

The Plain Language Adaptation of Biomedical Abstracts (PLABA) track @ TREC

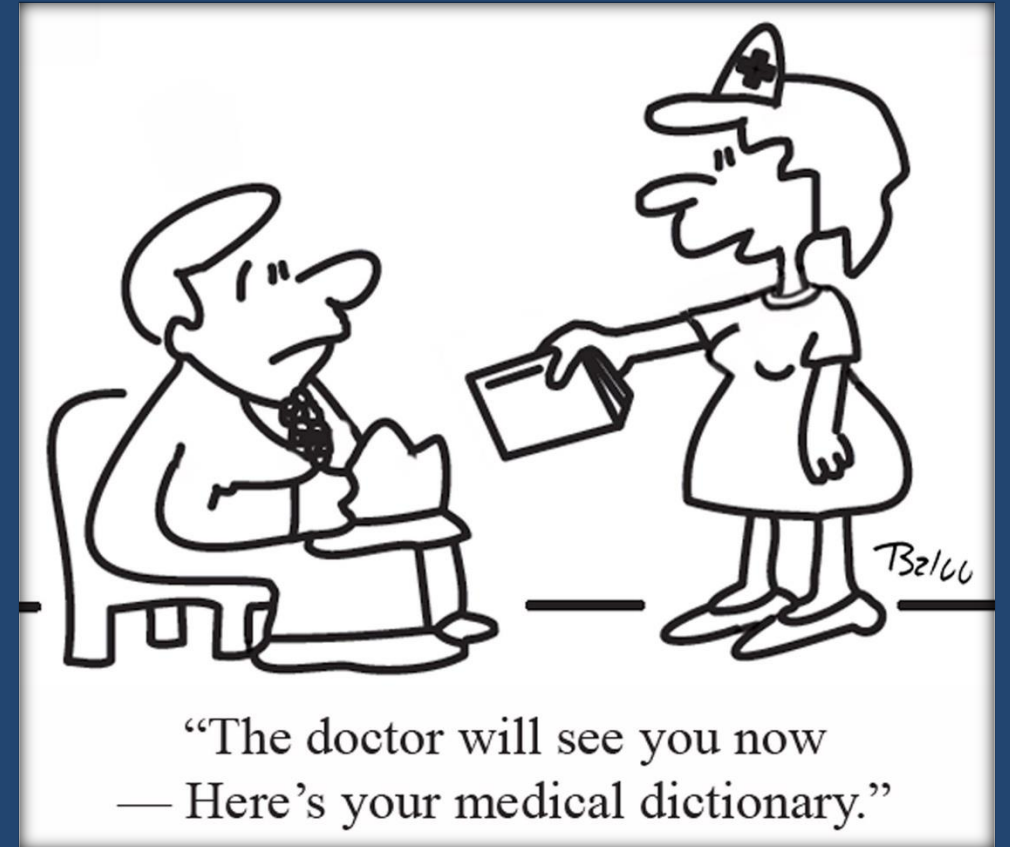
Brian Ondov, PhD

Department of Biomedical Informatics & Data Science

Yale SCHOOL OF MEDICINE

Outline

1. Why Rewrite Abstracts?
2. Creating a High-quality Dataset
3. Sentence-level Evaluation
4. Baselines and Team Systems
5. Manual and Automatic Results
6. Assessing Automatic Metrics
7. New Subtask: Term Replacement



1. Why Rewrite Abstracts?

Health Literacy

Annals of Internal Medicine

REVIEW

Low Health Literacy and Health Outcomes: An Updated Systematic Review

Nancy D. Berkman, PhD; Stacey L. Sheridan, MD, MPH; Katrina E. Donahue, MD, MPH; David J. Halpern, MD, MPH; and Karen Crotty, PhD, MPH

Conclusion: Low health literacy is associated with poorer health outcomes and poorer use of health care services.

“Cyberchondria”

“...escalation is associated with the amount and distribution of medical content viewed by users, the presence of escalatory terminology in pages visited, and a user's predisposition to escalate versus to seek more reasonable explanations for ailments...”

White RW, Horvitz E. Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems (TOIS)*. 2009 Nov 30;27(4):1-37.

merckmanuals.com

MERCK MANUAL Professional Version

HOME MEDICAL TOPICS RESOURCES DRUG INFO NEWS PROCEDURES

PROFESSIONAL / ... / ATRIAL FIBRILLATION

Atrial Fibrillation

(A Fib)

By **L. Brent Mitchell**, MD, Libin Cardiovascular Institute of Alberta, University of Calgary

Last full review/revision Jan 2021

[CLICK HERE FOR PATIENT EDUCATION](#)

Atrial fibrillation is a rapid, irregularly irregular atrial rhythm. Symptoms include palpitations and sometimes weakness, effort intolerance, dyspnea, and presyncope. Atrial thrombi may form, causing a significant risk of embolic stroke. Diagnosis is by electrocardiography. Treatment involves rate control with drugs, prevention of thromboembolism with anticoagulation, and sometimes conversion to sinus rhythm by drugs or cardioversion.

(See also [Overview of Arrhythmias.](#))

merckmanuals.com

MERCK MANUAL Consumer Version

HOME HEALTH TOPICS SYMPTOMS EMERGENCIES DRUG INFO RESOURCES

HOME / ... / ATRIAL FIBRILLATION AND ATRIAL FLUTTER

Atrial Fibrillation and Atrial Flutter

By **L. Brent Mitchell**, MD, Libin Cardiovascular Institute of Alberta, University of Calgary

Last full review/revision Feb 2021 | Content last modified Feb 2021

[CLICK HERE FOR THE PROFESSIONAL VERSION](#)

[GET THE QUICK FACTS](#)

Atrial fibrillation and atrial flutter are very fast electrical discharge patterns that make the atria (upper chambers of the heart) contract very rapidly, with some of the electrical impulses reaching the ventricles and sometimes causing them to contract faster and less efficiently than normal.

- These disorders often result from conditions that cause the atria to enlarge.
- Symptoms depend on how fast the ventricles contract and may include palpitations, weakness, dizziness or light-headedness, shortness of breath

Search
How long will it take to achieve good vision after retina detachment surgery?

PMID 17011955

...at 1 day, 2 months, and last follow-up examination...visual recovery after retinal reattachment was most dependent on macular involvement..

PMID 26943151

...evaluate best-corrected VA (BCVA) at 1, 2, 3, 6, 9, and 12 months after vitrectomy...best-corrected visual acuity significantly improved over time ($P < 0.001$), and the improvement correlated with EZ-RPE thickness ($r = -0.45$, $P = 0.021$)...

PMID 33974368

...the mean CDVA (logMAR) improved significantly from 1.35 ± 1.38 to 0.29 ± 0.37 at 6 weeks and remained stable at 12 months postoperatively in monofocal eyes ($P = .001$) and from 1.16 ± 1.2 to 0.37 ± 0.29 (6 weeks) and 0.20 ± 0.36 (12 months) in mIOL eyes ($P = .001$)...

Overcoming the Language Barrier

PMID 26943151

...evaluate best-corrected VA (BCVA) at 1, 2, 3, 6, 9, and 12 months after vitrectomy...best-corrected visual acuity significantly improved over time ($P < 0.001$), and the improvement correlated with EZ-RPE thickness ($r = -0.45$, $P = 0.021$)...



...used standard eyesight charts to measure vision clarity (with glasses or contacts if needed) at 1, 2, 3, 6, 9, and 12 months...vision improved over time, and the improvement corresponded with thickness of a particular layer of the retina called the ellipsoid zone-retinal pigment epithelium (EZ-RPE)...

Source (PMID 34362836)

Background: Little population-based data exist about adults with primary nephrotic syndrome.

Nephrologists reviewed medical records for clinical presentation, laboratory findings, and biopsy results to confirm primary nephrotic syndrome and assigned etiology.

We identified a 1:100 time-matched cohort of adults without diabetes, diagnosed nephrotic syndrome, or proteinuria as controls to compare rates of ESKD, cardiovascular outcomes, and death through 2014, using multivariable Cox regression.

Removal of section headers

Source (PMID 34362836)

Background: Little population-based data exist about adults with primary nephrotic syndrome.

Nephrologists reviewed medical records for clinical presentation, laboratory findings, and biopsy results to confirm primary nephrotic syndrome and assigned etiology.

We identified a 1:100 time-matched cohort of adults without diabetes, diagnosed nephrotic syndrome, or proteinuria as controls to compare rates of ESKD, cardiovascular outcomes, and death through 2014, using multivariable Cox regression.

Replacement of jargon

Generalization

Adaptation by human expert

Nephrotic Syndrome (NS) is a combination of symptoms relating to the kidneys. There is not much information about how NS exists in the population.

Kidney specialists looked at medical records to confirm NS and determined causes.

For comparison, we also found a group of adults with no diabetes, NS, or proteinuria (too much protein in the urine) through 2014. This group was to compare rates of kidney failure, heart disease, and death using statistical methods.

Background for first use of central term

Consistent abbreviation

Splitting of long sentences

Parenthetical explanation on first use

Plain Language Adaptation

Original Two days after starting these feeds, he developed respiratory arrest requiring intubation.

Simplified Two days after starting PediaSure, he stopped breathing and needed a tube inserted in his throat to help him breathe.

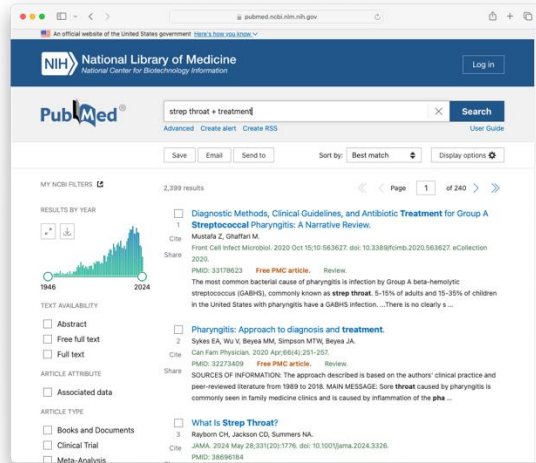
Anaphora: In the original, “these feeds” refers to a mention of PediaSure in the previous sentence.

Part of Speech: a noun phrase has changed to a verb phrase, and “developed” has been dropped.

Verb form: A gerund has been replaced with a past tense verb due to restructuring of sentence clauses.

Contextual pronouns: A single word has been replaced with a longer phrase that refers to the patient.

2. Creating a High-quality Dataset



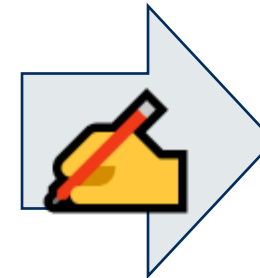
75
MedlinePlus
Questions

X

10
PubMed
results

=

750
abstracts

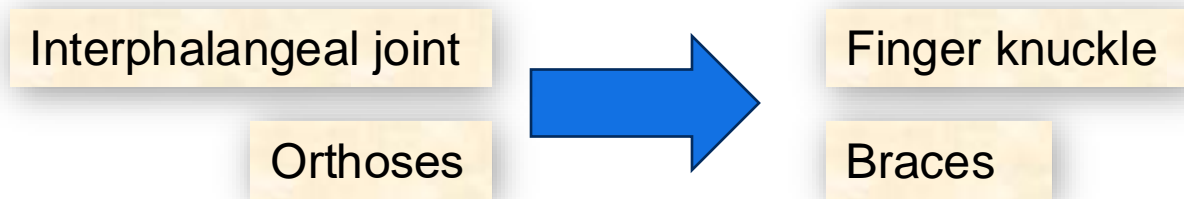


921
adaptations

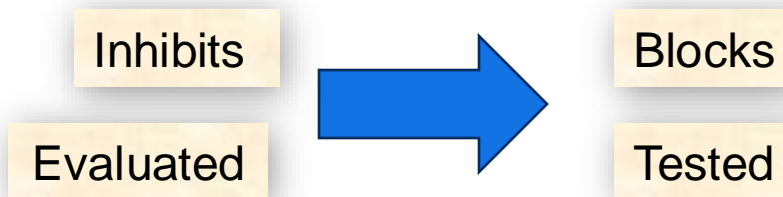
Science writers
&
NLM Index Section

Annotator Guidelines

- Replace professional jargon with common, consumer-friendly terms



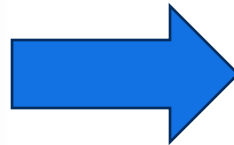
- Substitute longer, more arcane words for shorter, more common synonyms



Annotator Guidelines

- Splitting sentences: if a sentence is long and contains two or more complete thoughts, it should be split into multiple sentences that are simpler.

Treatment guidelines state that there is no difference in efficacy between the various first-generation antipsychotics, **however**, low-potency first-generation antipsychotic drugs are sometimes perceived as less efficacious than high-potency first-generation compounds by clinicians, **and they** also seem to differ in their side effects.

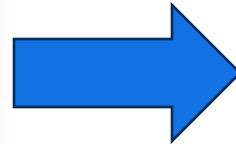


Treatment recommendations say there is no difference in how well older drugs used to treat psychotic disorders work. **However**, doctors sometimes think older drugs with higher doses do not work as well as those with lower doses. **The higher-dose drugs** also seem to have different side effects.

Annotator Guidelines

- If the jargon or a named entity does not have plain synonyms, leave as is in the first mention but explain it with parentheses or nonrestrictive clauses

Duloxetine is a combined **serotonin/norepinephrine** reuptake inhibitor currently under clinical investigation for the treatment of women with stress urinary incontinence.



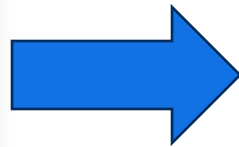
Duloxetine (a common antidepressant) blocks removal of **serotonin/norepinephrine (chemical messengers)** and is studied for treating women with bladder control loss from stress.

- Subsequent mentions in the same abstract do not need to be explained

Annotator Guidelines

- Change passive voice to active voice when possible

Results: eleven trials published between 1965 and 1999 including a total of 2081 participants fulfilled the inclusion criteria and **were further analysed.**

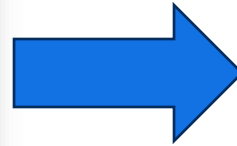


We looked at eleven studies published between 1965 and 1999 with a total of 2081 people.

Annotator Guidelines

- If a subheading is followed by a complete sentence, omit

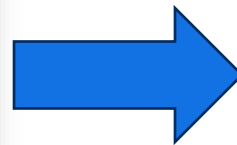
Background: Perphenazine is an old phenothiazine antipsychotic with a potency similar to haloperidol.



Perphenazine is an old drugs used to treat psychotic disorders with strength similar to haloperidol, another drug to treat psychotic disorders.

- If a subheading is followed by an incomplete sentence, incorporate

Objectives: To examine the clinical effects and safety of perphenazine for those with schizophrenia and schizophrenia-like psychoses.

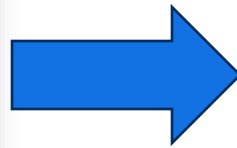


We aimed to look at effects in patients and safety of perphenazine in people with schizophrenia and schizophrenia-like disorders.

Annotator Guidelines

- Omit confidence intervals, p-values, and similar measurements

In children, perphenazine 0.07 mg kg was effective in preventing vomiting (RR, 0.31; 95% CI, 0.18-0.54), whereas in adults, a dose of about 5 mg was effective for the prevention of PONV (RR, 0.50; 95% CI, 0.37-0.67).



Perphenazine prevented nausea and vomiting in children and adults at different doses.

3. Sentence-level Evaluation

Automatic Evaluation

BLEU

Source	The male patient arrived at the hospital, presenting a case of hypertension.
Reference	The male patient arrived at the hospital. He had high blood pressure.
Output	The male patient arrived at the hospital. He presented a case of high blood pressure.

$$P_{keep} = \mathbf{0.63}$$

P_{keep} **Precision** for fraction of n-grams shared between the output and reference

SARI

Source	The male patient arrived at the hospital, presenting a case of hypertension.
Reference	The male patient arrived at the hospital. He had high blood pressure.
Output	The male patient arrived at the hospital. He presented a case of high blood pressure.

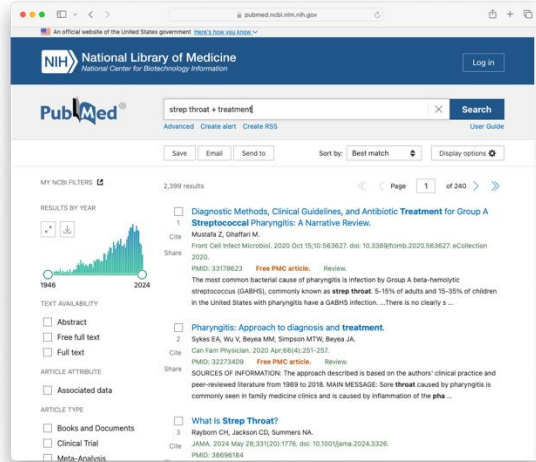
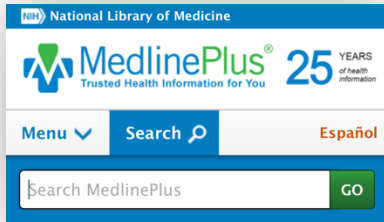
$$\frac{1}{3}F_{add} + \frac{1}{3}F_{keep} + \frac{1}{3}F_{del} = \mathbf{0.83}$$

F_{add} **F-score** for fraction of n-grams shared between the output and reference but not the source

F_{keep} **F-score** for fraction of n-grams shared between the output, reference, and source

F_{del} **F-score** for fraction of n-grams in the input not shared in the output or reference

Test data



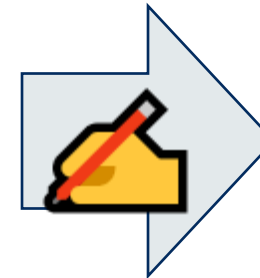
40
MedlinePlus
Questions

X

10
PubMed
results

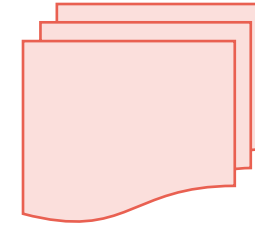
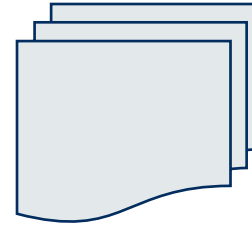
=

400
abstracts



4x

1,600
adaptations



Four-fold references

Median change from baseline was evaluated for procoagulant, anticoagulant, and fibrinolytic parameters, and for sex hormone-binding globulin (SHBG).

Results: Median change of endogenous thrombin potential (ETP) based activated protein C sensitivity resistance (APCr) at cycle 6 was +30% for E4/DRSP, +165% for EE/LNG (p-value <0.05 vs E4/DRSP), and +219% for EE/DRSP (p-value <0.05 vs E4/DRSP).

Changes to prothrombin fragment 1 + 2 and SHBG for E4/DRSP, EE/LNG, and EE/DRSP were +23%, +71%, and +64% (p-value <0.05 vs E4/DRSP); and +55%, +74% and +251% (p-value <0.05 vs E4/DRSP), respectively.

We measured the tendency of blood to clot with a lab test. The average increase in clotting tendency compared to the start of the study was 30% for E4/DRSP, 165% for EE/LNG, and 219% for EE/DRSP.

In the last sixth cycle the blood clotting chemicals changed significantly with the largest change in the EE/DRSP group, closely followed by the EE/LNG group.

For the study's results, the average change for a coagulation or bleeding-prevention test at cycle 6 was +30% for E4/DRSP, +165% for EE/LNG, which was much higher than E4/DRSP, and +219% for EE/DRSP, which was much higher than vs E4/DRSP.

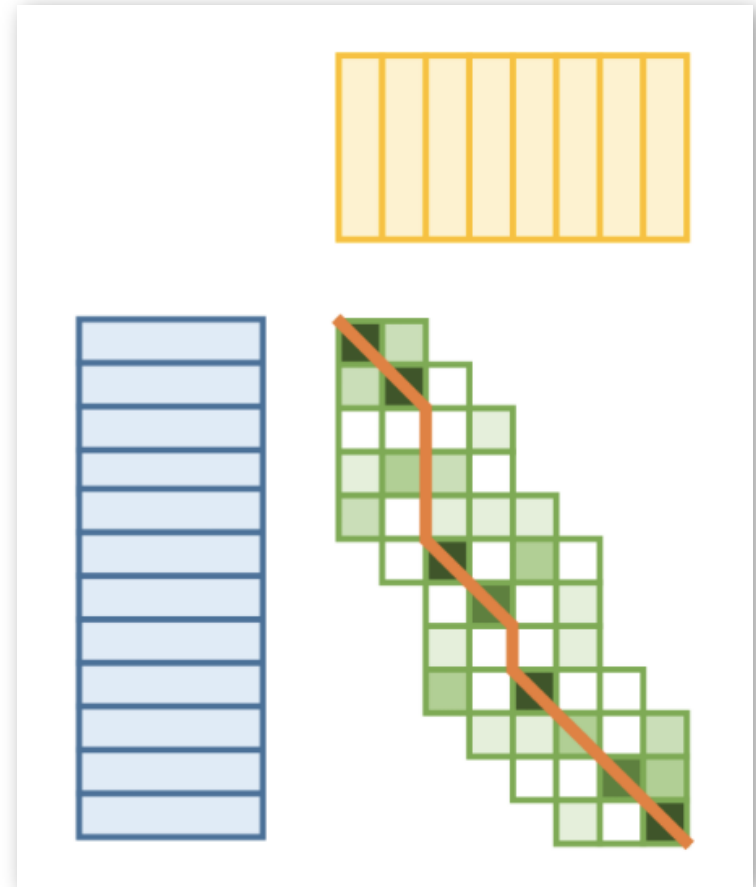
Change of internal thrombin potential (ETP) based activated protein C sensitivity resistance (APCr) at cycle 6 was increased for all groups. APCr is an increased tendency of the blood to clot.

Manual Evaluation

Simplicity		
SEN	Sentence simplicity	Are long, complex sentences appropriately split?
TRM	Term simplicity	Are expert terms in the source replaced with alternatives, or explained, either in this sentence or a previous sentence?
TAC	Term accuracy	Are substitutions and explanations of expert terms accurate?
FLU	Fluency	Does the output follow grammatical rules and read smoothly?
SIM	Final simplicity	Average(SEN,TRM,TAC,FLU)
Accuracy		
COM	Completeness	How much of the source information does the output provide?
FTH	Faithfulness	Do points made in the output match those of the source?
ACC	Final accuracy	Average(COM,FTH)
Final		
FIN	Final score	Average(SIM,ACC)

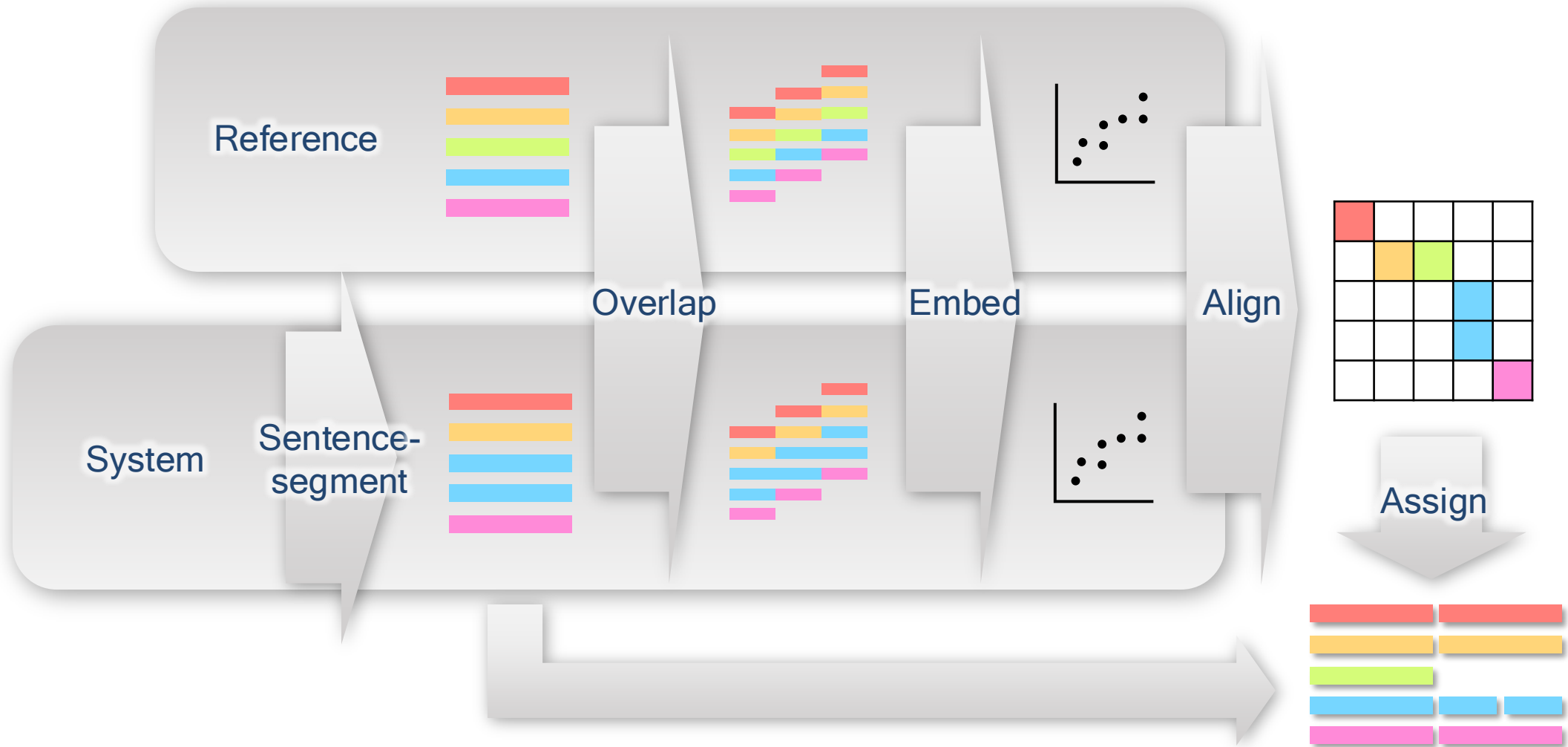
Handling document-level submissions

- Sentence segmentation with Stanza (Qi 2020)
- Overlaps
 - Windows of up to 10 sentence, to allow one-to-many alignment
- LASER (Artetxe 2019) sentence embeddings
- VecAlign (Thompson 2019)
 - `--alignment_max_size 10`
 - `--one_to_many`

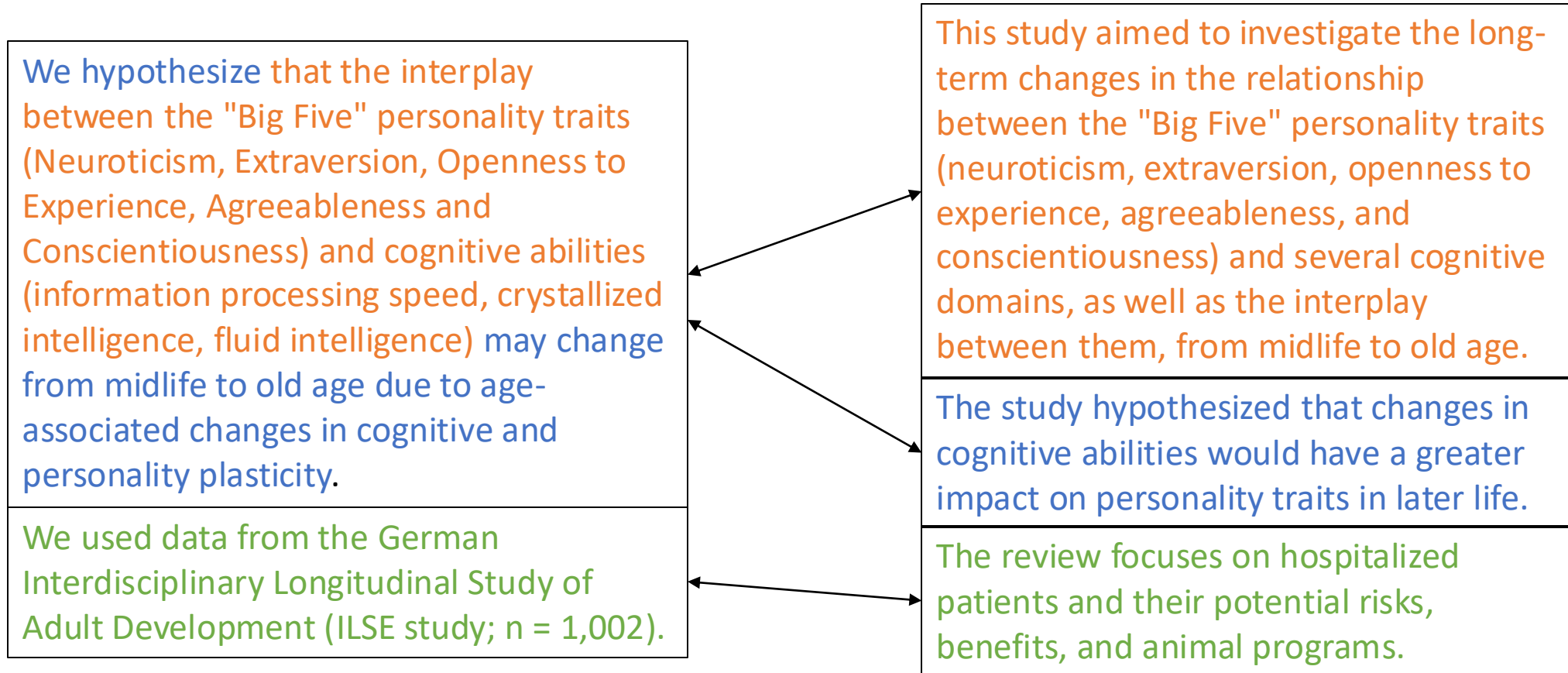


Thompson 2019

Handling document-level submissions



Handling document-level submissions

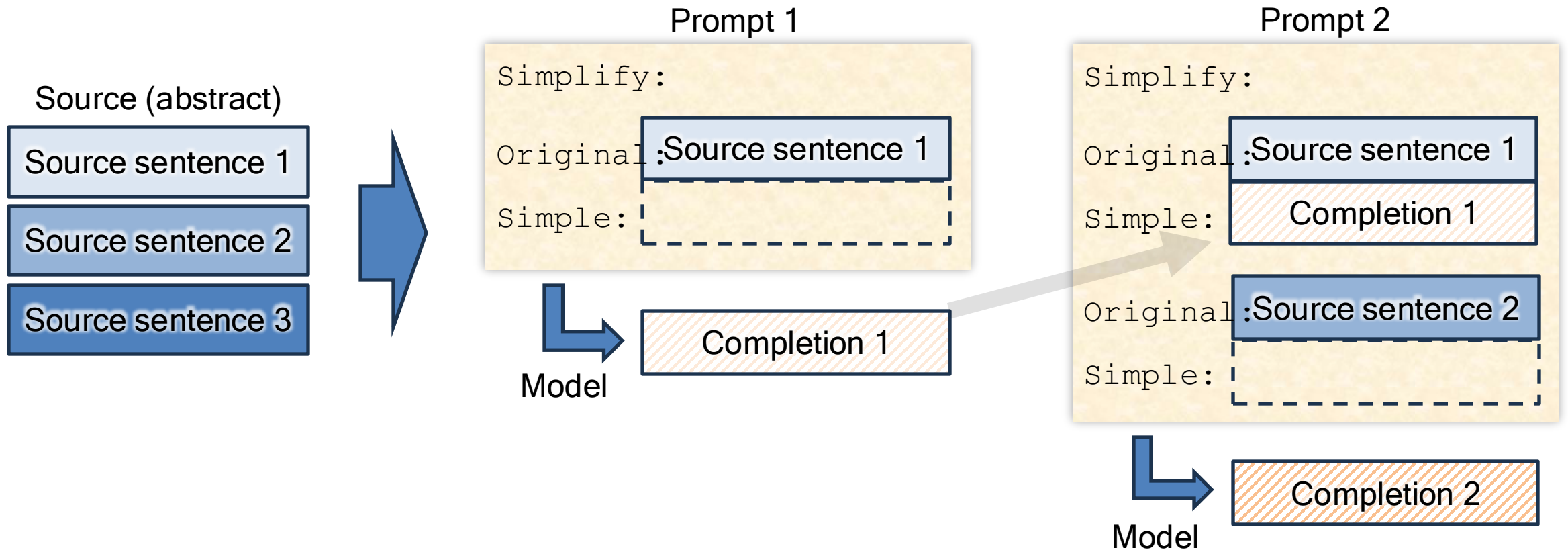


Source

Target

4. Baselines and Team Systems

PLABA_base_1: GPT-3.5, zero-shot, autoregressive



PLABA_base_2

- GPT 3.5
- Zero-shot
- Document-level
 - Prompted to rewrite each sentence

PLABA_base_3

- Falcon 7B instruct
- Zero-shot
- Document-level
 - Simpler prompt

Teams



Team	Location	Institute
BoschAI	Germany	Bosch Center for Artificial Intelligence
Mason_NLP	Virginia, USA	George Mason University
Bee_Manc	United Kingdom	University of Manchester
PTM3	West Bank/Spain	Palestine Tech. Univ. - Kadoorie; Univ. Carlos III de Madrid

Submissions

Team	Sub.	Description
valeknappich	1	“LLama 13B trained with token-based weight losses derived from diffs between source and target. For every sentence, a simplification is generated with varying temperatures. The best simplification is then selected using ChatGPT.”
	2	“LLama 13B trained with token-based weight losses derived from diffs between source and target.”
	3	“Zero-shot prompting of ChatGPT in a sentence-by-sentence fashion.”
Mason_NLP	1	“One shot, instruction tuning on GPT-4. Instructions were custom-designed for this task. This had a top preference in our human evaluation.”
	2	“Predictions from a GPT3.5 instruction-tuned model on the training set and had a second preference in our human evaluation.”
	3	“Predictions from the model LLaMa13B-chat fine-tuned on the PLABA training set. This model had 3rd preference in our human evaluation.”
Bee_Man	1	“Bart-large with control tokens”
	2	“T5”
	3	“sci-T5”
PTM3	1	“Zero-shot adaptation for biomedical abstracts with biobart model version 2”

5. Manual and Automatic Results

Results Summary

Submission	Auto (SARI)	Manual (0-100)	Strategies
PLABA_base_1	35.97	92.82	GPT 3.5 zero-shot w/ autoregressive prompting
PLABA_base_2*	36.96	89.54	GPT 3.5 zero-shot single prompt
PLABA_base_3*	23.48	44.69	Falcon7B zero-shot single prompt
MasonNLP_1	39.51	91.90	“GPT4 one-shot w/ in-context learning”
BoschAI_1	44.65	91.87	“LLama 13B FT w/ token-weighted loss, ChatGPT ranking”
Bee_Manc_1	40.59	78.87	“Bart-large with control tokens”
PT3M_1*	34.55	20.40	“Biobart version 2, zero-shot”

*Automatically aligned document-level system

Manual evaluation

Submission	Simplicity					Accuracy			Avg.
	Sent.	Term.	Term acc.	Fluency	Avg.	Comp.	Faith.	Avg.	
PLABA_base_1	91.45	86.84	91.22	93.53	90.76	95.73	94.02	94.87	92.82
PLABA_base_2*	94.33	81.94	87.50	95.25	89.76	90.17	88.46	89.32	89.54
PLABA_base_3*	84.67	63.67	42.67	87.00	69.50	20.94	18.80	19.87	44.69
MasonNLP_1	91.63	91.74	88.26	93.49	91.28	94.44	90.60	92.52	91.90
BoschAI_1	91.11	77.25	94.11	92.96	88.86	95.30	94.44	94.87	91.87
Bee_Manc_1	92.84	82.33	64.20	91.57	82.74	79.49	70.51	75.00	78.87
PT3M_1*	38.19	34.38	21.99	18.17	28.18	16.24	8.97	12.61	20.40

*Automatically aligned document-level system

PLABA_base_1 (GPT-3.5 zero-shot autoregressive)

Source (PMID: 32956694)	PLABA_base_1
Objective: To assess the effect on hemostasis parameters of a new combined oral contraceptive (COC).	Goal: To determine how a new birth control pill affects blood clotting.
Study design: In this randomized, single centre, open-label, exploratory study, healthy women received either 15 mg estetrol/3 mg drospirenone (E4/DRSP) (n = 39), 30 mcg ethinylestradiol/150 mcg levonorgestrel (EE/LNG) (n = 30), or 20 mcg ethinylestradiol/3 mg drospirenone (EE/DRSP) (n = 32) for six 28-day cycles.	Study Design: In this study, healthy women were randomly assigned to receive either a new birth control pill containing 15 mg estetrol and 3 mg drospirenone, a commonly used birth control pill containing 30 mcg ethinylestradiol and 150 mcg levonorgestrel, or another commonly used birth control pill containing 20 mcg ethinylestradiol and 3 mg drospirenone. They took the pills for six cycles, each lasting 28 days.
Blood was collected at baseline, cycle 3, and cycle 6.	Blood samples were collected at the beginning of the study and at the third and sixth cycles.
Median change from baseline was evaluated for procoagulant, anticoagulant, and fibrinolytic parameters, and for sex hormone-binding globulin (SHBG).	We looked at the average changes in blood clotting, blood thinning, and blood clot breakdown, as well as the levels of sex hormone-binding globulin, compared to the beginning of the study.

...

PLABA_base_1 (GPT-3.5 zero-shot autoregressive)

...

<p>Results: Median change of endogenous thrombin potential (ETP) based activated protein C sensitivity resistance (APCr) at cycle 6 was +30% for E4/DRSP, +165% for EE/LNG (p-value <0.05 vs E4/DRSP), and +219% for EE/DRSP (p-value <0.05 vs E4/DRSP).</p>	<p>Results: At the end of the study, the average increase in blood clotting was 30% for women taking the new birth control pill, 165% for women taking the commonly used birth control pill, and 219% for women taking another commonly used birth control pill. These increases were statistically significant compared to the new birth control pill.</p>
<p>Changes to prothrombin fragment 1 + 2 and SHBG for E4/DRSP, EE/LNG, and EE/DRSP were +23%, +71%, and +64% (p-value <0.05 vs E4/DRSP); and +55%, +74% and +251% (p-value <0.05 vs E4/DRSP), respectively.</p>	<p>Changes in blood clotting markers and sex hormone-binding globulin levels were observed in women taking the different birth control pills. These changes varied depending on the type of pill, with some pills causing larger increases in blood clotting and sex hormone-binding globulin compared to others.</p>
<p>At cycle 6, changes to other hemostasis parameters for E4/DRSP were similar or smaller than for EE/LNG or EE/DRSP.</p>	<p>By the sixth cycle of the study, the changes in other blood clotting parameters were similar or smaller in women taking the new birth control pill compared to the other commonly used birth control pills.</p>

...

PLABA_base_1 (GPT-3.5 zero-shot autoregressive)

...

Conclusions: In this study, changes in hemostasis parameters after treatment with 6 cycles of E4/DRSP were smaller or similar to those observed for EE/LNG .	Conclusions: This study found that the new birth control pill had similar or smaller effects on blood clotting compared to the commonly used birth control pill .
Similar, but more pronounced changes were also observed versus EE/DRSP , which supports the hypothesis that the effect of COCs on hemostasis parameters is mainly mediated by the estrogenic component.	Similar, but stronger changes were also seen compared to the other commonly used birth control pill , which suggests that the estrogen in birth control pills is the main factor affecting blood clotting .
Further studies are needed to provide more insight into the venous thromboembolic risk of E4/DRSP .	More studies are needed to understand the risk of blood clots with the new birth control pill .
Implications statement: This study reports that the effects on hemostasis parameters of a COC containing 15 mg E4/3 mg DRSP are less or similar to those for EE/LNG or EE/DRSP .	Implications: This study found that the new birth control pill has similar or smaller effects on blood clotting compared to other commonly used birth control pills .
It also demonstrates that the choice of estrogen modulates the effects of COCs on hemostasis parameters .	It also shows that the type of estrogen in birth control pills affects how they affect blood clotting .

PLABA_base_1 hallucination and inaccuracy

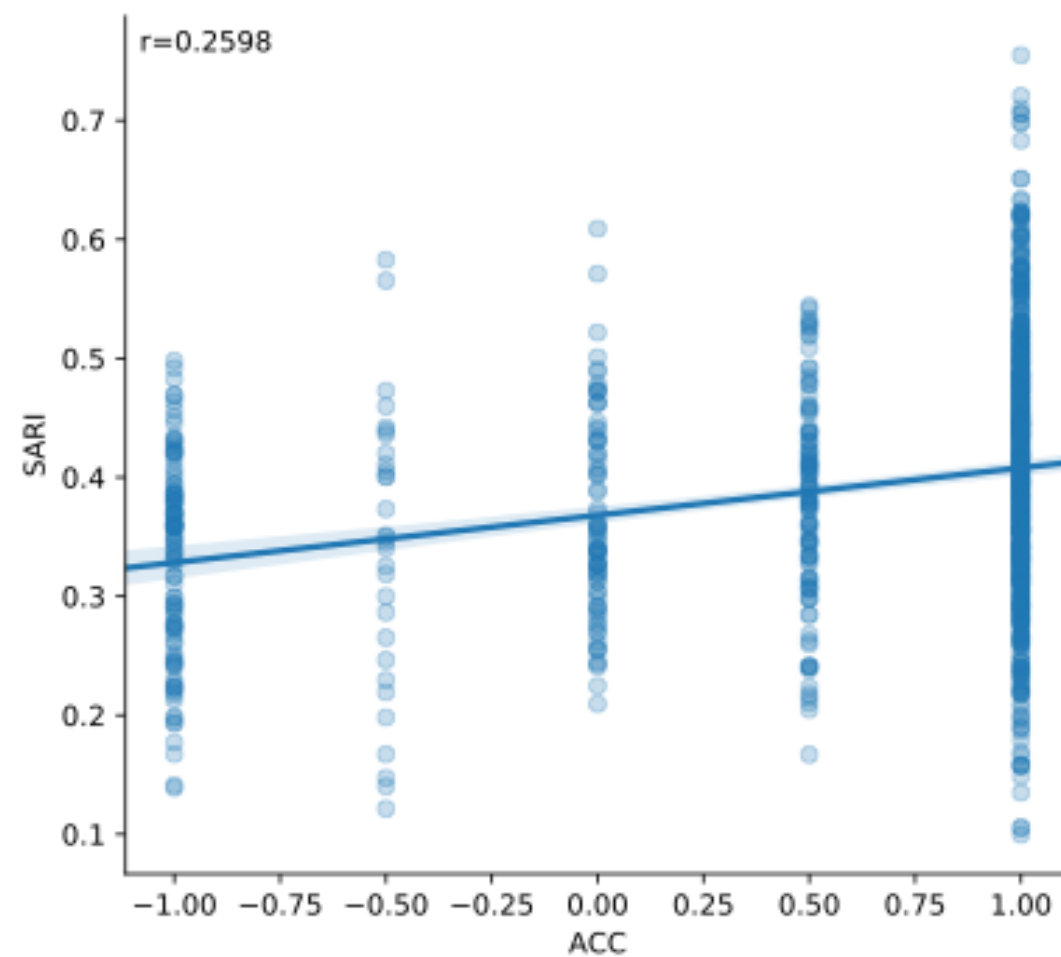
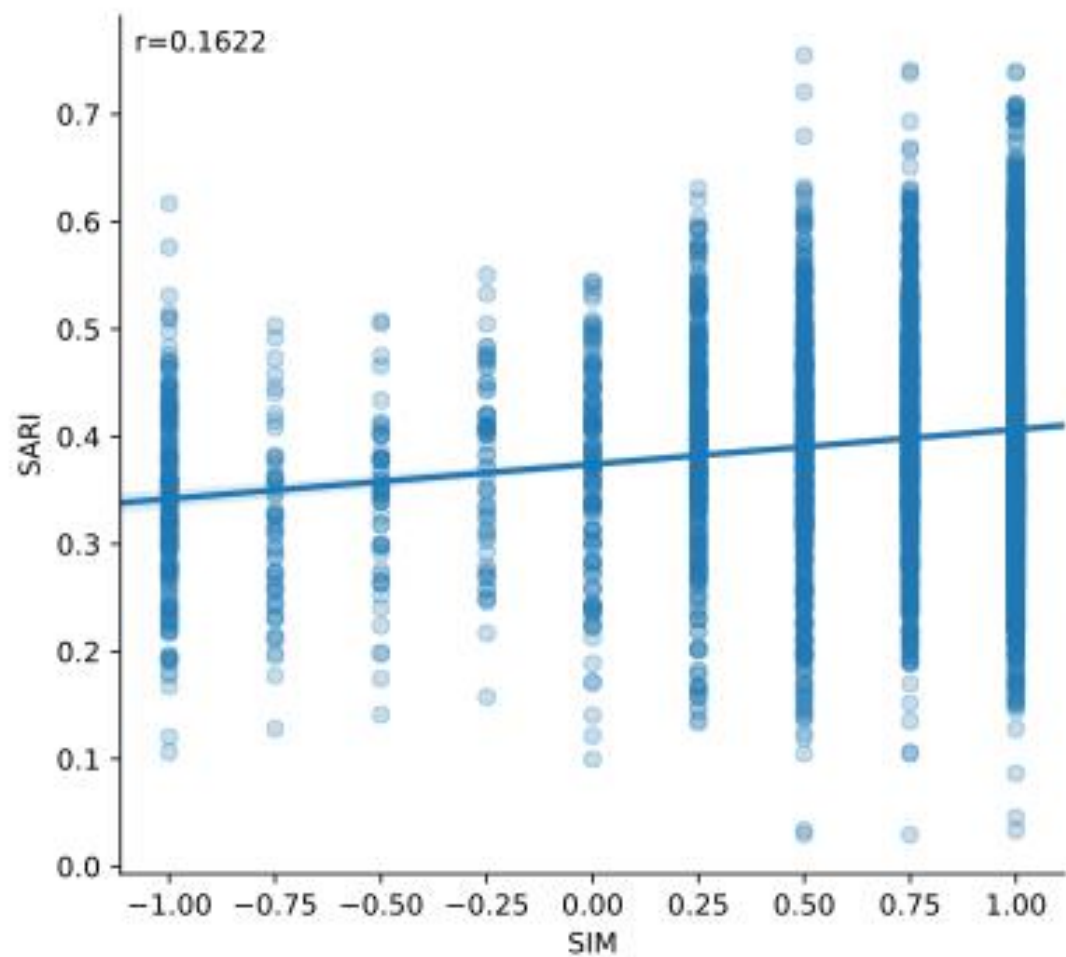
Source (PMID: 25864273)	PLABA_base_1
<p>Materials and methods: The study involved 308 postpartum women who had a medical examination one month after delivery.</p>	<p>For this study, researchers looked at 308 women who had given birth and had a medical check-up one month later. <i>Original: Results: Neck and shoulder pain was experienced by 61.4% of the women, with the highest severity in the neck area. The pain had a negative impact on the women's quality of life. Factors such as age, history of NSP, and work-related factors were found to be associated with NSP. Simple: The study found [...]</i></p>
Source (PMID: 25864273)	PLABA_base_1
<p>The disturbance of daily life due to NSP in postpartum women with past history of PMS and Hiesho were significant higher than that for women without those.</p>	<p>Postpartum women who had a history of premenstrual syndrome (PMS) and anemia during pregnancy experienced more disruption in their daily lives due to neck and shoulder pain compared to women without these conditions.</p>

PLABA_base_3 (Falcon 7B zero-shot document-level)

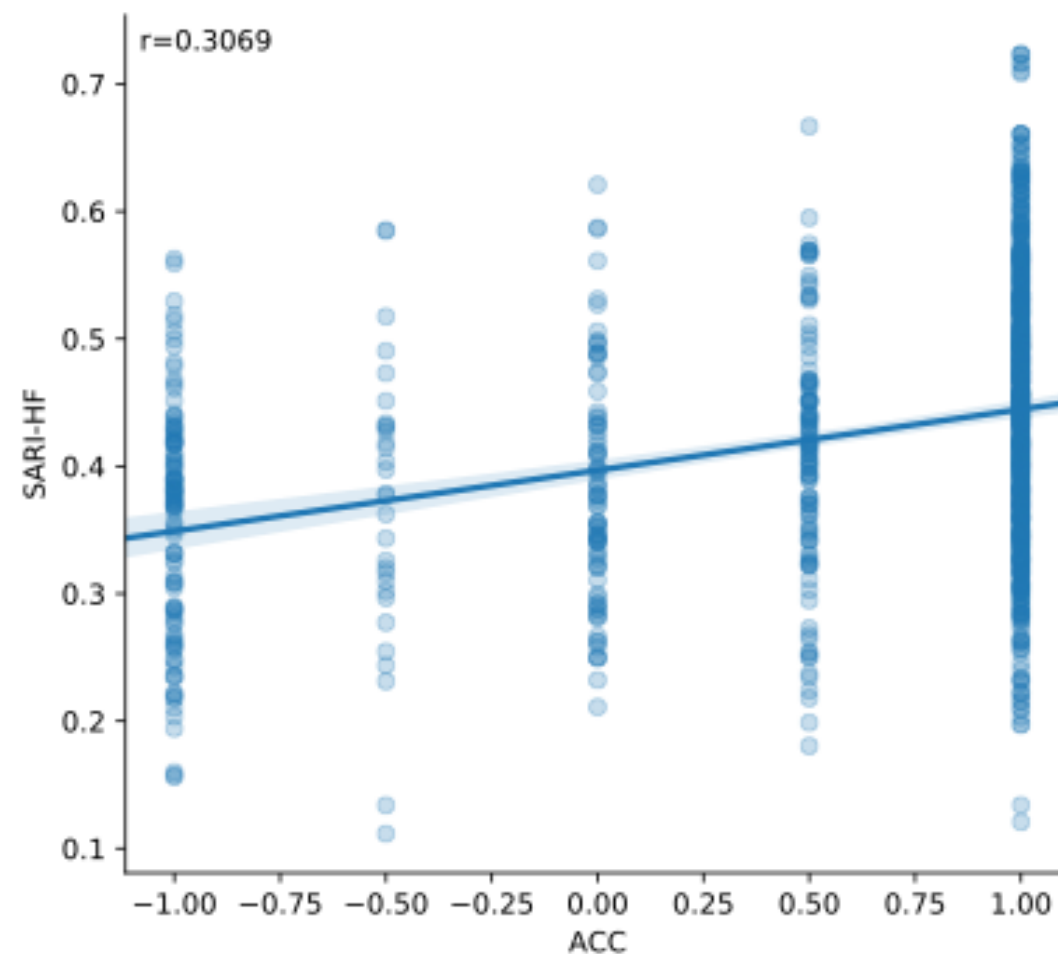
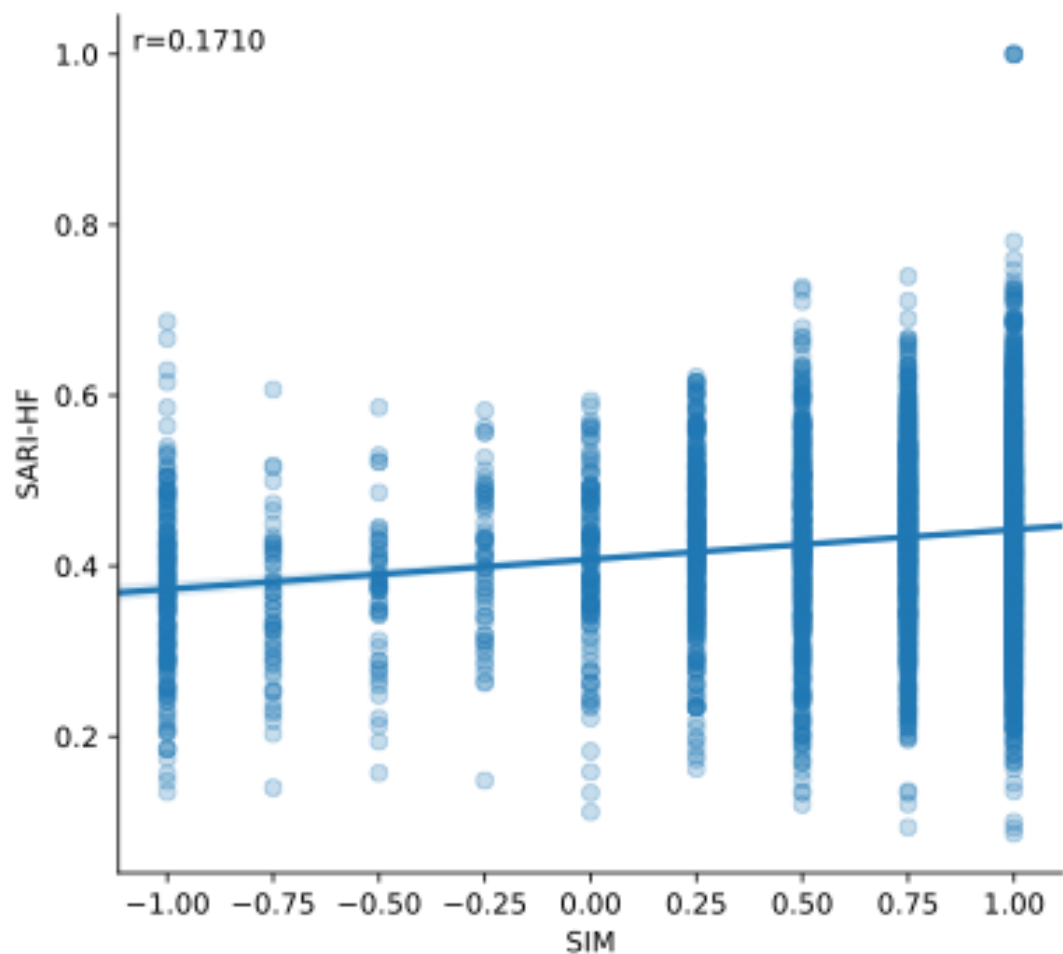
Source (PMID: 31768813)	Falcon 7B Instruct (zero-shot)
<p>Purpose: To determine predictors of best-corrected postoperative visual acuity (VA) in patients who underwent surgical intervention for macula-off rhegmatogenous retinal detachment.</p>	<p>Purpose: We report an analysis of the visual acuity in a cohort of 100 consecutive eyes that underwent surgical intervention for macula-off rhegmatogenous retinal detachment in a 4-year period.</p>
<p>Materials and methods: Primary macula-off retinal detachments from the University of Colorado Primary Rhegmatogenous Retinal Detachment Database (2012-2017) were reviewed.</p>	<p>A total of 2-year visual outcomes and postoperative predictors for visual recovery were analyzed to identify the factors that may predict visual outcome.</p>
...	
<p>Chi-square or Fisher's exact test was used to compare categorical variables, and Wilcoxon rank sum test was used for continuous variables.</p>	<p>Results: A total of 100 eyes underwent surgical intervention for macula-off rhegmatogenous retinal detachment.</p>
...	

6. Assessing Automatic Metrics

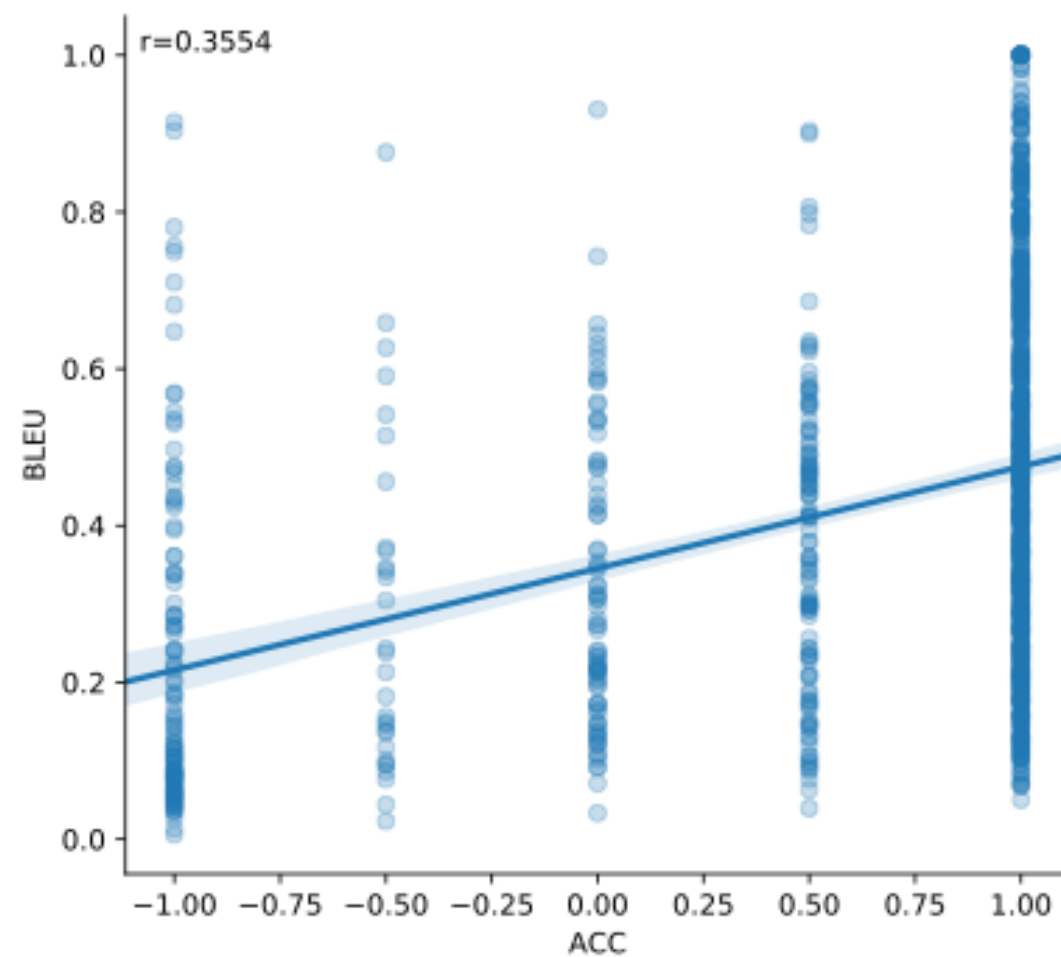
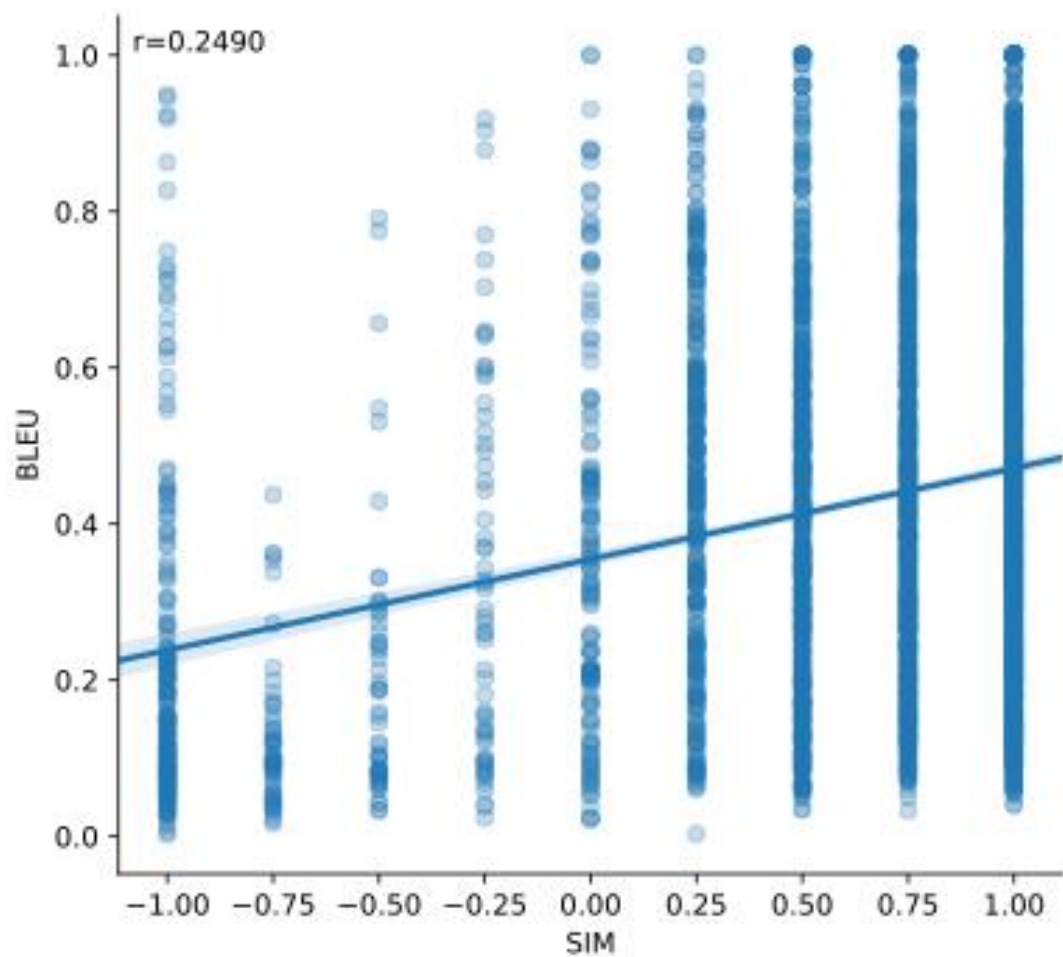
Manual evaluations vs. SARI



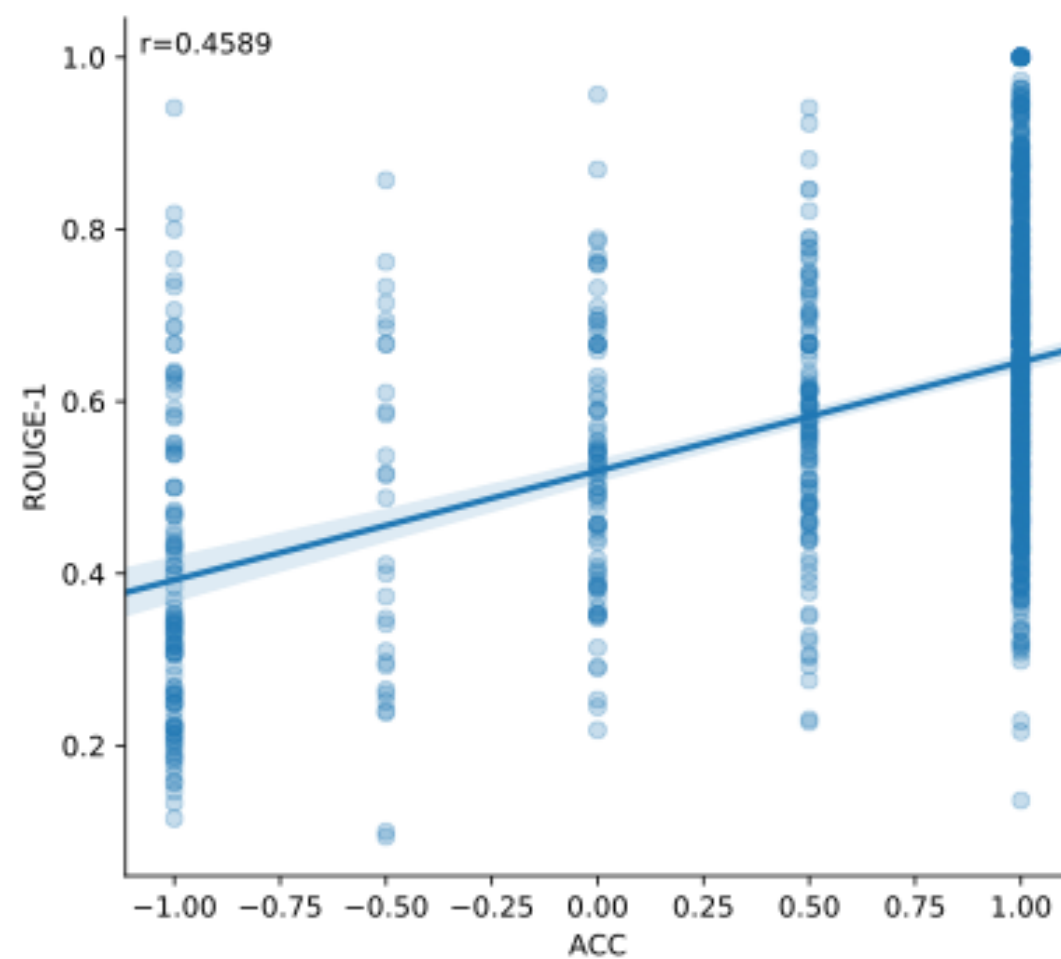
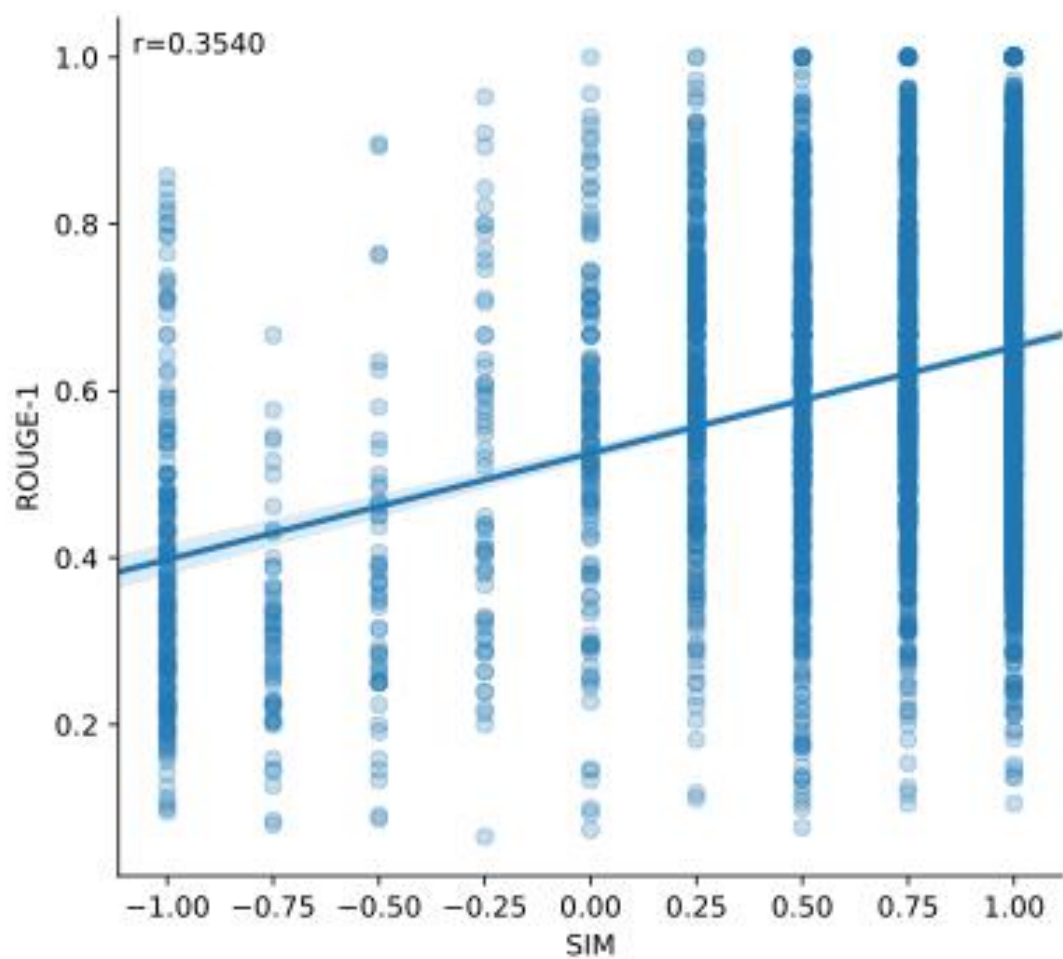
Manual evaluations vs. SARI (huggingface)



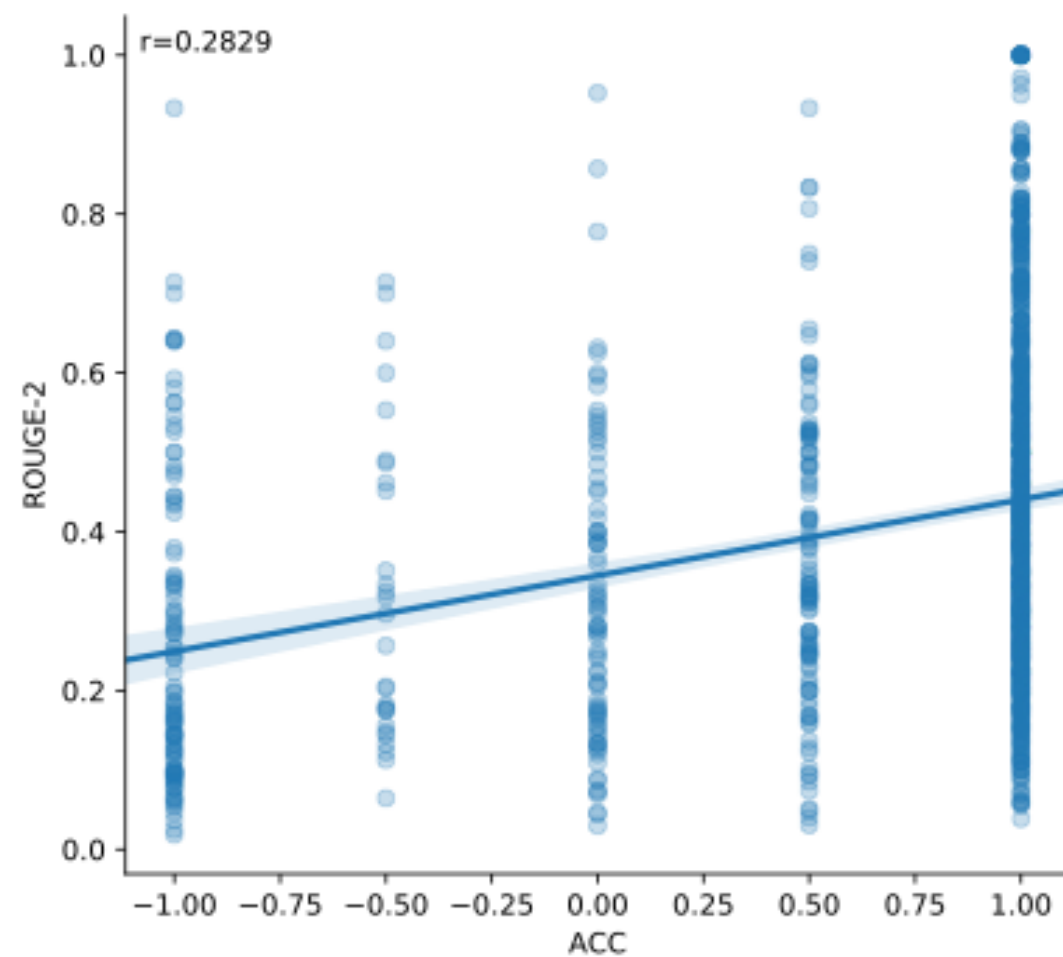
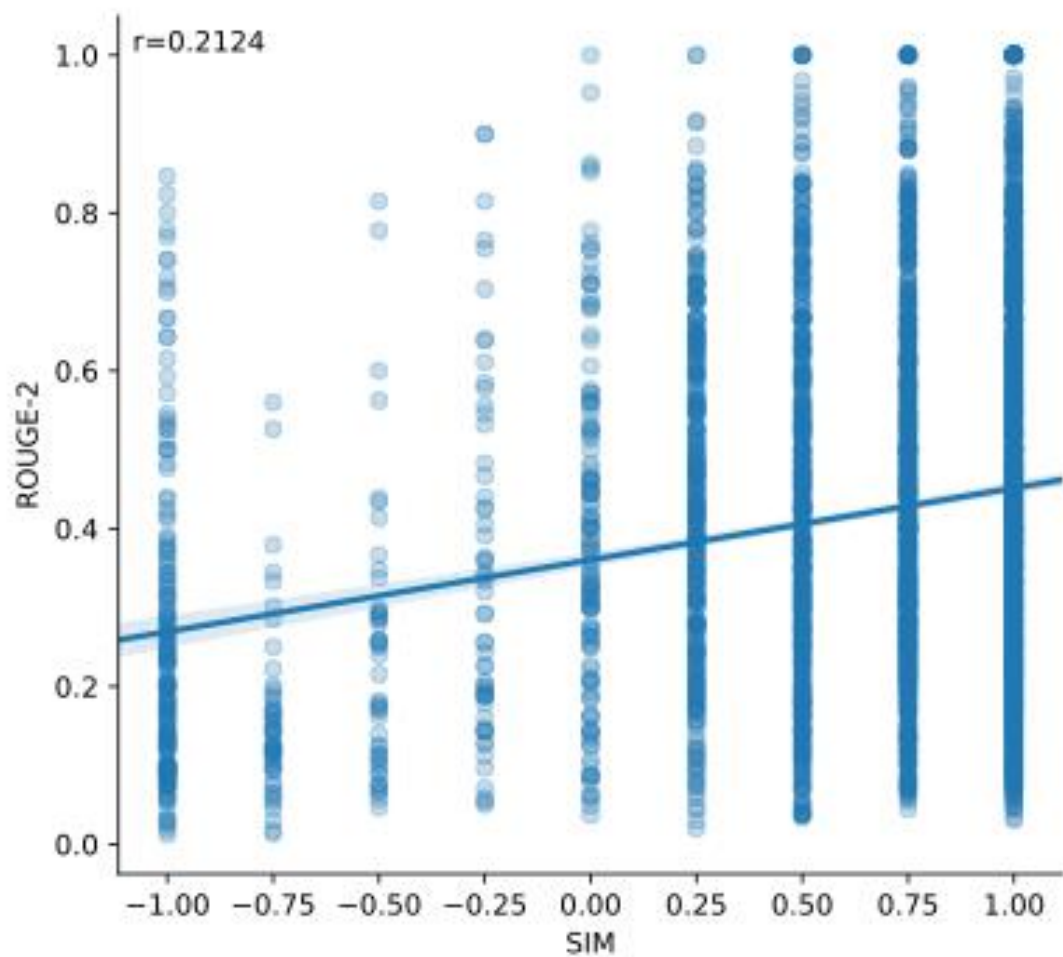
Manual evaluations vs. BLEU



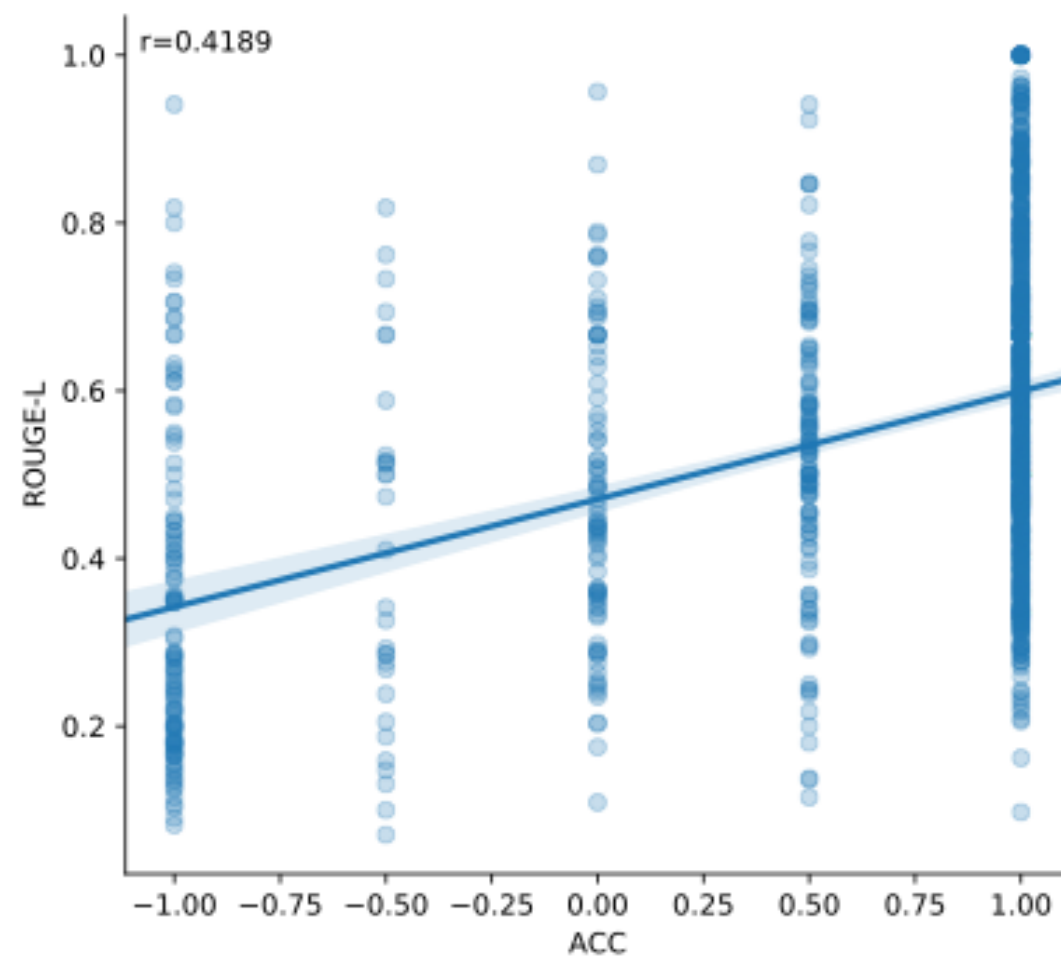
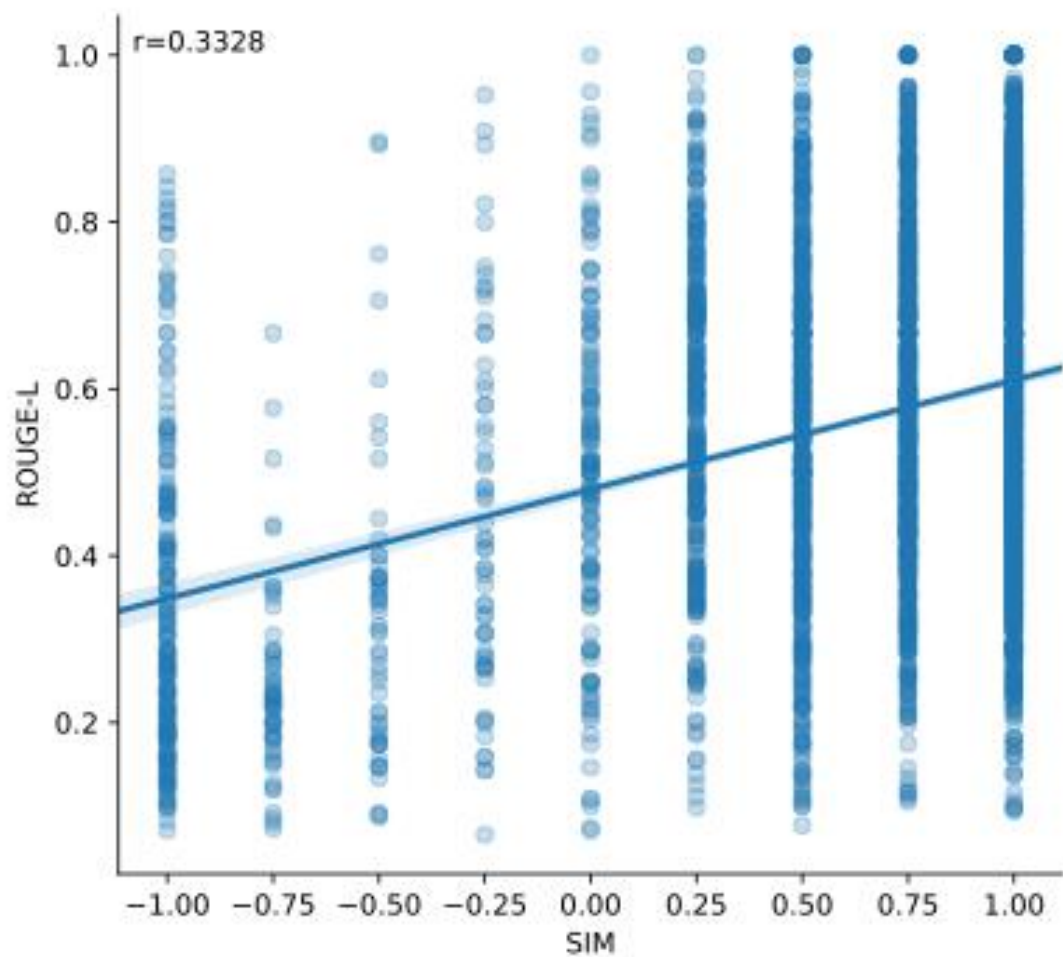
Manual evaluations vs. ROUGE-1



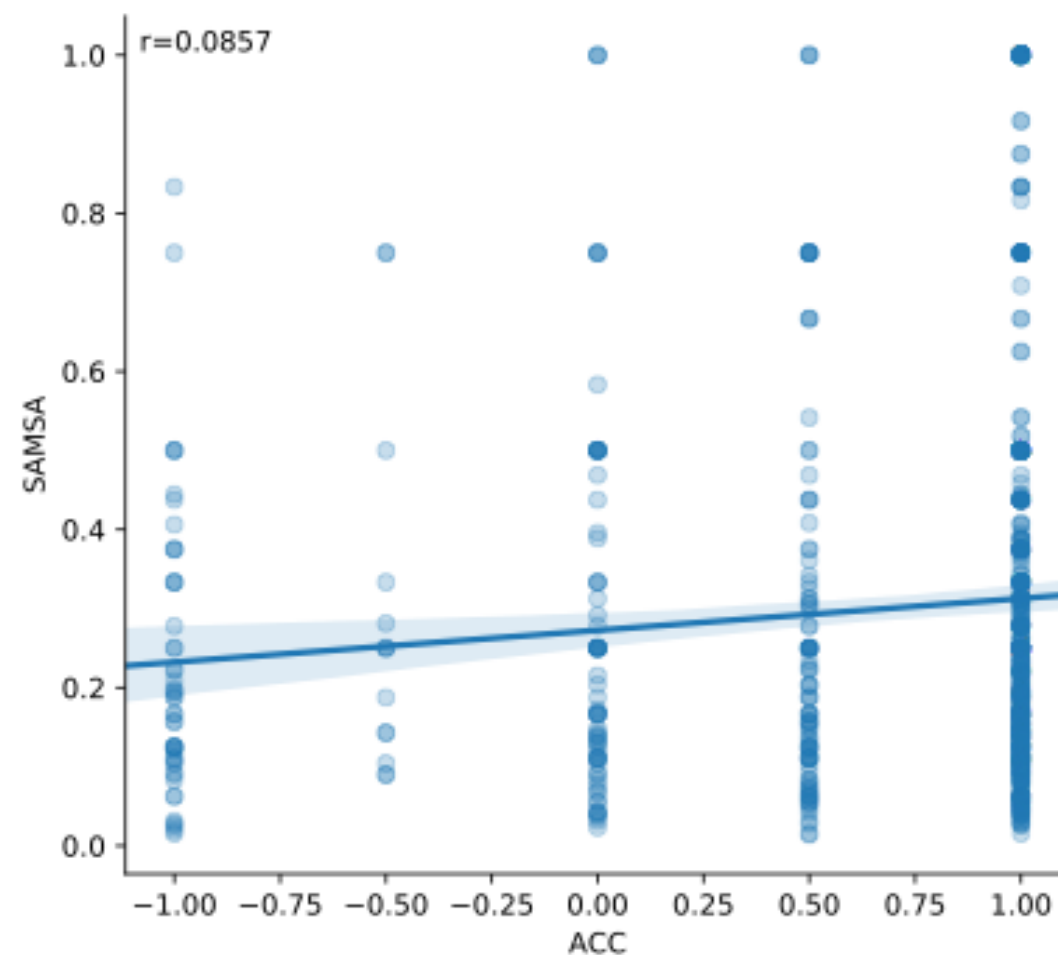
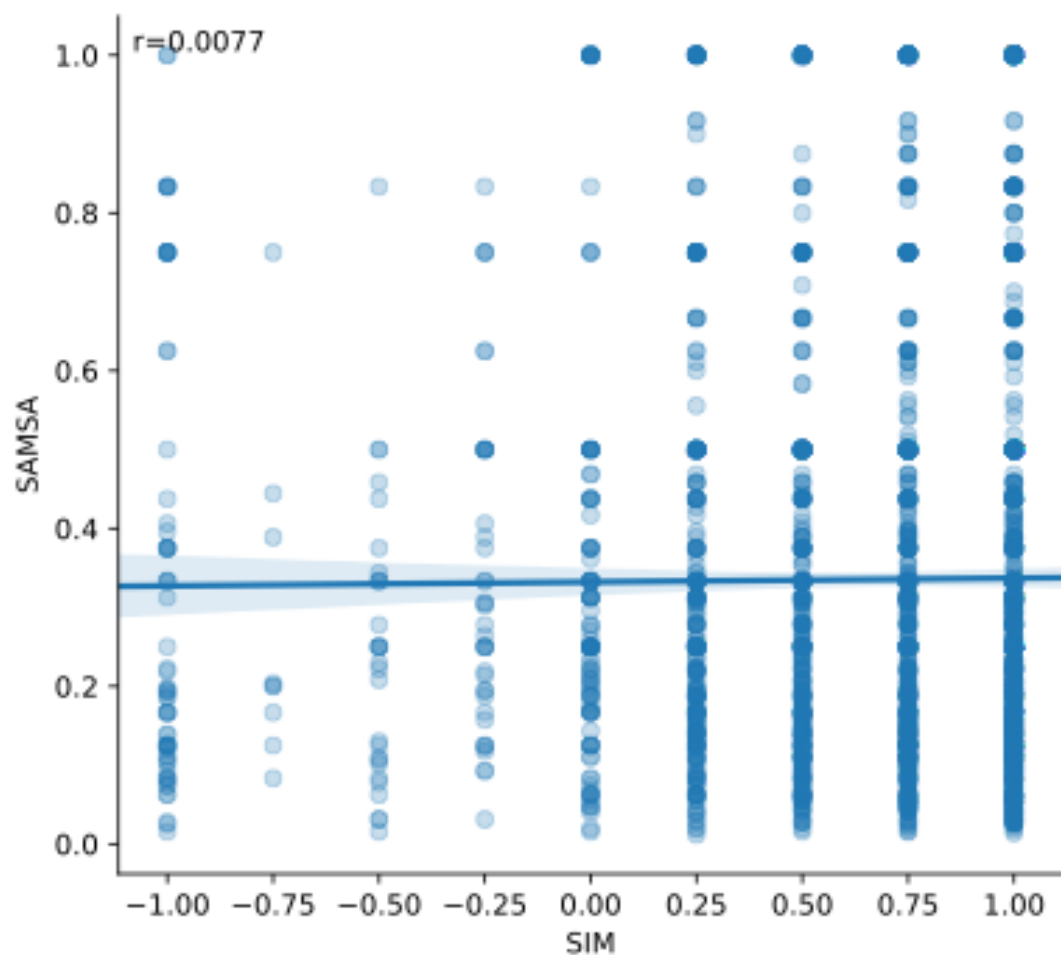
Manual evaluations vs. ROUGE-2



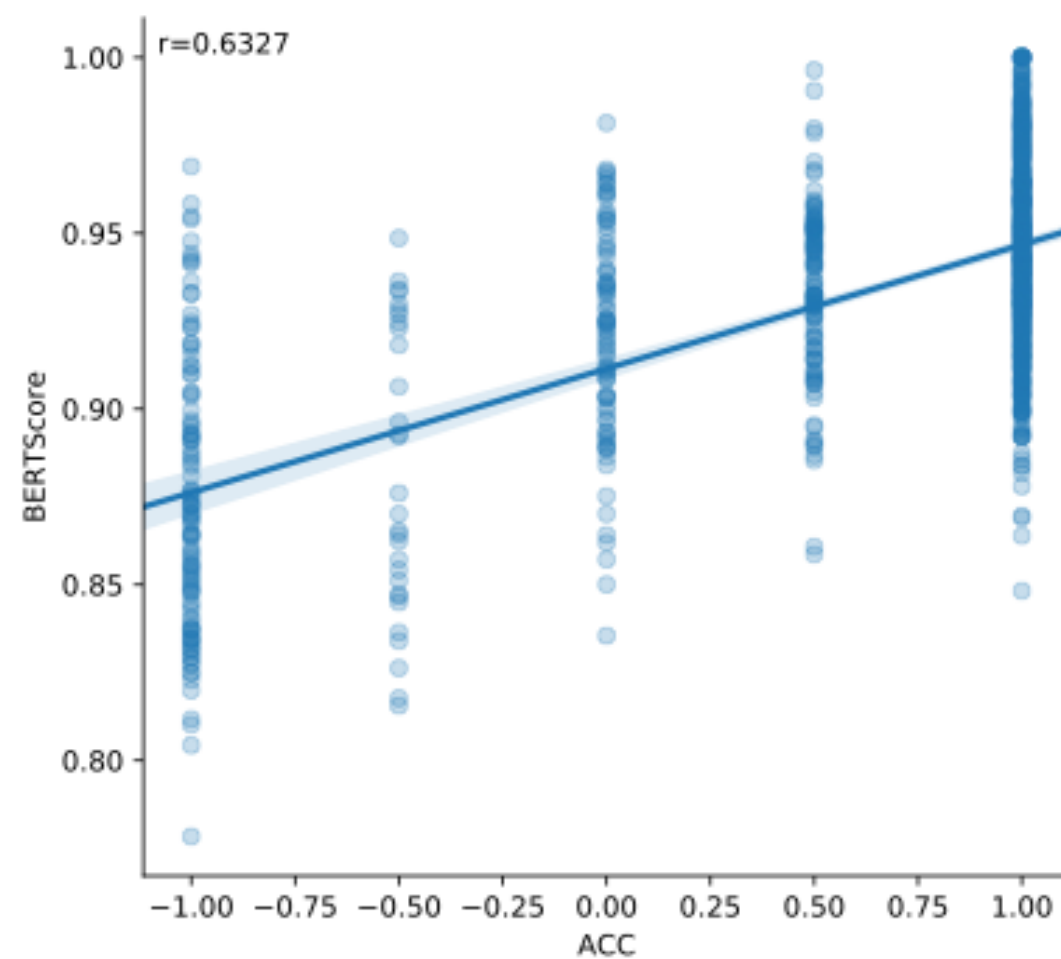
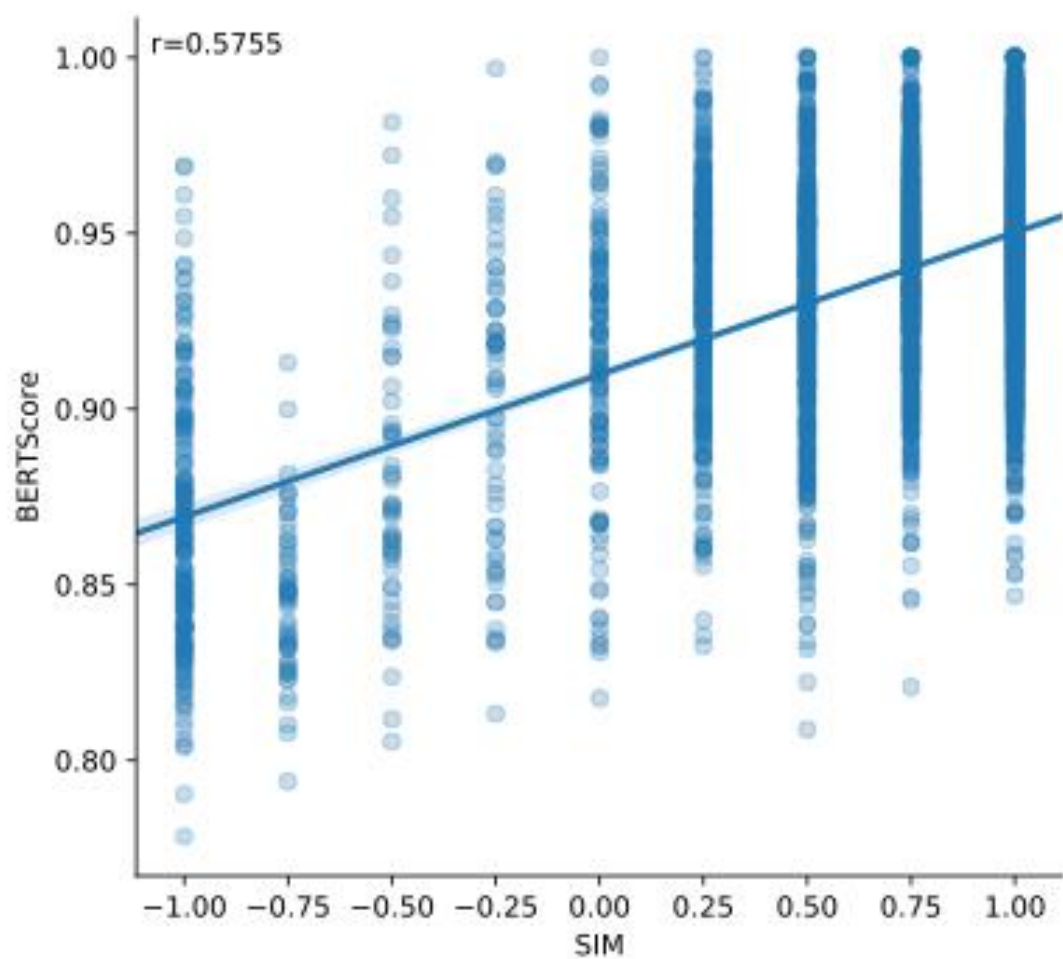
Manual evaluations vs. ROUGE-L



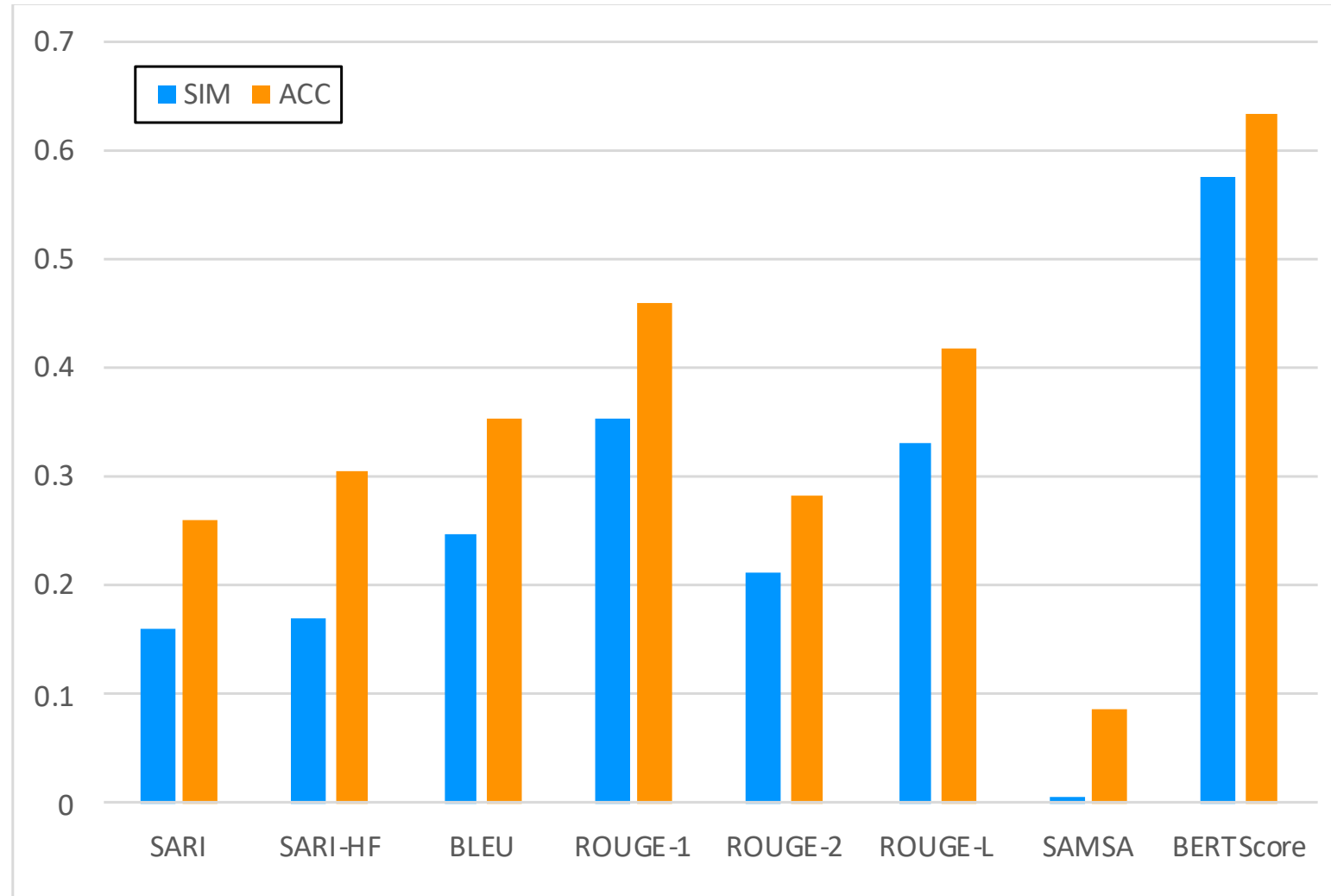
Manual evaluations vs. SAMSA



Manual evaluations vs. BERTScore



Summary of manual vs. automatic correlations



7. New Subtask: Term Replacement

PLABA 2024 @ TREC

New subtask: expert term replacement

[Sentence 9]

expert term

However, Arruga's ring suture method induced cataract more frequently than did other surgical procedures.

← term ——— term

EXPLAINS

← term ——— term

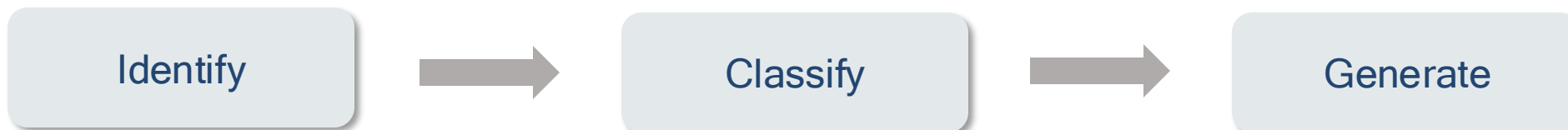
However, a particular procedure called "Arruga's ring suture method" caused cataract more often than other surgical procedures.

← term ———

GENERALIZES

← term ———

However, a specific suture method caused cataract more frequently than did other surgeries.
However, Arruga's ring suture method caused more cataracts than other surgeries



{substitute, explain, generalize, exemplify, omit}

Acknowledgements



Kush Attal



Dina Demner-
Fushman

This research was supported by the Intramural Research Program of the National Library of Medicine (NLM), National Institutes of Health.



National Library
of Medicine

Attal, K., Ondov, B. and Demner-Fushman, D., 2023. A dataset for plain language adaptation of biomedical abstracts. *Scientific Data*, 10(1), p.8.

Berkman, N.D., Sheridan, S.L., Donahue, K.E., Halpern, D.J. and Crotty, K., 2011. Low health literacy and health outcomes: an updated systematic review. *Annals of internal medicine*, 155(2), pp.97-107.

Artetxe, M. and Schwenk, H., 2019. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Transactions of the Association for Computational Linguistics*, 7, pp.597-610.

Ondov, B., Attal, K. and Demner-Fushman, D., 2022. A survey of automated methods for biomedical text simplification. *Journal of the American Medical Informatics Association*, 29(11), pp.1976-1988.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J. and Manning, C.D., 2020, July. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 101-108).

Thompson, B. and Koehn, P., 2019, November. Vecalign: Improved sentence alignment in linear time and space. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp.1342-1348.

White, R.W. and Horvitz, E., 2009. Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems (TOIS)*, 27(4), pp.1-37.