04732	-0.5798	0.563
n_patients3	-0.02832	0.06002	-0.4719	0.6378
seen_trans2	-0.02857	0.04557	-0.6269	0.5318
seen_trans3	0.132	0.1821	0.7247	0.4699
race1	0.4591	0.2726	1.684	0.09456
race2	0.4666	0.3012	1.549	0.1238
race5	0.4574	0.3215	1.422	0.1573
region1	-0.06101	0.2064	-0.2956	0.768
region2	0.1003	0.1973	0.5082	0.6122
region3	0.1132	0.2001	0.5659	0.5724
region4	0.1142	0.1978	0.5772	0.5648
region5	0.1509	0.2063	0.7314	0.4659
region6	0.1251	0.2066	0.6057	0.5458
(Intercept)	0.0872	0.4508	0.1934	0.8469

Case Sutdy 4 Linear Model

Table 3: Fitting linear model: Q9_p[complete] ~ training + factor(Age_Band) + years + gender + n_patients + seen_trans + race + region

	Estimate	Std. Error	t value	Pr(>|t|)
training2	-0.1032	0.05384	-1.916	0.05759
factor(Age_Band)25 - 34	0.2114	0.09549	2.214	0.02859
factor(Age_Band)35 - 44	0.1905	0.1198	1.59	0.1143
factor(Age_Band)45 - 54	0.3054	0.1606	1.902	0.05937
factor(Age_Band)55 - 64	0.3342	0.161	2.076	0.03987
years1	-0.2387	0.3311	-0.7209	0.4722
years2	-0.26	0.3362	-0.7735	0.4406
years3	-0.2903	0.3433	-0.8456	0.3994
years4	-0.3112	0.3581	-0.8692	0.3863
gender2	0.09138	0.08731	1.047	0.2973
gender5	-0.4031	0.2941	-1.371	0.1728
gender6	0.3083	0.293	1.052	0.2947
gender7	0.00663	0.2258	0.02936	0.9766
n_patients2	-0.01881	0.05157	-0.3648	0.7159
n_patients3	0.1299	0.06541	1.985	0.04921
seen_trans2	0.01767	0.04967	0.3557	0.7226
seen_trans3	0.3778	0.1984	1.904	0.05916
race1	0.8038	0.2971	2.706	0.007734
race2	0.9851	0.3283	3.001	0.003231
race5	1.067	0.3504	3.044	0.002834
region1	0.4791	0.225	2.129	0.03511
region2	0.4255	0.215	1.979	0.04995
region3	0.4161	0.218	1.908	0.05856
region4	0.4611	0.2156	2.139	0.03432
region5	0.2966	0.2249	1.319	0.1895
region6	0.4325	0.2251	1.921	0.05691
(Intercept)	-0.4791	0.4913	-0.9753	0.3313

Case Sutdy 5 Linear Model

Table 4: Fitting linear model: Q10_p[complete] ~ training + factor(Age_Band) + years + gender + n_patients + seen_trans + race + region

	Estimate	Std. Error	t value	Pr(>|t|)
training2	-0.03683	0.03269	-1.127	0.262
factor(Age_Band)25 - 34	0.03551	0.05798	0.6125	0.5413
factor(Age_Band)35 - 44	-0.07424	0.07274	-1.021	0.3093
factor(Age_Band)45 - 54	0.01302	0.09749	0.1335	0.894
factor(Age_Band)55 - 64	-0.05156	0.09774	-0.5275	0.5988
years1	0.2908	0.201	1.447	0.1504
years2	0.2882	0.2041	1.412	0.1604
years3	0.3423	0.2085	1.642	0.1031
years4	0.3824	0.2174	1.759	0.08095
gender2	-0.01433	0.05302	-0.2703	0.7874

	Estimate	Std. Error	t value	Pr(>|t|)
gender5	0.04083	0.1786	0.2286	0.8195
gender6	0.1453	0.1779	0.8166	0.4157
gender7	0.1564	0.1371	1.141	0.2561
n_patients2	-0.04529	0.03132	-1.446	0.1505
n_patients3	0.01227	0.03972	0.3088	0.758
seen_trans2	0.01681	0.03016	0.5574	0.5783
seen_trans3	0.08481	0.1205	0.7039	0.4828
race1	0.1188	0.1804	0.6589	0.5112
race2	-0.073	0.1993	-0.3662	0.7148
race5	0.2238	0.2128	1.052	0.2948
region1	-0.1765	0.1366	-1.292	0.1986
region2	-0.09294	0.1306	-0.7118	0.4779
region3	-0.1899	0.1324	-1.434	0.1539
region4	-0.1632	0.1309	-1.247	0.2148
region5	-0.1708	0.1366	-1.251	0.2132
region6	-0.1941	0.1367	-1.42	0.1581
(Intercept)	0.6154	0.2983	2.063	0.04112

Case Sutdy 6 Linear Model

Table 5: Fitting linear model: Q35_p1[complete] ~ training + factor(Age_Band) + years + gender + n_patients + seen_trans + race + region

	Estimate	Std. Error	t value	Pr(>|t|)
training2	-0.01991	0.0382	-0.5211	0.6032
factor(Age_Band)25 - 34	0.006398	0.06776	0.09443	0.9249
factor(Age_Band)35 - 44	0.043	0.085	0.5058	0.6139
factor(Age_Band)45 - 54	0.05571	0.1139	0.489	0.6257
factor(Age_Band)55 - 64	-0.1115	0.1142	-0.9761	0.3308
years1	0.4106	0.2349	1.748	0.08287
years2	0.3827	0.2385	1.604	0.1111
years3	0.4336	0.2436	1.78	0.07749
years4	0.473	0.2541	1.862	0.06493
gender2	-0.03916	0.06196	-0.632	0.5285
gender5	0.05499	0.2087	0.2635	0.7926
gender6	-0.4858	0.2079	-2.336	0.02101
gender7	-0.1373	0.1602	-0.8569	0.3931
n_patients2	0.01712	0.0366	0.4678	0.6407
n_patients3	0.01863	0.04641	0.4014	0.6888
seen_trans2	0.003184	0.03524	0.09035	0.9281
seen_trans3	-0.2359	0.1408	-1.675	0.09629
race1	-0.09244	0.2108	-0.4385	0.6617
race2	-0.0801	0.2329	-0.3439	0.7315
race5	0.0439	0.2487	0.1766	0.8601
region1	-0.1604	0.1596	-1.005	0.3168
region2	-0.1927	0.1526	-1.263	0.2088
region3	-0.1513	0.1547	-0.9777	0.33
region4	-0.124	0.153	-0.8109	0.4189
region5	-0.04344	0.1596	-0.2722	0.7859
region6	-0.2026	0.1597	-1.268	0.207

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.6876	0.3486	1.972	0.05071

Aggregate Linear Model

Table 6: Fitting linear model: y ~ training + factor(Age_Band) + years + gender + n_patients + seen_trans + race + region

	Estimate	Std. Error	t value	Pr(>|t|)
training2	-0.08385	0.08901	-0.942	0.3475
factor(Age_Band)>= 65	-0.08082	0.5403	-0.1496	0.8813
factor(Age_Band)25 - 34	0.1466	0.1484	0.9881	0.3246
factor(Age_Band)35 - 44	0.1579	0.1871	0.8436	0.4001
factor(Age_Band)45 - 54	0.06743	0.2459	0.2742	0.7842
factor(Age_Band)55 - 64	-0.05189	0.2404	-0.2159	0.8293
years1	0.1992	0.5946	0.335	0.7381
years2	0.2436	0.6017	0.4049	0.6861
years3	0.2417	0.6111	0.3956	0.6929
years4	0.08155	0.6337	0.1287	0.8978
gender1	-0.27	0.9802	-0.2755	0.7833
gender2	-0.09476	0.9717	-0.09752	0.9224
gender5	0.1861	1.099	0.1694	0.8657
gender6	-2.459	1.105	-2.225	0.02742
gender7	-0.7216	1.054	-0.685	0.4943
n_patients2	-0.09115	0.08311	-1.097	0.2743
n_patients3	0.02341	0.1036	0.2259	0.8215
seen_trans2	0.03514	0.08109	0.4333	0.6653
seen_trans3	0.1535	0.2625	0.5847	0.5596
race1	-0.1341	0.527	-0.2545	0.7994
race2	-0.05804	0.5871	-0.09886	0.9214
race3	-0.7603	0.7253	-1.048	0.296
race5	-0.1896	0.6222	-0.3047	0.761
region1	0.2924	0.3352	0.8723	0.3843
region2	0.1054	0.3244	0.3249	0.7457
region3	0.1515	0.3273	0.4629	0.6441
region4	0.1115	0.3249	0.3433	0.7318
region5	0.157	0.3411	0.4601	0.646
region6	0.2909	0.3413	0.8523	0.3953
(Intercept)	3.035	0.5208	5.829	2.862e-08

As we can see from the tables above, none of the potential covariates in any of the models are significant at the level of 0.05, which means we fail to reject all the null hypotheses. Thus, we can conclude that neither clinician experience (including number of patients seen each month, number of years spent working in a cancer setting, seen any transgender patients, got any transgender specific education/training), nor clinician demographics (age, gender, sexual orientation, race, region of practice) is significant in predicting clinicians' scores in each case.