Lecture 3: Supervised methods

Julian Ashwin

London Business School

Heidelberg, 2022

Road Map

- What are supervised methods?
- Feature extraction: joint or unsupervised
- Application: Qualitative Analysis at Scale with Rohingya refugees in Bangladesh.
- Focus in application is the use of supervised NLP models in very small samples - a common issue for social science applications

Supervised

Supervised versus unsupervised distinction:

1. Supervised methods aim to use some input to predict an output (e.g. regression or classification)
2. Unsupervised methods aim to summarise and identify patterns in an unlabelled dataset (e.g. principle components analysis).

Supervised learning

Supervised learning uses some input data to predict the value of some target value (or “label”).

Models range from OLS to massive neural networks, but fundamentally boil down to either regression of classification.

In an NLP context, this usually involves using the text to predict some numerical variable(s).

There are exceptions (e.g. STM [Roberts et al., 2013]) uses numerical data to predict text).

Two approaches

1. Extract features first
   - Dictionary methods
   - ngrams
   - Term counts, tf-idf
   - Topics, embeddings
2. Joint extraction and prediction
   - sLDA, SCHOLAR
   - fine-tune embeddings like BERT by training for a classification task.

Worked examples

In the 3_supervised_methods.R script we use a LASSO regression on term frequencies to predict GDP growth.

Feature extraction simply unigram term-frequencies, sparsity enforced by regularisation.

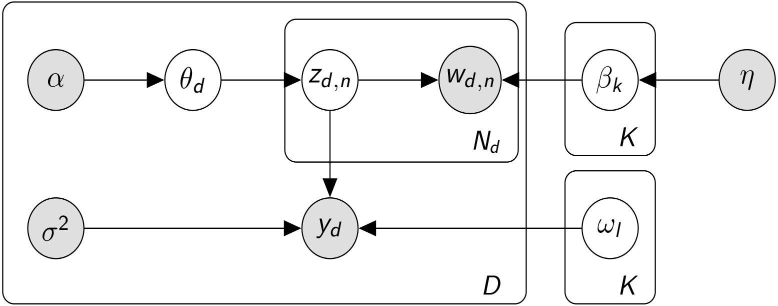
Some interpretability - coefficient on each term - but can get tricky when regression gets very high dimensional.

Perhaps some predictive value in the text when used in this way, but could we do any better?

supervised LDA

1. **for** $k = 1, \ldots, K$:
    1. $\beta_k \sim \text{Dir}(\eta)$
2. **for** $d = 1, \ldots, D$:
    1. $\theta_d \sim \text{Dir}(\alpha)$
    2. **for** $n = 1, \ldots, N_d$:
        1. topic assignment $z_{d,n} \sim Mult(\theta_d)$
        2. term $w_{d,n} \sim Mult(\beta_{z_{d,n}})$
3. $y \sim \mathcal{N}\left(z\omega, \sigma^2 I\right)$

Graphical model



Gibbs EM algorithm

- Can collapse out θ and β, following [Griffiths and Steyvers, 2004].
- Posterior distribution is then:

$$p(Z, \omega, |W, y) = p(Z|W, y, \omega, \sigma^2)p(\omega, |Z, y). \tag{1}$$

- E-step: approximate $p(Z|W, y, \omega)$ by Gibbs sampling.

$$p(z_{d,n} = k|z_{-(d,n)}, W, X, y, \alpha, \eta, \omega, \sigma^2) \propto$$
$$(s_{d,k,-n} + \alpha)\frac{m_{k,v,-(d,n)} + \eta}{\sum_v m_{k,v,-(d,n)} + V\eta} \exp\left\{\frac{1}{2\sigma^2}\left(\frac{2\tilde{\omega}_{z,d,k}}{N_d}\left(y_d - \frac{\tilde{\omega}_{z,d}^{\mathsf{T}}}{N_d}s_{d,-n}\right) - \left(\frac{\tilde{\omega}_{z,d,k}}{N_d}\right)^2\right)\right\}.$$

- M-step: given $\mathbb{E}(Z|\cdot)$, we can use Maximum Likelihood to find $p(\omega|Z, y)$.

Worked example

In the 3_supervised_methods.R script we use a sLDA model to predict GDP growth.

For larger models, sLDA gives more easily interpretable results than a LASSO.

But it's a very high dimensional model, so out-of-sample performance poor for small datasets (see next lecture for performance in larger datasets).

Motivation

Qualitative and quantitative methods each have strengths and weaknesses.

- Qualitative:
  - Strength: nuanced and complex signals, based on expert reading
  - Weakness: small, unrepresentative samples
- Quantitative
  - Strengths: large, representative samples
  - Weaknesses: can hide important complexity

- This paper:
  - Use NLP to scale-up small sample qualitative work to larger, representative samples.
  - Applications to **aspirations**, belonging and well-being among Rohingya refugees in Bangladesh (Cox's Bazaar).

Why not use established NLP tools?

Many NLP methods we could use that broadly fall into two categories:

1. Unsupervised (e.g. topic models, embeddings):
    - Explain variation in text: not geared towards question.
    - In small samples, we can't go that granular.
    - Pre-trained language models may not be relevant (context, language).
    - Interpretability a challenge - may not give well-suited metrics.

Why not use established NLP tools?

Many NLP methods we could use that broadly fall into two categories:

1. Unsupervised (e.g. topic models, embeddings):
    - Explain variation in text: not geared towards question.
    - In small samples, we can't go that granular.
    - Pre-trained language models may not be relevant (context, language).
    - Interpretability a challenge - may not give well-suited metrics.
2. Supervised
    - Train model to predict a label (e.g. movie review text to predict stars).
    - Interpretable labels often not available, e.g. crowd-sourcing larger samples unreliable for complex coding trees.
    - Training model on a labelled dataset from another context may be a bad match.
    - Dictionary-based methods (e.g. sentiment) unsuited to question and/or have a WEIRD bias.

Our contribution

Traditional qualitative analysis on a sub-sample to develop a coding tree and get high quality "gold standard" annotations.

Use NLP to predict annotations on larger sample.

Advantages:

- Careful, expert reading: structure comes from the data
- Specialised coding gives a training set suited to question
- Stronger signals offset smaller dataset
- Interpretability built in!

Drawbacks:

- Introduces an additional source of measurement error...
- Units for qualitative annotations often not clearly comparable

▸ Literature

Relevant literature

▸ Back

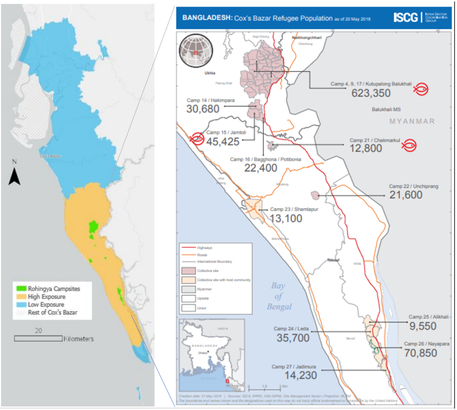
- Use of manually coded text data
  [Michalopoulos and Xue, 2021, Jayachandran et al., 2021]
- Unsupervised methods
  [Parthasarathy et al., 2019, Hansen et al., 2018, Larsen et al., 2021]
- Dictionary methods [Loughran and McDonald, 2011, Apel and Grimaldi, 2012, Baker et al., 2016, Shapiro et al., 2020]
- Expanding manually classified sub-sample
  [Mann and Püttmann, 2018, Yordanova et al., 2019]
- Assisting human annotation with NLP
  [Wiedemann, 2019, Karamshuk et al., 2017, Chen et al., 2018]

Data

- Rohingya refugees in the Cox's Bazaar camp and local Bangladeshi residents. Context
- **2019.** Baseline survey (finances, education, gender, age, trauma) for all members of 5,020 households. Details
- Two rounds of open-ended interviews on aspirations
  - **2020.** R2: 1,020 10-min interviews with household head on aspirations
  - **2021.** R3: 2,040 30-min interviews with household head on aspirations, belonging and well-being.
- 400 from each round annotated according to a coding tree with 25 hierarchical categories (by MB, AH and AK).

Context ▸ Back

- Around 725,000 Rohingya have entered Bangladesh since 2017.
- Doubled population in Cox's Bazar upazilas of Teknaf and Ukhia.
- Rohingya are younger than hosts and males under-represented.
- 62% of Rohingya have no schooling (33% of hosts)
- 58% of primary school aged Rohingya children in education (95% of host children).



Questions

Well-being (R3 only):

- In your opinion, what is a good life?
- How is your life at the moment?
- Why do you think are you in this state?
- Describe the happiest moment you experienced most recently?
- Describe the saddest moment you experienced most recently?
- What measures could the government or other authority take to help you improve the quality of your life?

Aspirations:

- **What are your other hopes and dreams for {eld_child_name}?**
- **What steps are you taking to fulfill your children's dreams?**

Belonging (R3 only):

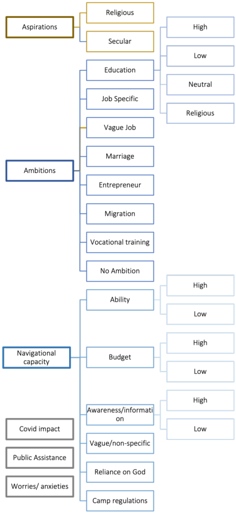
- What words would you use to describe hosts/refugees (H/R)?
- Tell me about a time where you felt H/R benefited your life?
- Tell me about a time where you felt that H/R negatively affected you?
- How can the government help with improving the relationships between H and R?

Ambition, Aspirations and Navigational Capacity

- Large literature on aspirations and development [Genicot and Ray, 2020, Fruttero et al., 2021]
- This largely focuses on "ambition" rather than "aspiration" [Callard, 2018]:
  - **Ambition** - material goals that you want to achieve
  - **Aspiration** - transformative goals that change values, and how one sees and conducts oneself in life
- The important idea of the "capacity to aspire" [Appadurai et al., 2004] is also neglected:
  - **Navigational Capacity** - capacity of individuals to conceptualize goals, and to have the capability to navigate their way towards them
- Very salient for policy, if they prove to be important

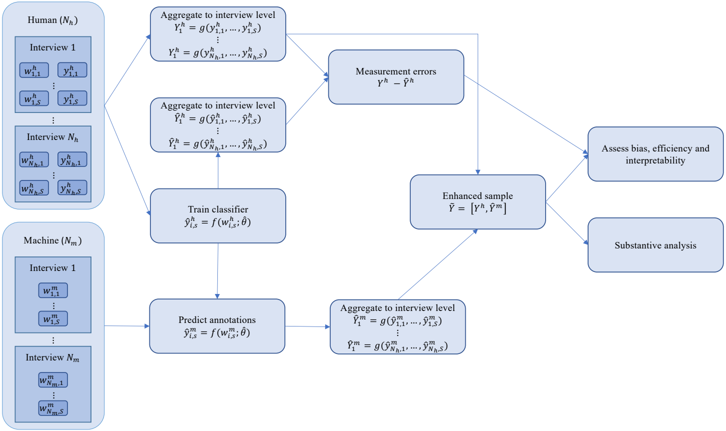
Qualitative analysis

▸ Summary statistics



- **Ambition.Education.Low**: "God willing, I will teach my son up to 10th class."
- **Ambition.Education.High**: "My daughter's dream is to study ... I will educate my daughter so she can get a job in administration"
- **Capacity.Ability.Low**: "I don't do much at home. I help her as much as I can."
- **Capacity.Ability.High**: "The school is still closed for Corona. So, by selling some of my food, I have arranged for private teacher by paying at minimum."
- **Aspirations.Secular**: "They will become well behaved, good human beings. Will have a respectable job."
- **Aspirations.Religious**: "I don't want make my son work. I want him to become a religious cleric (hujur).."

Methodology



▸ Differences from paper

Modelling choices

- Text representation $w_{i,s}$ (unsupervised):
  - ngram tf-idf vectors, embeddings, translations, topics, clusters, sentiment, include questions ...
- Classifier $f()$:
  - pre-filter observations
  - Cross-validate model type and hyperparameters ▸ Options
  - Most successful are logistic regression, random forest and stochastic gradient descent classifier ▸ Details
- Aggregation $g()$:
  - Prediction or probability
  - Maximum or mean
- Uncertainty
  - Bootstrap by sampling without replacement
  - Quantify model error and prediction error

Modelling choices ▸ Back

Features:

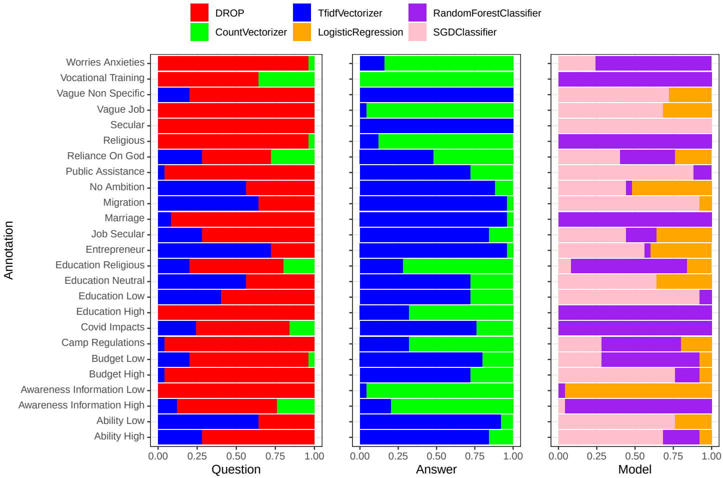
- tf-idf vectors
- tf vectors
- with/without questions
- unigram/bigrams
- maximum number of features
- minimum/maximum document frequency

Models:

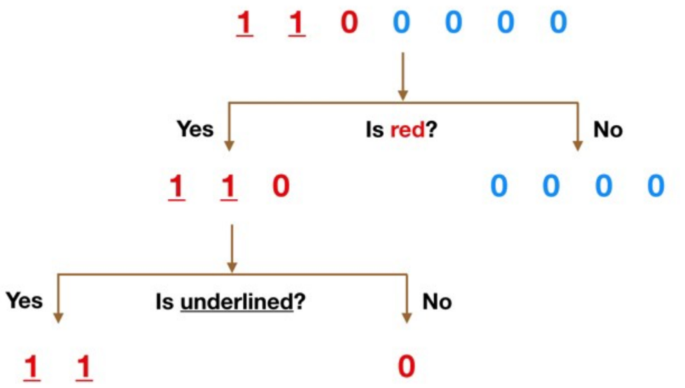
- Logistic regression (L1/L2 regularisation, penalty parameter)
- Random Forest (Number of estimators, maximum depth)
- SGDClassifier (regularisation parameter α, loss function)
- Also available: key phrase matching, KNN, linear SVC, MLP.
- In all cases, cross validate threshold probability

Models selected

Back



Random Forest



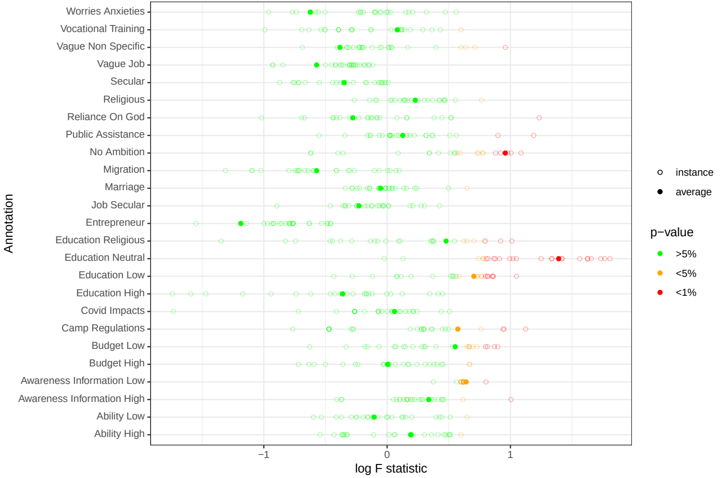
What we care about

- Bias ▸ Details:
  - Use prediction errors to test for bias with respect to HH characteristics
  - Include dummy variables for annotation status in analysis
- Efficiency ▸ Details:
  - Bootstrap to quantify model and prediction errors
  - Depends on use case, change in sample size and error magnitude
- Interpretability:
  - Validation set performance ▸ Details
  - Supervised topic model on predictions ▸ Details
  - Regression F-statistics on HH characteristics

Bias

Back

Are the prediction errors correlated to the household characteristics?



Efficiency ▸ Back

Two types of measurement error in the enhanced sample:

1. Model error $\sigma^2_{\hat{y}}$ - estimate by bootstrapping estimates
2. Prediction error σ^2_{ϵ} - estimate on held-out observations from each bootstrap (out-of-bag error)

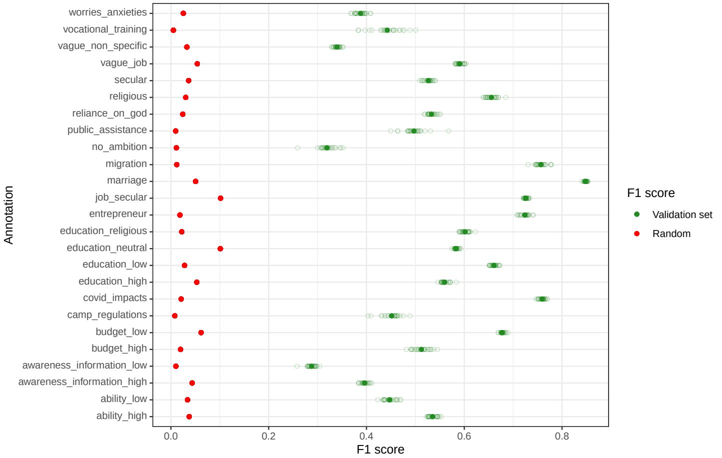
Is the extra noise worth it? That depends...

For example, standard error of mean in the enhanced sample is weighted average of variance in human and machine samples.

Standard error is smaller in enhanced sample if

$$\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_h} < \frac{N_m + 2N_h}{N_h}$$

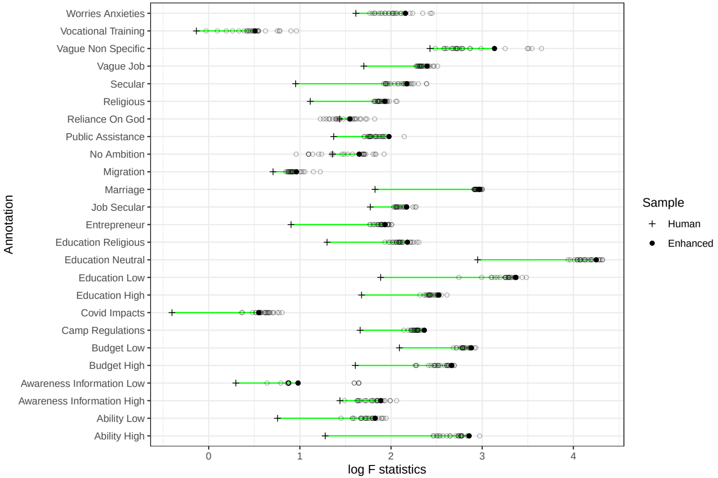
Validation set performance ▸ Back



Interpretability

Other specs Performance Bias

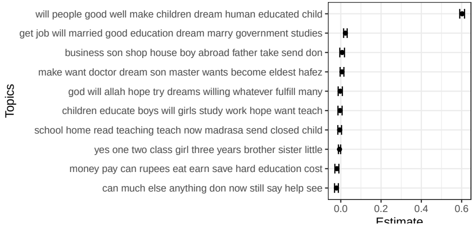
Assuming text-based variables should be related to household characteristics, compare F statistics for human and enhanced samples.



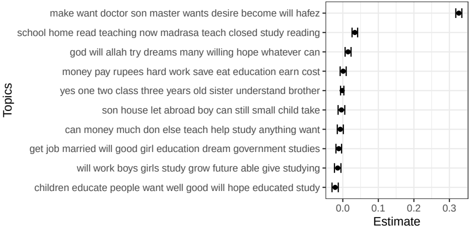
Interpretability ▸ Back

Topics

Aspirations:Secular



Aspirations:Religious



Selected results

Relate annotations to household characteristics

- **Ambition:** lower ambitions for refugees and female eldest children, refugees more likely to show no or neutral ambitions. Qualitative variables can add richness to quantitative measures. ▸ Results
- **Capacity:** Refugees less likely to show low ability, reliance on God, give vague answers, and mention lack of money ▸ Results
- **Aspirations:** More educated parents show more secular aspirations, and parents express fewer aspirations for female children ▸ Results
- Religious aspirations less likely to get/keep child into school between rounds

Education status

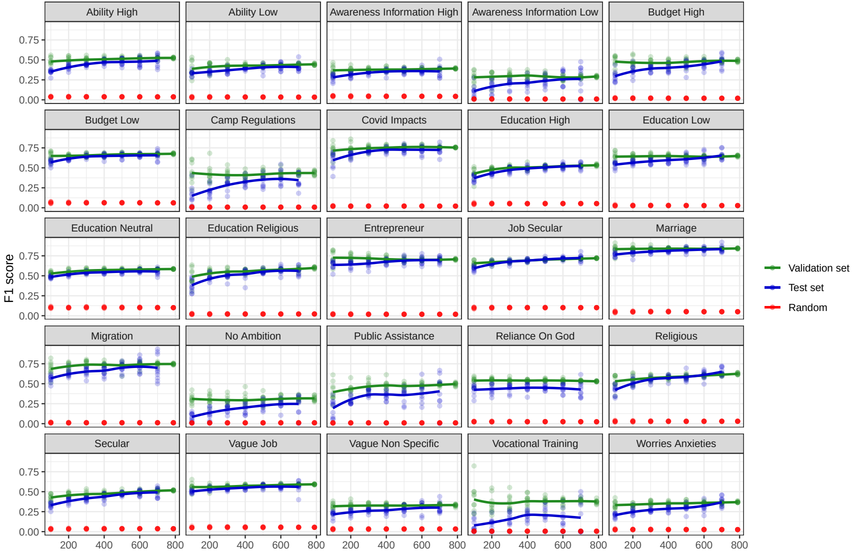
- Subjects asked in both rounds whether eldest child is currently in school
- Do interview codes predict changes in education status between rounds?
- Religious aspirations and low budget in R2 associated with negative outcomes in R3

	Δedu_status	
	(Human)	(Enh)
secular	-0.331	-0.403^{*}
religious	-1.024^{**}	-0.506^{**}
budget_high	0.201	0.211
budget_low	-0.599^{**}	-0.592^{***}
aware_info_high	0.297	0.113
aware_info_low	-2.498^{**}	-1.449^{*}
refugee	-0.047	0.002
num_child	-0.030	0.023
hh_head_sex	0.074	0.027
hh_head_age	-0.0004	-0.002
parent_eduyears	0.012	0.022^{***}
parent_reledu	0.176	0.060
eld_sex	-0.040	-0.075
eld_age	0.004	0.004
hh_asset_index	-0.052	-0.032^{*}
hh_income	0.018	0.014
machine_annot		-0.070
Constant	-0.030	-0.150^{*}
Observations	248	640
F Statistic	1.470	2.855^{***}

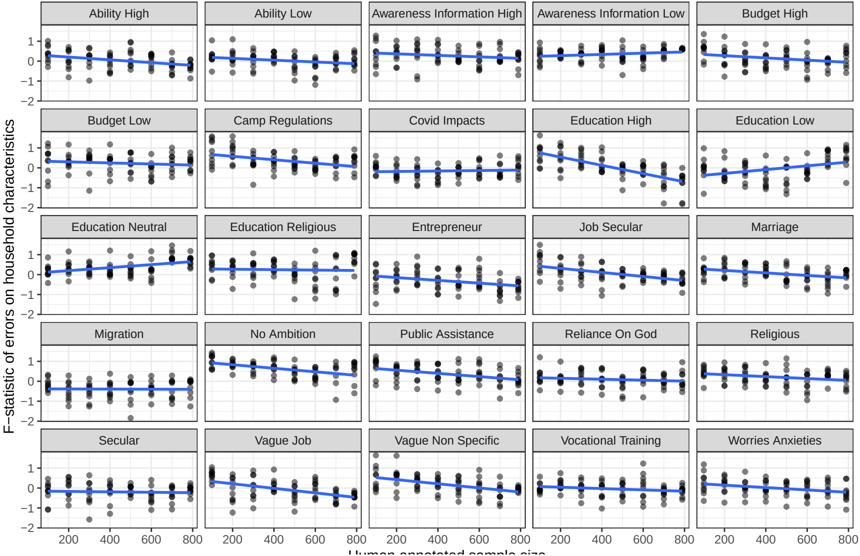
Boundaries Membership and Belonging

	Refugee		Host		
	mean	s.d.	mean	s.d.	
suspicion	0.003	0.012	0.011	0.030	Results
empathy/sympathy	0.010	0.031	0.013	0.034	Results
humans are equal	0.002	0.012	0.004	0.023	Results
expressions of gratitude	0.119	0.123	0.005	0.018	Results
prejudice	0.002	0.010	0.007	0.023	Results
quarrel/dispute	0.013	0.036	0.013	0.041	Results
fear/concern	0.010	0.028	0.030	0.062	Results
security concern	0.017	0.043	0.046	0.085	Results
reduced wages	0.001	0.007	0.009	0.036	Results
perceived lack in morals	0.005	0.016	0.009	0.032	Results
more aid	0.002	0.010	0.020	0.051	Results

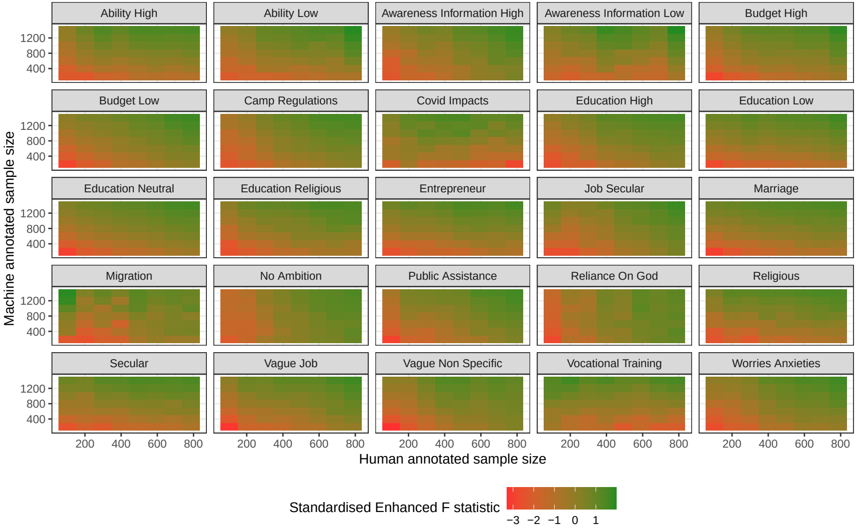
Performance for increasing N_h



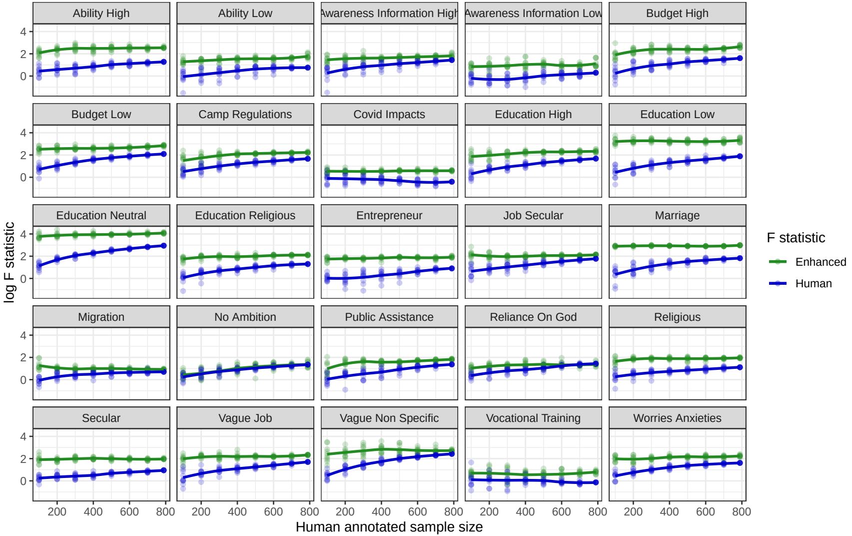
Bias not really affected by N_h



Interpretability increases with both N_h and N_m



Interpretability increases with N_h, holding N fixed



More labels or more interviews?

Collecting and labelling more data may be possible, but what should we focus on?

- 100 additional human annotated interviews increases the enhanced sample F statistic by 8.4% on average,
- 100 additional machine annotated interviews increases it by 6.1% on average.
- So annotating 100 existing interviews increases the F statistic by 2.3%.

Of course, these effects will be non-linear and vary greatly from one context to the next.

Conclusions and Next Steps

Conclusions:

- NLP has promising application to bridge the gap between qualitative work and quantitative social science.
- New method with very general applications of expanding human coding to larger samples.
- For this paper: important differences in navigational capacity and aspiration across households - hard to capture without open-ended interviews.

Next steps:

- Expand focus to well-being and belonging
- Python package coming soon

References I

Apel, M. and Grimaldi, M. (2012).
The information content of central bank minutes.
Riksbank Research Paper Series No. 92.

Appadurai, A., Vijayendra, R., and Michael, W. (2004).
The capacity to aspire: Culture and the terms of recognition.
Culture and Public Action, ed. Vijayendra Rao and Michael Walton, Stanford, California: Stanford University Press, pages 59–84.

Baker, S. R., Bloom, N., and Davis, S. J. (2016).
Measuring economic policy uncertainty.
The Quarterly Journal of Economics, 131(4):1593–1636.

Callard, A. (2018).
Aspiration: The agency of becoming.
Oxford University Press.

Chen, N.-C., Drouhard, M., Kocielnik, R., Suh, J., and Aragon, C. R. (2018).
Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity.
ACM Transactions on Interactive Intelligent Systems (TiiS), 8(2):1–20.

Fruttero, A., Muller, N., and Calvo-Gonzalez, O. (2021).
The power and roots of aspirations.
Technical report.

References II

Genicot, G. and Ray, D. (2020).
Aspirations and economic behavior.
Annual Review of Economics, 12.

Griffiths, T. L. and Steyvers, M. (2004).
Finding scientific topics.
Proceedings of the National Academy of Sciences, 101(suppl 1):5228–5235.

Hansen, S., McMahon, M., and Prat, A. (2018).
Transparency and deliberation within the fomc: a computational linguistics approach.
The Quarterly Journal of Economics, 133(2):801–870.

Jayachandran, S., Biradavolu, M., and Cooper, J. (2021).
Using machine learning and qualitative interviews to design a five-question women's agency index.
Technical report, Northwestern Global Poverty Research Lab Working Paper.

Karamshuk, D., Shaw, F., Brownlie, J., and Sastry, N. (2017).
Bridging big data and qualitative methods in the social sciences: A case study of twitter responses to high profile deaths by suicide.
Online Social Networks and Media, 1:33–43.

Larsen, V. H., Thorsrud, L. A., and Zhulanova, J. (2021).
News-driven inflation expectations and information rigidities.
Journal of Monetary Economics, 117:507–520.

References III

Loughran, T. and McDonald, B. (2011).
When is a liability not a liability? textual analysis, dictionaries, and 10-ks.
The Journal of Finance, 66(1):35–65.

Mann, K. and Püttmann, L. (2018).
Benign effects of automation: New evidence from patent texts.
Available at SSRN 2959584.

Michalopoulos, S. and Xue, M. M. (2021).
Folklore.
The Quarterly Journal of Economics, 136(4):1993–2046.

Parthasarathy, R., Rao, V., and Palaniswamy, N. (2019).
Deliberative democracy in an unequal world: A text-as-data study of south india's village assemblies.
American Political Science Review, 113(3):623–640.

Roberts, M. E., Stewart, B. M., Tingley, D., Airoldi, E. M., et al. (2013).
The structural topic model and applied social science.
In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*, volume 4. Harrahs and Harveys, Lake Tahoe.

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2020).
Measuring news sentiment.
Journal of Econometrics.

References IV

Wiedemann, G. (2019).
Proportional classification revisited: Automatic content analysis of political manifestos using active learning.
Social Science Computer Review, 37(2):135–159.

Yordanova, K. Y., Demiray, B., Mehl, M. R., and Martin, M. (2019).
Automatic detection of everyday social behaviours and environments from verbatim transcripts of daily conversations.
In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, pages 1–10. IEEE.

Quant data ▸ Back

Table

Statistic	N	Mean	St. Dev.	Min	Max
refugee	2,407	0.460	0.498	0	1
eld_sex	2,294	0.525	0.499	0	1
eld_age	2,295	24.936	14.285	3	80
hh_head_sex	2,407	0.185	0.388	0	1
num_child	2,405	2.675	1.463	0	9
parent_eduyears	2,398	3.548	3.837	0	11
parent_reledu	2,398	0.036	0.186	0	1
hh_asset_index	2,406	0.147	1.820	−2.633	8.191
hh_income	2,407	1.125	2.340	0.000	60.000
int_trauma_exp	2,287	2.641	2.410	0	11

Annotations in each round ▸ Back

annot	Mean			Std dev			Prop annot		
	r2	r3	pool	r2	r3	pool	r2	r3	pool
religious	0.036	0.028	0.030	0.185	0.164	0.171	0.208	0.234	0.230
secular	0.053	0.028	0.036	0.224	0.166	0.187	0.332	0.332	0.360
no.ambition	0.021	0.013	0.016	0.144	0.113	0.124	0.129	0.129	0.137
vague.job	0.081	0.041	0.054	0.272	0.199	0.226	0.430	0.386	0.421
vocational.training	0.007	0.004	0.005	0.086	0.062	0.071	0.041	0.041	0.045
entrepreneur	0.031	0.012	0.018	0.173	0.111	0.134	0.154	0.122	0.150
education.low	0.013	0.035	0.028	0.114	0.183	0.164	0.094	0.475	0.312
education.neutral	0.185	0.062	0.101	0.388	0.240	0.301	0.772	0.574	0.691
education.high	0.064	0.048	0.053	0.245	0.213	0.224	0.375	0.454	0.427
education.religious	0.035	0.016	0.022	0.184	0.125	0.147	0.210	0.168	0.198
marriage	0.082	0.036	0.050	0.274	0.185	0.219	0.385	0.396	0.418
migration	0.022	0.007	0.012	0.147	0.084	0.109	0.104	0.079	0.097
vague.non.specific	0.066	0.017	0.033	0.248	0.129	0.178	0.420	0.234	0.349
reliance.on.god	0.039	0.017	0.024	0.194	0.130	0.154	0.243	0.228	0.253
ability.high	0.048	0.032	0.037	0.215	0.177	0.190	0.311	0.424	0.391
ability.low	0.035	0.033	0.034	0.185	0.180	0.181	0.230	0.360	0.321
budget.high	0.033	0.013	0.020	0.178	0.115	0.139	0.200	0.188	0.212
budget.low	0.111	0.039	0.062	0.314	0.193	0.241	0.522	0.401	0.492
awareness.information.high	0.070	0.031	0.043	0.255	0.173	0.204	0.387	0.297	0.367
awareness.information.low	0.007	0.012	0.010	0.086	0.107	0.101	0.061	0.145	0.114
camp.regulations	0.016	0.004	0.008	0.124	0.065	0.089	0.094	0.066	0.088
covid.impacts	0.022	0.021	0.021	0.145	0.143	0.144	0.149	0.272	0.226
public.assistance	0.020	0.005	0.010	0.141	0.071	0.099	0.137	0.058	0.107
worries.anxieties	0.049	0.014	0.025	0.216	0.118	0.157	0.268	0.175	0.242

Ambition ▸ Back

	Dependent variable:			
	eld_edu_ambition			
	(Human)	(Enh)	(Human)	(Enh)
annotation_statusunannotated		0.217**		0.092
no_ambition	−10.972***	−9.467***	−4.469*	−5.050***
job_secular	2.663***	1.552***	1.793***	1.155***
vocational_training	−3.879*	−3.336**	−2.416	−2.429*
entrepreneur	0.550	−0.663	−1.615*	−0.165
education_low	−4.184***	−5.357***	−1.429	−3.303***
education_neutral	−0.860*	−0.587**	−0.015	0.143
education_high	3.636***	3.691***	2.606***	2.197***
education_religious	−3.264***	−3.328***	−1.385	−1.692***
marriage	−1.853***	−1.777***	−1.911***	−1.793***
migration	−0.045	−1.071	1.560	−0.348
HH characteristics			✓	✓
Constant	4.039***	4.172***	3.587***	3.911***
Observations	426	1,267	392	1,184
R^2	0.286	0.204	0.515	0.465
F Statistic	16.610***	29.308***	18.672***	45.835***
Note:				*p<0.1; **p<0.05; ***p<0.01

Capacity ▸ Back

	Dependent variable:							
	AbilityLow		RelianceOnGod		VagueNonSpecific		BudgetLow	
	(Human)	(Enh)	(Human)	(Enh)	(Human)	(Enh)	(Human)	(Enh)
refugee	−0.015*	−0.018***	−0.028***	−0.014***	−0.031***	−0.011***	−0.036***	−0.030***
num_child	0.004*	0.004***	−0.001	−0.001	−0.003	−0.001	0.007**	0.006***
hh_head_sex	0.006	0.004	0.010*	0.002	−0.008	−0.002	0.019*	0.004
hh_head_age	0.0003	0.0001	−0.0004	−0.0002	−0.0002	−0.0003*	0.0001	0.0001
parent_eduyears	−0.001*	−0.002***	−0.002**	−0.001***	−0.001	0.0003	−0.003***	−0.003***
parent_reledu	−0.016	−0.017**	−0.011	−0.003	0.004	0.006	0.006	−0.001
eld_sex	−0.006	0.003	0.007	0.003	0.003	−0.001	0.008	0.008*
eld_age	−0.0003	0.0001	0.0001	0.0001	−0.001	−0.0001	0.0003	0.0003
hh_asset_index	−0.005*	−0.004***	−0.003	−0.0001	0.001	0.001	−0.005	−0.006***
hh_income	−0.002	−0.001	0.001	0.0001	−0.001	−0.001	−0.004	−0.002*
int_trauma_exp	−0.002**	−0.001*	0.0001	0.0003	−0.0002	0.0003	0.001	−0.0002
roundR3	0.004	−0.005	−0.023**	−0.014***	−0.037***	−0.040***	−0.070***	−0.062***
machine_annot		0.002		0.0002		0.003		−0.003
Constant	0.041***	0.041***	0.069***	0.052***	0.116***	0.084***	0.106***	0.099***
Observations	696	2,177	696	2,177	696	2,177	696	2,177
R^2	0.036	0.036	0.069	0.028	0.166	0.121	0.125	0.097
F Statistic	2.127**	6.205***	4.210***	4.705***	11.330***	22.981***	8.104***	17.841***

Aspiration ▸ Back

	Dependent variable:			
	Secular		Religious	
	(Human)	(Enh)	(Human)	(Enh)
refugee	−0.005	−0.005	−0.00002	−0.002
num_child	0.002	−0.001	−0.002	−0.001
hh_head_sex	0.002	−0.002	−0.012	0.001
hh_head_age	−0.0005	−0.0004**	0.001*	0.0002
parent_eduyears	0.002*	0.002***	−0.0002	−0.001
parent_reledu	0.003	0.008	0.015	0.017*
eld_sex	−0.013*	−0.012***	−0.021***	−0.022***
eld_age	0.0002	0.00003	−0.001**	−0.001***
hh_asset_index	−0.001	−0.001	−0.005*	−0.002
hh_income	0.0002	0.001	−0.001	−0.001
int_trauma_exp	−0.001	0.00003	0.0002	0.001
roundR3	−0.035**	−0.029***	0.016	0.012
machine_annot		0.008**		0.006*
Constant 0.074***	0.078***	0.043***	0.060***	
Observations	696	2,177	696	2,177
R^2	0.044	0.050	0.051	0.040
F Statistic	2.595***	8.790***	3.046***	6.893***
Note:			*p<0.1; **p<0.05; ***p<0.01	

Suspicion ▸ Back

	suspicion					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.011**	-0.006***				
distancefromcamp					-0.197	-0.055
human_annot		0.005***		0.001		0.009***
int_sex	-0.005	0.000	0.001	0.000	-0.011	0.001
int_age	0.000	0.000	-0.000	-0.000	0.000	0.000
int_eduyears	-0.000	0.000	-0.000	0.000	0.000	0.000
int_reledu	0.010	0.006*	0.002	0.003*	0.050*	0.012*
int_trauma_exp	-0.003***	-0.001***	-0.001*	-0.000**	-0.005**	-0.002***
Constant	0.025***	0.009***	0.008	0.004***	0.032*	0.008*
Observations	386	1941	191	961	195	980
R2	0.060	0.043	0.023	0.010	0.067	0.030
F-statistic	4.015***	12.432***	0.858	1.615	2.233**	4.356***

Empathy ▸ Back

	empathy_sympathy					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	0.003	-0.001				
distancefromcamp					0.257	0.104**
human_annot		0.011***		0.015***		0.008***
int_sex	-0.002	-0.000	-0.007	-0.001	0.003	0.001
int_age	-0.000	0.000	0.000	0.000	-0.001	0.000
int_eduyears	0.001	0.001**	0.001	0.001**	-0.000	0.000
int_reledu	0.012	0.005	0.010	0.005	0.030	0.005
int_trauma_exp	0.001	0.000	-0.001	-0.000	0.004**	0.001*
Constant	0.016	0.006*	0.018	0.006	0.021	0.003
Observations	386	1941	191	961	195	980
R2	0.008	0.024	0.017	0.044	0.050	0.018
F-statistic	0.526	6.654***	0.633	7.374***	1.638	2.474**

Equality ▸ Back

	humans_are_equal					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.003	-0.001				
distancefromcamp					0.135	0.047
human_annot		0.003***		0.002**		0.005**
int_sex	0.002	0.000	0.000	0.000	0.003	0.000
int_age	0.000***	0.000***	0.000*	0.000	0.001**	0.000***
int_eduyears	0.000	0.000	-0.000	0.000	0.001	0.000*
int_reledu	0.025***	0.005**	0.010	0.002	0.088***	0.013**
int_trauma_exp	0.000	-0.000	-0.000	-0.000	0.001	-0.000
Constant	-0.012	-0.001	-0.005	0.001	-0.027**	-0.007**
Observations	386	1941	191	961	195	980
R2	0.048	0.015	0.032	0.008	0.117	0.022
F-statistic	3.175***	4.283***	1.213	1.262	4.167***	3.166***

Gratitude ▸ Back

	expressions_of_gratitude					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	0.120***	0.118***				
distancefromcamp					-0.159	-0.075***
human_annot		0.010*		0.015		0.004***
int_sex	-0.031**	-0.013***	-0.065***	-0.029***	-0.003	0.001
int_age	-0.000	0.000**	-0.000	0.001*	-0.000	0.000*
int_eduyears	-0.001	0.001	-0.005	0.000	0.000	0.000
int_reledu	0.003	0.012	-0.004	0.018	-0.003	-0.003
int_trauma_exp	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001**
Constant	0.051*	-0.004	0.196***	0.114***	0.025**	0.004
Observations	386	1941	191	961	195	980
R2	0.244	0.305	0.039	0.021	0.034	0.024
F-statistic	20.442***	121.125***	1.510	3.334***	1.112	3.415***

Prejudice ‣ Back

	prejudice					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.010***	-0.005***				
distancefromcamp					-0.426***	-0.108***
human_annot		0.002*		-0.001		0.005***
int_sex	0.000	-0.001	0.003	0.000	-0.004	-0.002
int_age	-0.000	0.000	0.000	-0.000	-0.000	0.000
int_eduyears	0.000	0.000	-0.000	0.000	0.001	0.000**
int_reledu	0.010	0.002	-0.002	-0.002	0.071***	0.011**
int_trauma_exp	0.000	-0.000	0.001**	0.000	-0.000	-0.000
Constant	0.013	0.007***	-0.006	0.002*	0.031**	0.007**
Observations	386	1941	191	961	195	980
R2	0.032	0.025	0.032	0.004	0.091	0.026
F-statistic	2.122*	6.937***	1.227	0.711	3.126***	3.760***

Quarrel

▸ Back

	quarrel_dispute					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	0.006	0.003				
distancefromcamp					-0.645***	-0.156***
human_annot		0.000		0.001		-0.001
int_sex	-0.007	-0.003*	0.002	0.002	-0.015**	-0.008***
int_age	-0.000	0.000	-0.000	0.000	0.000	0.000
int_eduyears	0.001	0.000	0.001	0.001	0.002**	0.001*
int_reledu	0.007	0.010**	0.010	0.017***	0.005	-0.004
int_trauma_exp	-0.002*	-0.001**	-0.001	-0.000	-0.003*	-0.002***
Constant	0.023*	0.011***	0.029*	0.009**	0.032*	0.019***
Observations	386	1941	191	961	195	980
R2	0.018	0.009	0.016	0.011	0.132	0.028
F-statistic	1.139	2.384**	0.584	1.772	4.764***	3.992***

Fear ▸ Back

	fear_concern					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.021***	-0.018***				
distancefromcamp					-0.359	-0.421***
human_annot		0.005*		0.002		0.008
int_sex	-0.010	-0.007***	0.010*	0.002	-0.027**	-0.016***
int_age	0.000	0.000	-0.000	0.000	0.001	0.000
int_eduyears	-0.000	0.001**	0.000	0.001**	0.001	0.002***
int_reledu	0.011	0.003	-0.005	0.003	0.103**	0.006
int_trauma_exp	-0.002	-0.000	-0.001	-0.000	-0.002	-0.001
Constant	0.034**	0.029***	0.014	0.006**	0.037	0.042***
Observations	386	1941	191	961	195	980
R2	0.051	0.052	0.029	0.007	0.080	0.044
F-statistic	3.426***	15.019***	1.097	1.167	2.716**	6.451***

Security ▸ Back

	security_concern					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.041***	-0.025***				
	(0.011)	(0.004)				
distancefromcamp					-0.607	-0.674***
human_annot		0.009**		-0.001		0.017***
int_sex	-0.006	-0.013***	0.001	-0.007**	-0.010	-0.018***
int_age	-0.000	0.000	-0.000	0.000	0.000	0.000
int_eduyears	-0.001	0.001	0.000	0.000	-0.000	0.002**
int_reledu	-0.000	-0.002	0.016	0.007	-0.056	-0.013
int_trauma_exp	-0.003	-0.002**	-0.002	-0.001	-0.004	-0.003**
Constant	0.078***	0.048***	0.034**	0.020***	0.089**	0.063***
Observations	386	1941	191	961	195	980
R2	0.059	0.060	0.022	0.012	0.025	0.055
F-statistic	3.959***	17.604***	0.839	1.858*	0.799	8.115***

Wages ▸ Back

	reduced_wages					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.019***	-0.012***				
distancefromcamp					-0.491***	-0.322***
human_annot		0.003*		0.000		0.005
int_sex	-0.001	-0.000	0.001	0.000	-0.002	0.002
int_age	-0.000	-0.000*	0.000	0.000	-0.001	-0.000
int_eduyears	-0.002***	-0.001***	-0.000	-0.000	-0.002**	-0.001***
int_reledu	-0.008	-0.007**	-0.001	-0.001	-0.010	-0.016**
int_trauma_exp	0.001	0.000	0.000	-0.000	0.003*	0.000
Constant	0.028***	0.017***	-0.002	0.001	0.059***	0.031***
Observations	386	1941	191	961	195	980
R2	0.054	0.043	0.008	0.001	0.100	0.070
F-statistic	3.625***	12.327***	0.302	0.126	3.466***	10.394***

Morals ▸ Back

	perceived_lack_in_morals					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.010**	-0.004***				
distancefromcamp					-0.190	-0.031
human_annot		0.004***		0.001		0.008***
int_sex	-0.008*	-0.003**	-0.004	-0.001	-0.013	-0.005**
int_age	-0.000	0.000	-0.000	0.000	-0.000	0.000
int_eduyears	-0.000	0.000	0.000	0.000	0.000	0.000
int_reledu	0.036***	0.013***	0.004	0.005*	0.194***	0.032***
int_trauma_exp	-0.001	-0.000	-0.001	-0.000	-0.001	0.000
Constant	0.027**	0.008***	0.020**	0.006***	0.030	0.005
Observations	386	1941	191	961	195	980
R2	0.049	0.023	0.027	0.007	0.180	0.038
F-statistic	3.265***	6.479***	1.031	1.109	6.891***	5.434***

More aid ▸ Back

	more_aid					
	All		Refugee		Host	
	Human	Enh.	Human	Enh.	Human	Enh.
refugee	-0.028***	-0.020***				
distancefromcamp					-1.025***	-0.562***
human_annot		0.004*		-0.000		0.008*
int_sex	-0.009*	-0.002	0.003	0.001**	-0.018**	-0.004
int_age	-0.000	-0.000	0.000	0.000	-0.000	-0.000
int_eduyears	-0.001*	-0.000*	-0.000	0.000	-0.000	0.000
int_reledu	-0.010	-0.001	-0.002	-0.001	-0.009	0.003
int_trauma_exp	-0.000	-0.000	0.001*	0.000**	-0.001	-0.000
Constant	0.040***	0.025***	-0.005	-0.000	0.083***	0.043***
Observations	386	1941	191	961	195	980
R2	0.082	0.064	0.040	0.010	0.151	0.065
F-statistic	5.650***	18.788***	1.546	1.556	5.577***	9.719***