


Exploring the effects of linguistic and cultural similarities between interviewer and respondents' behavior in large scale surveys

S. A. Cheema ^a, I. L. Hudson ^{b,a} , Q. Aslam ^c, M. Shafqat ^d, T. Kifayat ^d and Z. Hussain ^e

^a School of Mathematical and Physical Sciences, University of Newcastle, Australia, ^b Department of Mathematical Sciences, College of Science, Engineering and Health, Royal Melbourne Institute of Technology (RMIT), Melbourne, Australia, ^c Lahore Business School, The University of Lahore, Lahore, Pakistan, ^d Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan, ^e Department of Social & Allied Sciences, Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Pakistan.
Email: irene.hudson@rmit.edu.au

Abstract: In this paper, we explore the factors affecting respondents' degree of cooperation in large scale household surveys, with a focus on a multi-linguistic society. A detailed examination of the issue is conducted by using the Pakistan social and living standards measurement (PSLM) survey, 2014-15 data, gathered through a nationwide sample of 78,635 households. We observe that respondents' attention is not only affected by their socio-economic conditions, but also the linguistic similarities between interviewer and interviewee, which play a significant role in this matter. We considered *respondent behavior* (RB), assessed by enumerator and quantified on a Likert scale with six categories such as, "Co-operative = 1, Normal = 2, Reluctant = 3, Non-serious = 4, Refusal = 5 and Non-contact = 6" codes, as the study variable. The initial exploratory analysis revealed that at a national level, 29.57% of the respondents were co-operative and 62.19% were categorized with a normal attitude by interviewers. Further, 6.18% of the total 78,635 respondents were assessed as reluctant and the remaining 2.04% respondents were reported with non-serious behavior. Based on literature and experts' opinion, a range of explanatory variables are also taken into account in this study. Inspired by the American Statistical Association recommendations, we took *language of the interview* (L) as one of the explanatory variables. The PSLM 2014-15 questionnaire recognized nine languages including "Urdu = 1, Punjabi = 2, Sindhi = 3, Pushto = 4, Balochi = 5, Kashmiri = 6, Balti = 7, Hindko = 8, Siraiki = 9", whereas code 10 was assigned to other minor languages. We observed that most of the interviews were conducted in *Punjabi* (almost 26% of total sample). The five languages, *Urdu*, *Punjabi*, *Sindhi*, *Pushto* and *Siraiki* cumulatively contributed almost 97.1% and the contribution of *Urdu* (national language) remained almost 23%. Next, we used *household education level* (Edu) as a potential explanatory variable. The cumulative contribution of categories between class 1 to class 5 (inclusive) in terms of percentage, in defining educational level of a household remained almost 67.7%. Whereas, class 6 to class 10 (inclusive) contributed almost 15.4%. *Household health status* (Sick) is considered as a possible determinant of respondents' attitude. Exploratory analysis of the PSLM data revealed that almost 71% of the sampled households reported that no one in household was sick or injured during the last two weeks, whereas 22.5% of households documented one sick or injured member. Given the importance of economic factors as an important driver of social conduct, we also considered *household economic progression* (Eco) as an explanatory factor in this study. The PSLM (2014-15) survey questionnaire allowed for the assessment of relative progression of economic conditions of households with respect to the previous year. Based on the data, we observed that the most frequently reported response remains "Like before", where almost 48% of the respondents of the survey remained. Cumulatively, almost 19% of the respondents felt that they are better off as compared to the previous year, whereas, almost 36% (cumulatively) of households suffered a worse economic situation than last year. Only 0.17% of the survey respondents chose the option of "Do not know". Lastly, the locality of the interviewer (rural or urban) was considered as a determinant of respondents' attitude towards survey exercise. We found that, on average rural participants are more likely to have a cooperative attitude when compared to their urban fellows. At the modeling stage, the usual logistic regression, ignoring varying cultural streams of the country, led to un-interpretable and un-cohesive estimates of the factors potentially impacting respondents' response. Based on these observations and to meet the complexity of the socio-cultural diversity of the country, we propose a multi-level generalized linear modeling strategy to investigate the explanatory power of literature based factors which may impact the respondents' behavior in large scale survey exercises.

Keywords: Household survey data, Multi-level modeling, respondents' behavior

1. INTRODUCTION

Surveys have been long cherished as a go-to method to collect self-reported data in multidisciplinary research areas, such as economics, social and political sciences. Research communities from diverse fields use survey-based data as a source of scientific knowledge. For example, Hudson et al. (2010) used historical New Zealand election data (1893-1919) to study the gender effect over voter's turnout. Also, Castro (1995) used meta-analysis of 26 demographic surveys to study the nature and extent of association of female education with the family planning attitudes of couples. Moreover, note Cave et al.'s (1998) study quantifying the degree of the sophistication of polymerase chain reaction (PCR) technique in assessing the likelihood of leukaemia relapse in children, see also Reaman and Smith (2011). For a detailed and interesting account of the utility of surveys in scientific inquiries, one may consult Gideon (2012), Fowler (2013) and Warkentin (2014). In this recent age of information and data, we have witnessed an explosion of surveys directed to administrative purposes. For instance, the Pakistan social and living standards measurement (PSLM) survey by the Pakistan Bureau of Statistics (PBS) to assess the living conditions of the country's population, the European health interview survey by Euro Stat to explore health conditions and measure the credibility of health care systems and the American community survey of the US Census Bureau which addresses the issue of gender equality and racial discrimination.

Given the significance of surveys in the research literature, many studies have made respondent behavior and the extent of respondent co-operation of prime interest. From a survey practitioner perspective, Gill (1993) listed ten factors that must be taken into account to ensure the validity of survey results, especially in developing or under-developed countries. In the list, a respondent was associated with six factors, such as, degree of familiarity with the rationale of the study, educational background and social standing of the respondent, the warmth of rural communities towards strangers (interviewer) and language of the interview. Increasing the respondents' co-operation and once acquired, then maintaining it, remains the primary focus of numerous academic discussions originated from a wide range of research fields. Pedersen and Nielsen (2016) emphasized the importance of the need for a high extent of co-operation from respondents in a survey to reduce bias, thus increasing the generality and validity of results, see also Dillman et al. (2009) for a review of the issue. Stecklov et al (2015) noted that the extent of engagement of the survey respondents can be enhanced by using a community familiar person as the interviewer. Additionally, several scholars suggested the use of incentives such as cash prizes, lottery and charity donation, see for example Rose et al. (2007), Edwards et al. (2009) and Fan and Yan (2010).

Our research, instead of advising on a strategy to enhance the response rate directly, focuses on empirical evaluations of the respondent's socio-cultural attributes that may impact the extent of his/her co-operation in a survey, especially in multi-linguistic societies. The objectives of the research are tested by studying the PSLM 2014-15 household survey data assembled at the district level. This micro-level assembly of information thus provides the opportunity to explore these issues in a more homogenous cultural and administrative context. In summary, our findings are based on a nationwide representative sample of 78,635 households from 114 districts of Pakistan, covering a variety of covariates, ranging from personal factors to external factors.

2. BACKGROUND

As our prime focus lies on district level analysis so as to gain the maximum available extent of homogeneity (administratively and culturally), we proceed by exploring respondents' behavior with reference to districts. Table 1 displays the percentages of respondents judged by interviewers according to their tendencies towards the household survey 2014-15, while taking into account 114 districts. Table 1 reveals interesting information about the extent of cooperativeness of respondents at the district level. The relevant information is italicized in the table. The maximum percentage of *co-operative* tendencies is 90.82% reported from the district of *Batagram* – a district of KPK province. Whereas, the minimum percentage of *co-operative* respondents was recorded as 0.49% from the *Chaghi* district of Balochistan province. On the other hand, this same district *Chaghi*, reveals maximum respondents attributed with *normal* behavior towards the survey – the percentage remains 98.76%, whereas, minimum for this category is documented in KPK's district of *Batagram*, where the reported percentage remains 8.64%. Similarly, the maximum percentage of *reluctant* behavior, with respect to the PSLM 2014-15 survey, is 43.65%, observed in the *Barkhan* district of Balochistan. Moreover, 44 districts exhibit a higher percentage of *co-operative* respondents as compared to the overall national level percentage of 29.57%. Thus, overall estimate at the 60th percentile of the districts, that is, 40% of the districts have a higher extent of co-operative attitude than overall. Similarly, the overall percentage of *normal* behavior of respondents during the survey, remains at the 40th percentile with respect to districts, thus 60% of the districts display a higher percentage of normal attitude of respondent compared with the national level percentage

Table 1. Summary of the respondents' behavior during the PSLM 2014-15 survey

Province	District	Behavior Categorization				Province	District	Behavior Categorization			
		Co-operative	Normal	Reluctant	Non-serious			Co-operative	Normal	Reluctant	Non-serious
KPK	Chitral	31.68	52.59	11.21	4.53	Punjab	Jhang	56.12	41.61	1.89	0.38
	Upper Dir	21.64	69.03	8.40	0.93		T. T. Singh	53.60	43.53	2.61	0.27
	Lower Dir	19.84	61.72	16.03	2.40		Gujran-Wala	11.84	76.84	10.13	1.18
	Swat	13.50	73.23	10.18	3.10		Hafizabad	9.33	73.38	15.38	1.91
	Shangla	25.57	66.35	4.31	3.77		Gujrat	15.98	76.12	6.53	1.37
	Buner	9.18	71.72	14.43	4.66		MandiBahauddin	44.50	53.88	1.46	0.16
	Mala-kand	19.21	63.54	13.32	3.93		Sialkot	73.80	19.65	3.67	2.88
	Kohistan	19.38	43.30	34.20	3.12		Narowal	45.15	39.64	14.89	0.32
	Man-sehra	56.52	42.17	0.87	0.43		Lahore	40.29	52.23	7.13	0.36
	Bata-gram	90.82	8.64	0.53	0.00		Kasur	26.31	69.01	4.50	0.18
	Abbottabad	56.76	41.69	1.33	0.22		Sheikhu-pura	39.91	57.22	2.41	0.46
	Haripur	81.03	11.85	3.88	3.23		Nankana Sahib	34.52	62.63	2.67	0.18
	Tor Ghar	64.57	35.43	0.00	0.00		Okara	33.88	64.96	0.99	0.17
	Mardan	20.54	61.61	14.51	3.35		Sahiwal	56.67	42.34	0.82	0.16
	Swabi	13.01	72.15	12.56	2.28		Pakpattan	48.69	49.93	0.83	0.55
	Char-sadda	18.53	67.82	10.79	2.85		Vehari	18.66	71.03	8.18	2.13
	Pesha-war	25.56	57.14	14.04	3.26		Multan	36.79	51.70	10.49	1.01
	Now-shera	8.05	75.86	11.49	4.60		Lodhran	38.85	48.12	11.90	1.14
	Kohat	17.78	68.82	10.39	3.00		Khanewal	37.10	51.14	10.65	1.11
	Hangu	13.15	63.85	15.73	7.28		D. G. Khan	73.83	24.63	0.86	0.68
Karak	4.85	90.31	3.96	0.88	Rajanpur	46.75	34.74	8.12	10.39		
Bannu	6.76	84.02	7.17	2.05	Layyah	83.28	15.53	1.04	0.15		
LakkiMarwat	7.07	88.38	3.54	1.01	Muzaffar-garh	14.84	68.21	15.73	1.22		
D. I. Khan	30.69	64.48	3.22	1.61	Bahawalpur	3.03	84.82	9.81	2.35		
Tank	32.06	64.35	2.63	0.96	Bahawal-nagar	87.74	10.92	1.09	0.24		
Punjab	Attock	3.66	73.48	16.79	6.06	Rahim Yar Khan	23.30	71.21	4.73	0.76	
	Rawal-pindi	1.38	74.43	19.60	4.59	Jacobabad	8.63	87.82	0.96	2.59	
	Jhelum	0.91	83.64	12.36	3.09	Kashmore	7.00	89.21	0.66	3.12	
	Chakwal	1.39	70.83	22.40	5.38	Shikarpur	24.53	71.93	2.02	1.52	
	Sargodha	46.98	48.00	4.71	0.31	Larkana	23.04	71.62	3.34	2.00	
	Bhakkar	24.10	64.61	9.53	1.77	Shahdadt	26.12	67.75	2.68	3.46	
	Khushab	47.66	48.79	3.46	0.09	Sukkur	16.18	79.92	3.51	0.39	
	Mian-wali	41.37	49.95	7.77	0.91	Ghotki	11.83	79.67	4.67	3.84	
	Faisalabad	20.88	76.50	2.04	0.58	Khairpur	53.05	43.17	0.58	3.20	
	Chiniot	18.03	81.18	0.79	0.00	NaushahroFeroze	20.43	71.91	5.30	2.36	

Province	District	Behaviour Categorization				Province	District	Behaviour Categorization			
		Co-operative	Normal	Reluctant	Non-serious			Co-operative	Normal	Reluctant	Non-serious
Sindh	Shaheed Benazirabad	2.81	96.79	0.40	0.00	Balochistan	Musakhel	82.37	17.63	0.00	0.00
	Dadu	19.17	64.49	9.80	6.54		KillaSaifullah	1.09	93.75	4.08	1.09
	Jamshoro	22.95	71.17	3.96	1.91		Zhob	21.49	53.95	6.80	17.76
	Hyderabad	23.35	74.71	1.56	0.39		Sheerani	46.96	44.76	6.59	1.69
	Tando Allah Yar	16.89	82.87	0.12	0.12		Sibbi	29.05	67.57	1.35	2.03
	Tando M. Khan	16.43	82.15	1.10	0.33		Harnai	5.04	80.34	1.44	13.19
	Matiari	19.10	80.15	0.60	0.15		Ziarat	5.94	78.75	3.44	11.88

	Badin	20.14	78.87	0.77	0.23		Kohlu	31.42	46.56	10.09	11.93
	Thatta	37.92	54.03	7.32	0.73		DeraBugti	90.13	9.65	0.00	0.22
	Sujawal	34.78	57.69	6.25	1.28		Bolan/ Kachhi	6.32	93.42	0.26	0.00
	Sanghar	12.20	87.32	0.32	0.16		Jaffarabad	21.17	70.84	1.51	6.48
	MirpurKhas	17.67	81.19	1.14	0.00		Nasirabad/ Tamboo	17.07	77.61	0.44	4.88
	Umerkot	20.68	78.62	0.70	0.00		JhalMagsi	32.00	56.53	5.33	6.13
	Tharpar- kar	48.69	48.12	3.10	0.09		Kalat	22.97	69.59	3.38	4.05
	Karachi	33.60	57.07	8.20	1.13		Mastung	20.83	76.60	1.92	0.64
Baloch- istan	Quetta	43.97	49.65	3.55	2.84	Khuzdar	21.04	71.04	2.74	5.18	
	Pishin	10.34	78.06	2.82	8.78	Awaran	19.17	80.42	0.42	0.00	
	Killa Abdullah	8.61	77.78	3.06	10.56	Kharan	51.02	42.52	0.00	6.46	
	Chagai	0.50	98.76	0.50	0.25	Washuk	32.75	63.77	0.25	3.23	
	Nushki	10.25	89.40	0.35	0.00	Lasbela	44.69	48.23	7.07	0.00	
	Loralai	76.18	23.82	0.00	0.00	Gwadar	2.84	89.01	1.77	6.38	
	Barkhan	9.32	37.71	43.64	9.32	Capital	Islamabad	3.51	71.00	22.14	3.34

of the category under consideration. Next, with respect to *reluctant* respondents, 43 districts show higher percentages in comparison to the overall percentage of 6.18%, which is almost at the 63rd percentile of the district extent of a reluctant attitude. Lastly, 48 districts demonstrate a higher percentage of a *non-serious* respondent than the national level percentage of 2.04, which is at the 58th percentile with respect to the district level percentages. Thus 42% of the 114 districts with higher percentages of respondents labeled as non-serious, than overall category. Next, we present our strategy to cope with the issue of highly volatile respondents' attitude by modeling the different cultural streams existent at country level.

3. PROPOSED STRATEGY

We propose a multi-level logistic regression model which is capable of maintaining the distinctive homogenous structures at the contextual (a district) level. This is achieved by incorporating a random intercept allowed to vary across the groups or districts in our study. The rationale of our proposed approach heavily relies on the understanding that the cultural and administrative homogeneities of districts ensure demographic similarities between interviewer and interviewee. Let us say Y_{ijk} is the perceived cooperativeness of the i 'th respondent associated with the j 'th covariate recorded in k 'th district, where $Y_{ijk} = 1$ if attitude is cooperative and '0' otherwise. In the first phase, we define the i 'th respondent behavior in a k 'th district. The model is as follows,

$$g(\pi_{ik}) = \beta_{ok} + \sum_j \beta_j x_{ijk}, \quad (1)$$

where, $g(\pi_{ik}) = \log(\pi_{ik}/1 - \pi_{ik})$, π_{ik} denotes the probability of i 'th respondent in k 'th district being cooperative and β_{ok} is the intercept permitted to vary across k districts while $k = 1, 2, \dots, 114$. In the next phase, the district level variability prevalent in the respondents' behavior is captured by treating micro-level intercepts (within each district) separately through a linear model as follows,

$$\beta_{ok} = \gamma_k + \vartheta_{ok}, \quad (2)$$

where, $\vartheta_{ok} \sim N(0, \sigma_{kk})$. Moreover, γ_k is the vector of indicator values of residence status of the i 'th respondent in the k 'th district, with the assumption that for $k \neq k^*$, $cov(\vartheta_{ok}, \vartheta_{ok^*}) = 0$. In the next section, we will delineate the gain of employing this multi-level strategy to handle distinctive social streams existing in a society.

The general model given in equation (1) is simplified with the modeling of respondents' behavior, including respondents' education, region, language of interview, household health conditions and household economic progression as explanatory variables, as follows,

$$g(\pi_{ik}) = \beta_{ok} + \beta_1(\text{Edu})_{ik} + \beta_2(\text{Region})_{ik} + \beta_3(\text{Urdu})_{ik} + \beta_4(\text{Punjabi})_{ik} + \beta_5(\text{Sindhi})_{ik} + \beta_6(\text{Pushto})_{ik} + \beta_7(\text{Siraiki})_{ik} + \beta_8(\text{Others - lang})_{ik} + \beta_9(\text{Sick})_{ik} + \beta_{10}(\text{Eco})_{ik}, \quad (3)$$

where, the intercept β_{ok} is allowed to vary across the 114 districts of the country. The variability associated with each district is further modeled by regressing intercepts through the following model,

$$\beta_{ok} = \gamma_k + \vartheta_{ok}, \quad (4)$$

where, ϑ_{ok} is normally distributed error term specified above.

4. RESULTS

We now demonstrate the utility of our proposed model in handling the heterogeneity in respondents' attitude towards the survey presented in the Table 1 at contextual levels (districts). First, we argue an important building block of our proposed model, that is, the assumption of normally distributed error in the model of equation (4). In Figure 1, we present the QQ plot verifying the feasibility of the normality assumption. Further, in Figure 2, we present the correlation matrix of the fixed effects. The extent of correlation that exists, between language in which interviews were conducted and the intercept, support the altering effects of linguistic diversity across the district in modeling the respondents' behavior. We next report estimates of fixed effects and their interpretation within the context of our study. Figure 3, provides graphical display depicting the strength, direction and significance of the relationship, estimated through multi-level logistic regression (equation 3), between the study variable and the explanatory variable, in terms of odds. The co-efficient of *education* (Edu) is positively and significantly associated with respondents' behavior, that is, the higher the education category of a household, the more likely is the respondent to be cooperative. The *region* of the respondent is negatively and significantly associated in explaining the respondent's tendencies towards the PSLM survey, which means respondents' from rural areas are more likely to cooperate than urban areas on average (see also Gill (1993)). Further, the *household economic progression* (Eco) remains statistically significant and positively associated with respondents' degree of cooperation. Those respondents who said they were better off

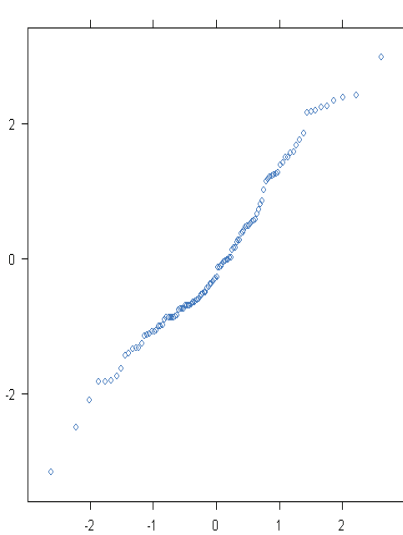


Figure 1. The QQ plot.

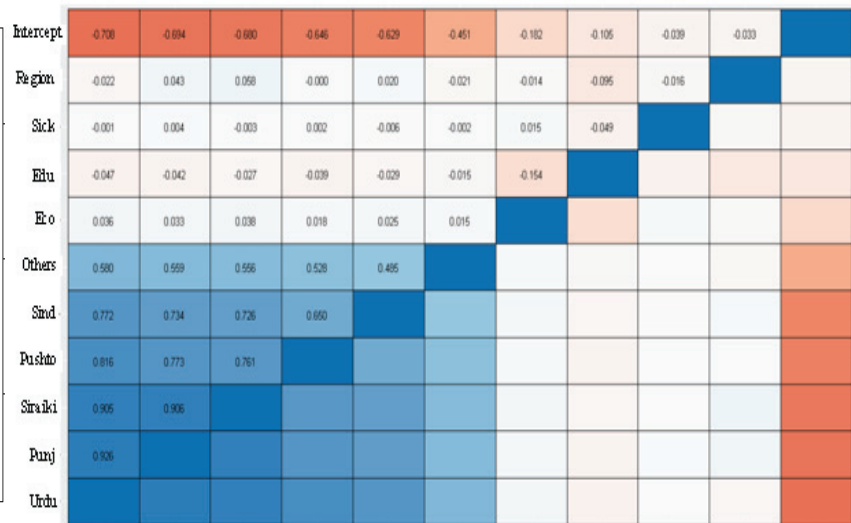


Figure 2. Correlation matrix of the fixed effects.

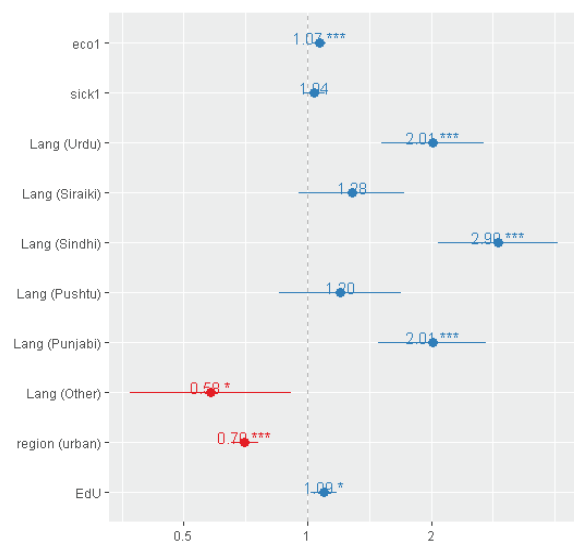


Figure 3. Estimated fixed effects in terms of odds.

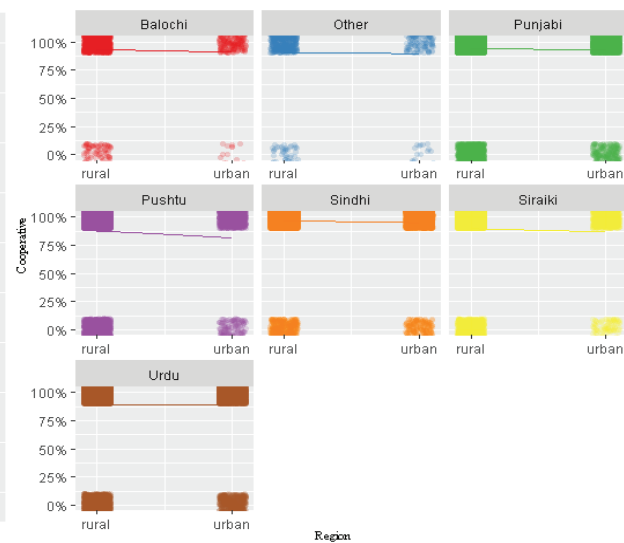


Figure 4. Predictive probabilities of a cooperative respondent.

Table 2. Estimates of random effect when multi-level modeling is employed

Province	District	Estimate	Province	District	Estimate	Province	District	Estimate	Province	District	Estimate
KPK	Chitral	-1.326	Punjab	Sargodha	-0.048	Punjab	Bahawalpur	-0.689	Baluchistan	Killa Abdullah	-0.877
	Upper Dir	-0.279		Bhakkar	-0.737		Bahawal-nagar	1.202		Chagai	2.172
	Lower Dir	-1.081		Khushab	0.275		Rahim Yar Khan	0.156		Nushki	2.416
	Swat	-0.644		Mian-wali	-0.605		Jacobabad	-0.041		Loralai	2.377
	Shangla	-0.138		Faisal-abad	0.658		Kashmore	-0.087		Barkhan	-3.163
	Buner	-1.135		Chiniot	1.668		Shikarpur	-0.026		Musakhel	2.339
	Mala-kand	-0.991		Jhang	0.724		Larkana	-0.427		KillaSaifullah	0.371
	Kohistan	-2.508		T.T. Singh	0.464		Shahdadkot	-0.613		Zhob	-1.410
	Man-sehra	1.232		Gujran-wala	-0.863		Sukkur	0.009		Sheerani	-0.649
	Bata-gram	2.159		Hafizabad	-1.439		Ghotki	-1.001		Sibbi	0.409
	Abbott-abad	1.257		Gujrat	-0.539	Khairpur	-0.120	Harnai		-1.118	
	Haripur	0.544		MandiBahauddin	1.004	NaushahroFeroze	-0.878	Ziarat		-1.139	
	Tor Ghar	2.980		Sialkot	-0.372	Shaheed Benazir-abad	1.842	Kohlu		-1.740	
	Mardan	-0.999		Narowal	-1.345	Dadu	-1.636	DeraBugti		2.251	
	Swabi	-0.757		Lahore	-0.304	Jamshoro	-0.508	Bolan/ Kachhi		2.236	
	Char-sadda	-0.691		Kasur	-0.017	Hyderabad	0.799	Jaffarabad		-0.911	
	Pesha-war	-0.875		Sheikhu-pura	0.509	Tando Allah Yar	2.186	Nasirabad/ Tamboo		-0.437	
	Now-shera	-0.847		Nankana Sahib	0.479	Tando M. Khan	0.843	JhalMagsi		-0.739	
	Kohat	-0.633		Okara	1.277	Matiari	1.382	Kalat		-0.333	
	Hangu	-1.255		Sahiwal	1.410	Badin	1.217	Mastung		0.578	
Karak	0.387	Pakpattan	1.139	Thatta	-0.675	Khuzdar	-0.511				
Bannu	-0.294	Vehari	-0.752	Sujawal	-0.569	Awaran	1.582				
LakkiMarwat	0.478	Multan	-0.694	Sanghar	1.755	Kharan	-0.133				
D. I. Khan	0.265	Lodhran	-0.869	MirpurKhas	1.171	Washuk	0.156				
Tank	0.569	Khanewal	-0.880	Umerkot	1.489	Lasbela	-0.369				
Attock	-1.805	D. G. Khan	1.497	Tharpar-kar	0.236	Gwadar	0.007				
Rawal-pindi	-1.819	Rajanpur	-1.071	Karachi	-0.498	Capital	Islamabad	-1.831			
Jhelum	-1.319	Layyah	1.559	Quetta	0.127						
Chakwal	-2.093	Muzaffar-garh	-1.079	Pishin	-0.689						

in the year 2014-15 in comparison to the previous year 2013-14 were the cooperative participants. Household *health status* remained insignificant in impacting respondents' tendencies. Figure 4 depicts the predictive probabilities of respondent being cooperative for all languages, while displaying the distinction of rural and urban participants. We, next report the results of the micro-level logistic regression, namely the results of regression within each district, modeling the district level diversity by permitting random intercepts (equation 4). The results compiled in Table 2 show clear evidence of the variations in respondents' attitude towards the 2014-15 PSLM survey due to the district level differences. By employing a multi-level modeling strategy and thus controlling the variability prevalent at the district level, our proposed model is in fact inducing homogeneous within group structure. The minimum estimate of random intercept equals -3.136 (Barkhan district), whereas the maximum intercept estimate of 2.980 (Torgarh district).

5. CONCLUSIONS

The major result is the role of the *language of interview* in defining respondents' tendencies of cooperativeness, when both respondent and enumerator possess cultural similarities. It is encouraging to observe that, once the district level diversity is being adjusted for and linguistic homogeneity within districts is allowed to play its role, the relevant estimates become interpretable. The effects of language of communication are interpretable with reference to *Baluchi* language as follows. For instance, the overall coefficient associated with the Baluchi language, when the given respondent is from the rural area (code 0), household education level minimal (no education - code 0), household economic conditions better off in year 2013-14 as compared to 2014-15 (code 0) and with no sick or injured household member (code 0), is 2.186. The co-efficient is then positively and significantly associated with respondents' attitude towards the PSLM 2014-15 survey. The odds of a cooperative respondent, when mode of interview was *Urdu* or *Punjabi*, is almost two times higher than if the mode was *Balichi*. Whereas, in the case of *Sindhi* language the odds of

cooperativeness is 2.91 times higher compared to Baluchi language. In comparison to other less prevalent languages, the odds of witnessing a more cooperative respondent, when the interview was conducted in Baluchi language is higher – the odds is 6.95 times higher. The *Pushto* and *Sariaki* languages reveal no significant effect in comparison to the Baluchi language on the respondents' degree of cooperation in PSLM 2014-15 household survey. The major result of this study is evidence of the impact of language of interview in relationship to a respondent's increased tendency to exhibit cooperativeness, this in the case when both respondent and enumerator possess cultural similarities. The most important source of a vast body of officially published statistics across the globe is generated as a result of surveys. It is therefore of no surprise that the study of respondent behavior is constantly gaining popularity among survey practitioners and applied researchers. In this article, we explore the impact of linguistic and cultural similarities of interviewer and respondents on the extent of cooperation of respondents in a large scale survey. The findings indicate that cultural homogeneity plays an important role in gaining higher cooperation in respondents for survey research. The issue however is complex in nature. The questionnaires of main stream developmental surveys are developed under the guidance of the UN statistics department to achieve maximum data consistency for comparative purposes and therefore the mother language of questionnaires remains to be English. The questionnaires are then translated by experts into local languages (Urdu - for Pakistan) for training and enumeration purposes. At the first stage, translation is itself a potential source of misunderstanding and misinterpretation. Further difficulty arises in multi-linguistic countries, like Pakistan, with lower literacy rates. In these situations, enumerators are bound to further translate questions into a local language to attain a response and to maintain response rates. This policy, without proper training is likely to dilute the true meanings of question(s). By exploring the importance of interview language in face to face survey research, we have suggested more rigorous training sessions focused on major languages of the country, accommodating for other situations that exist in various cultural settings. Though this process is rigorous and will impose time delays and additional cost, we believe this is the best path to attain valid information in a cohesive manner.

REFERENCES

- Castro, T. M. (1995). Women's education and fertility: Results from 26 demographic and health surveys. *Studies in Family Planning*, 26, 187-202.
- Cave, H. C., Bosch, V. W., Suciu, S., Guidal, C., Waterkeyn, C., Otten, J., Bakkus, M., Thielman, K., Grandchamp, B., Vilmer, E., Nelken, B., Fournier, M., Boutard, P., Lenrun, E., Mechinaud, E., Garand, R., Robert, A., Dastungne, N., Plouvier, E., Racodot, E., Ferster, A., Gyselinck, J., Fenneteau, O., Duval, M., Solbue, G. and Manel, A. M. (1998). Clinical significance of minimal residual disease in children acute lymphoblastic leukemia. *New England Journal of Medicine*, 339, 591-598.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Mail and internet surveys: The tailored design method*. 3rd ed. New York, NY: Wiley.
- Edwards, P. J., Roberts, I., Clarke, M. J., DiGiuseppi, C., Wenzl, R., Kwan, I., Cooper, R., Felix, L. M., & Pratap, S. (2009). Methods to increase response to postal and electronic questionnaires (review). *Cochrane Database of Systematic Reviews* 2009, Issue 3.
- Fan, W. and Yan, Z. (2010) Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior*, 26(2), 132-39.
- Fowler, J. F. (2013). *Survey Research Methods*. 5th ed. SAGE Publication, USA.
- Gideon, L. (2012). *Handbook of Survey Methodologies for Social Sciences*. City University of New York, NY, USA. Springer.
- Gill G. J. O.K. (1993). The data's lousy, but it's all we've got (being a critique of conventional methods). *London: Gatekeeper Series. Report No. 38*. Available: <http://pubs.iied.org/pdfs/6052IIED.pdf>
- Hudson, I.L., Moore, L. Beh, E.J. and Steel, D.G. (2010). Ecological inference techniques: an empirical evaluation using data describing gender and voter turnout at New Zealand elections 1893-1919. *Journal of the Royal Statistical Society, Series A*, 173, 185-213.
- Pedersen, M. J. and Nielsen, C. V. (2016). Improving survey response rates in online panels: effects of low-cost incentives and cost-free text appeal interventions. *Social Science Computer Review*, 34 (2), 229-243.
- Reaman, G. H. and Smith, F. O. (2011). *Childhood Leukaemia: A Practical Handbook*. Springer: New York.
- Rose, D. S., Sidle, S. D., & Griffith, K. H. (2007). A penny for your thoughts: Monetary incentives improve response rates for company-sponsored employee surveys. *Organizational Research Methods*. 10(2), 225-240.
- Stecklov, G., Weinreb, A. A., & Sana, M. (2015). Family planning for strangers: an experiment on the validity of reported contraceptive use. *PloS one*, 10(8). e0136972. <https://doi.org/10.1371/journal.pone.0136972>
- Warkentin, M. (2014). *Trends and Research in the Decision Sciences: Best Papers from the 2014 Annual Conference*. Pearson FT Press, USA.