

The Impact of Medical Interpretation Method on Time and Errors

Francesca Gany, MD, MS⁵, Luciano Kapelusznik, MD⁵, Kavitha Prakash, MD, MPH^{2,3},
Javier Gonzalez⁵, Lurmag Y. Orta, MD⁴, Chi-Hong Tseng, PhD¹,
Jyotsna Changrani, MD, MPH⁵

¹New York University School of Medicine, New York, USA; ²Harvard Medical School, Boston, USA; ³Cambridge Health Alliance, Windsor Street Health Center, Cambridge, USA; ⁴Mount Sinai School of Medicine, New York, USA; ⁵Center for Immigrant Health, New York University School of Medicine, New York, NY, USA.

BACKGROUND: Twenty-two million Americans have limited English proficiency. Interpreting for limited English proficient patients is intended to enhance communication and delivery of quality medical care.

OBJECTIVE: Little is known about the impact of various interpreting methods on interpreting speed and errors. This investigation addresses this important gap.

DESIGN: Four scripted clinical encounters were used to enable the comparison of equivalent clinical content. These scripts were run across four interpreting methods, including remote simultaneous, remote consecutive, proximate consecutive, and proximate ad hoc interpreting. The first 3 methods utilized professional, trained interpreters, whereas the ad hoc method utilized untrained staff.

MEASUREMENTS: Audiotaped transcripts of the encounters were coded, using a prespecified algorithm to determine medical error and linguistic error, by coders blinded to the interpreting method. Encounters were also timed.

RESULTS: Remote simultaneous medical interpreting (RSMI) encounters averaged 12.72 vs 18.24 minutes for the next fastest mode (proximate ad hoc) ($p=0.002$). There were 12 times more medical errors of moderate or greater clinical significance among utterances in non-RSMI encounters compared to RSMI encounters ($p=0.0002$).

CONCLUSIONS: Whereas limited by the small number of interpreters involved, our study found that RSMI resulted in fewer medical errors and was faster than non-RSMI methods of interpreting.

KEY WORDS: interpreting; medical errors; access to care; language; immigrant health.

J Gen Intern Med 22(Suppl 2):319–23

DOI: 10.1007/s11606-007-0361-7

© Society of General Internal Medicine 2007

BACKGROUND

Over 33 million U.S. residents are foreign-born.¹ Twenty-two million persons have limited English proficiency (LEP),² defined as “a limited ability to listen, speak, read, and write in

English, and speak[ing] English less than ‘very well’.”³ From 1990 to 2000, 46 states experienced an increase in their LEP populations, 15 of which had an increase of over 100% and 14 others of over 50%.² In New York City alone, nearly one-fourth of the population has LEP.⁴ Over 50% of the city’s Spanish-speaking population has LEP.⁴ Hospitals often call upon untrained staff or bystanders to interpret. Untrained interpreters are prone to editing, polishing, omissions, additions, substitutions, volunteered opinions, and confidentiality breaches.⁵ Patients who communicate through untrained, “ad hoc” interpreters are less satisfied with their patient-provider relationship compared with those in same language encounters.⁶ In encounters with bilingual nurses not trained in interpreting, approximately one-third of the uncomplicated and two-thirds of the complicated cases resulted in interpreting errors.⁷ Inadequately addressed language discordance between patient and provider inhibits LEP patients from receiving clinically indicated care, contributing to health disparities.^{8,9} Linguistic barriers limit patient education and adversely impact patient understanding and health outcomes.^{10,11} When information is not accurately understood or expressed, unnecessary, hazardous, or expensive diagnostic tests may be ordered, and indicated tests omitted.¹² Conversely, interpreter services have led to the improved use of preventive services resulting in cost savings.¹³ A systematic review of studies on professional interpreters in the medical encounter revealed that the use of professional interpreters results in improved clinical outcomes compared with ad hoc interpreters.¹⁴

Health care facilities are mandated to address interpreting needs. Title VI of the 1964 Civil Rights Act requires that all federally funded institutions make “reasonable” attempts to provide meaningful access to language services for LEP patients.¹⁵ Various strategies have been employed to bridge language gaps.^{16,17} Simultaneous interpreting is a near word-for-word running rendition performed within milliseconds of the original speech—nearly simultaneously—almost like a voiceover. In consecutive interpreting, the interpreter listens as the primary speaker speaks, and then interprets only once the primary speaker has finished. The person interpreting can either be located in the exam room or not. In proximate interpretation, the person interpreting is in the room with the provider and patient. In remote interpretation, the interpreter is located outside the interview room but is linked to the provider and patient through telecommunication. Remote simultaneous interpreting is most commonly associated with the United Nations, and is often referred to as UN-style

interpreting.¹⁸ Remote simultaneous interpreting has also been used effectively in the court system.¹⁹

Little is known about the impact of various interpreting strategies on interpreting speed and errors. This pilot study addresses this important gap by determining the accuracy and speed of four different medical interpreting strategies: remote simultaneous medical interpreting (RSMI), remote consecutive medical interpreting (RCMI), proximate consecutive medical interpreting (PCMI), and proximate ad hoc interpreting.

METHODS

Four scripted clinical encounters were run across the 4 interpreting methods, for a total of 16 interpreted encounters that generated the data for this study. Scripted encounters were used to enable the comparison of equivalent clinical content across interpreting modes. Six trained interpreters were randomly assigned to interpret in the 12 encounters involving RSMI, RCMI, and PCMI. In addition, 4 untrained, ad hoc personnel were each randomly assigned to interpret in encounters employing 1 of the 4 scripts (proximate ad hoc). The interpreters and ad hoc personnel were blind to the study design. Error analyzers, who were blinded to the method of interpreting they were analyzing, then scored each interpretation for linguistic and medical errors.

Script Development

Four patient–physician dialogues were prepared in English representing common primary care cases, including diabetes mellitus, tuberculosis testing, depression, and menopause. Each dialogue involved history taking, a brief physical exam, and discussions of diagnosis, testing, and treatment. The 4 dialogues were similar in length (1,366–1,500 words) and could be read in 6–9 minutes in English. The dialogues were constructed at similar levels of linguistic and medical difficulty. The patient portion was translated into Spanish and then backtranslated by the study team linguist to ensure accuracy.

Error Coding Methodology

Based on a review of the translation, interpreting, and linguistics literature,^{20–24} an error coding methodology was developed to measure both linguistic and medical errors (meaningful linguistic errors with medical information). Linguistic errors consisted of additions, omissions, or substitutions. Additions occurred when the interpreter added any language not uttered by the speaker(s). Omissions referred to lexical items present in the source language, which were left out by the interpreter. Substitutions occurred when the interpreter substituted material different from what was uttered by the speaker(s). A linguistic error was also considered to be a medical error if the language made reference to something that was medically related. Medical errors were considered clinically significant if they were likely to impact clinical decision-making and outcomes. Once the error was determined as medically related, the severity of the potential clinical consequences was noted. Five different categories were possible: clinically insignificant, mildly clinically significant, moderately clinically significant, highly clinically significant, and potentially life threatening. An error could fall into any of

the four “clinically significant” categories only if it altered the history in any way or had the potential to alter clinical outcomes. The severity of an error related to the level of harm that the misinterpretation could result in. For example, errors pertaining to dosages of medicine were considered highly clinically significant, although not life-threatening, as no medications were actually prescribed. An error in interpreting the age of a relative’s death from breast cancer at the age of 65 instead of 55 would have been considered mildly clinically significant.

Timing and Scoring for Errors

Ten people interpreted, including 6 trained interpreters who had received a standard, 64-hour training in consecutive and simultaneous interpreting—both proximate and remote—and 4 untrained personnel who served as ad hoc “interpreters”. Of the trained interpreters, the years of experience ranged from 0 to 2.75. Of the untrained “interpreters”, the years of experience ranged from 0 to 30. All 10 people interpret in a municipal clinic/hospital setting. Each of the trained interpreters was randomly assigned to 2 of the 3 trained interpreting methods (RSMI, RCMI, PCMI) across different scripts, resulting in 12 interpreted encounters using trained interpreters. No interpreter interpreted for the same script twice or in the same mode twice. Each of the 4 untrained “interpreters” participated once as a proximate ad hoc interpreter, each interpreting a different script. During each encounter, the assigned script was read aloud by an English-speaking physician and by a Spanish-speaking volunteer “patient”. Each physician–patient pair read each of the 4 scripts 4 times. Each time each script was read, it was interpreted by a different modality, resulting in 16 separate encounters. All encounters occurred in clinic-like settings and were audiotaped. A stopwatch was activated at the beginning and conclusion of each script reading. To ensure that the time-related learning curve by the doctor–patient pair would not influence study outcomes, we randomly alternated the order of the four interpreting modalities for each script.

A bilingual English–Spanish medical linguist who was blind to the study aims and design transcribed the 16 interpreted encounters. The transcriptions for each interpretation mode included the provider, patient, and interpreter portions. Two other blinded bilingual translators then checked the transcriptions. There was no indication in the transcriptions as to the interpreting mode or interpreter. All error reviewers underwent a half-day training on coding for errors.

Two of the authors (LO, KP) separately reviewed and scored all transcripts for linguistic and medical errors. A medically trained linguist (JG) was available to assist with any technical questions during linguistic error discussion, reviewed all linguistic decisions, and adjudicated any disagreements. A third bilingual clinician reviewer (FG) with extensive clinical, linguistic, and cultural competence experience participated in the medical error assessment, also to adjudicate disagreements. LO and KP are both bilingual. LO is a native Spanish-speaker and KP received formal education and medical training in Spanish and lived in a Spanish-speaking country. Both are fluent in the ability to speak, understand, read, and write in Spanish and English. JG is a native Spanish-speaker. All the error reviewers were blind to the interpreter and to the interpreting method in the transcripts.

Each transcript was divided into utterances, each of which constituted a phrase. Each utterance was scored using the described error coding methodology. Concordance between the medical error analyzers was 98.2%.

Data Analysis

Summary statistics (mean and standard deviation) were produced to summarize the outcome variables (time, number of linguistic errors, and number of potential medical errors). For the analysis of time, the analysis unit was encounters. For the number of linguistic errors and potential medical errors, we utilized utterances as units of analysis. Medical errors of greater than mild clinical significance were aggregated into a single outcome measure of “potential medical error”. Regression analysis was conducted to evaluate the effect of interpreting mode and identify risk factors associated with the number of medical errors. The units of analysis are not truly independent as they are part of a larger whole (same interpreter, or same script, or utterances that are in close proximity to one another in the script), all of which may influence their association with potential medical error. To handle this complex data structure, a log-linear mixed model was used in the regression analysis. The log-linear mixed model combines the log-linear model with random effects to facilitate analysis of counted data, whereas allowing structured dependence among units of analysis.²⁴ In this model, the number of potential medical errors was the outcome variable; the interpreter was included as a random effect; and interpreting mode (RSMI versus other), script (diabetes, menopause, depression, tuberculosis), complexity of the utterance (more than 1 concept, more than 10 words, moderate/complex), and interpreter-specific variables (completed training, more than 1 year experience) were considered as fixed effects. This model evaluated the effect of interpreting mode, while allowing correlations among the utterances from the same interpreter and adjusting for the effect associated with script, utterance complexity, and training and experience associated with the interpreters. The model fitting consisted of the following 3 steps, and the interpreter was included as a random effect throughout. We first examined the relation between each fixed effect and the outcome of potential medical error to obtain the unadjusted odds ratio. Second, a full model was constructed by including all fixed effects in the model. This allowed us to evaluate the effect of RSMI, in an attempt to correct for the impact that would be attributable to script type, script complexity, and

interpreters. Lastly, to identify the important risk factors associated with potential medical errors, a final model was obtained by the backward elimination method, and only those variables that were significant at the 5% level were included in the model.

Results

In terms of interpreting speed, RSMI encounters averaged 12.72 vs 18.24 minutes for the next fastest method of interpreting (proximate ad hoc) ($p=0.002$) (Table 1).

The 16 encounters yielded 1,909 utterances. Of these, 1,185 contained more than 1 meaningful concept. Across the different interpreting modes, RSMI produced far fewer errors than the other modes, which had error rates that were clustered at a significantly higher rate (Table 1). RSMI had a mean of 1.139 (SD=1.737) linguistic errors per utterance and 0.019 (SD=0.15) medical errors of moderate or greater clinical significance per utterance (Table 1).

The regression analysis showed that non-RSMI interpreting modalities were associated with a 12-fold greater rate of potential medical errors (of moderate or greater significance) per utterance compared to RSMI ($p=0.0002$), after adjusting for script type, interpreter’s experience and training, and utterance complexity. Other factors associated with potential medical errors included the number of concepts per utterance and interpreter experience. Utterances with more than 1 concept were associated with a 2.76 times greater rate of medical errors of moderate or greater clinical significance compared to utterances with only 1 concept. If an interpreter had less than 1 year of experience, she/he made 2.78 times more potential medical errors (of moderate or greater significance) per utterance than an interpreter with more than 1 year of experience ($p=0.0022$).

Examples of clinically significant medical errors included the following:

Mild clinical significance:

Doctor: Well, nowadays there are many different treatment options available. We could try using Wellbutrin 150 mg twice a day, and see what happens. This is a very effective drug and I think it would be very helpful. What do you think?

Interpreter: Sí, hoy en día hay diferentes opciones. Podemos tratar esta nueva que se llama Robutrin y coger 150 mg. diarios. ¿Cree usted que podemos tratar

Table 1. Time and Linguistic and Medical Errors of Moderate or Greater Clinical Significance, by Interpreting Method

	Time		Errors				
	N [†]	Mean (SD) in min	N [†]	Linguistic errors		Medical errors	
				Number of errors	Mean (SD) per utterance	Number of errors	Mean (SD) per utterance
RSMI	4	12.72 (3.23)	481	548	1.139 (1.737)	9	0.019 (0.15)
RCMI	4	23.7 (0.99)	454	892	1.965 (2.704)	67	0.148 (0.418)
PCMI	4	19.62 (1.1)	478	988	2.067 (2.494)	87	0.182 (0.528)
Ad hoc	4	18.24 (2.67)	496	861	1.736 (2.086)	47	0.095 (0.451)

RSMI: remote simultaneous medical interpreting with trained interpreters, RCMI: remote consecutive medical interpreting with trained interpreters, PCMI: proximate consecutive medical interpreting with trained interpreters, ad hoc: proximate consecutive medical interpreting with untrained “interpreters”.

*Encounters.

†Utterances.

esto? (Yes, there are different options nowadays. We can try this new one called Robutrin and get 150 mg daily. Do you think we can try this?)

Moderate clinical significance:

Doctor: You might otherwise get bad wounds without even feeling pain.

Interpreter: De otra forma, usted podría sentirse mejor sin sentir dolor. (Otherwise, you might feel better without feeling pain.)

High clinical significance:

Doctor: I'm going to have to introduce a small speculum to take a look at your cervix and do a new PAP smear.

Interpreter: Ella va a hago un... una instrumento... para chequear... para sacarlos unas cosas de ahí. (She is going to I do an... an instrument... to check... to take them out some things from there.)

Potentially life threatening:

Patient: Sí, pero me duele sentirme así. He cometido muchos errores en mi vida y supongo que tengo que pagar por ello ahora. No merezco vivir más. Quisiera estar muerto. (Yes, but I deserve to feel this way. I have made many mistakes in my life and I suppose that I have to pay for it now. I don't deserve to live anymore. I wish I were dead.)

Interpreter: Ah yes, but I suppose I ... it's due for me to feel this way. I've committed many...ahh... I've committed many bad things in my life.

CONCLUSIONS

We found that scripted encounters were more accurately and quickly interpreted with RSMI than with the more commonly used methods of RCMI, PCMI, and proximate consecutive ad hoc interpreting. RSMI resulted in fewer errors of clinical consequence compared with non-RSMI modes.

Whereas it is easy to understand why RSMI would take less time, given its simultaneous nature, RSMI was also more accurate. Our study does not allow us to understand why this may be the case, but it may be that clinical information is more accurately transmitted because there is minimal time lag. Interpreters do not need to recall large amounts of information. Furthermore, the minimal time allowed to interpret in RSMI may also inhibit editing or advocacy by the interpreters.

RSMI may present other advantages. The time savings afforded by RSMI may allow the clinician to be more patient-centered and address issues beyond the technical aspects of care. Limited previous research on RSMI demonstrates increased physician and patient engagement. Hornberger demonstrated increased physician and patient utterances, questions and explanations, and physician and patient satisfaction when comparing RSMI with PCMI.¹⁷

The degree of training that is needed for RSMI interpretation as opposed to consecutive methods is marginally more. Trainings at our institution are 48 hours for introduction to consecutive medical interpreting and 60 hours for introduction to simultaneous medical interpreting.

Our method of determining the clinical significance of potential medical errors was necessarily subjective, and it is not possible from our controlled study to estimate the impact that the coded errors might have had in actual clinical encounters. The high rate of agreement between coders suggests that the method was reliable. However, we believe that our use of standardized dialogues, whereas allowing for fair comparison across modes of interpretation may have resulted in an undercount of potential medical errors. Regardless of the number and severity of the errors in interpretation, the actors adhered to the scripted dialogue. In a real-life interpretation, such errors could have resulted in critical mistakes. Furthermore, there were multiple low-grade errors, which synergistically could produce errors of greater severity and consequence. The scripts were weighted more heavily on diagnostic as opposed to therapeutic aspects of the medical encounter, thereby limiting the potential for assessed clinical consequences. In the patient-physician dialogues, interpreters struggled with the names of medications and with dosages. Medication usage was discussed, but not actually initiated. As a result, there was a tendency to rank medication errors as being of lesser clinical significance than they would be in real-life circumstances, when medications are actually prescribed.

Despite the fact that a standard training and quality management program was used, the results from this study cannot necessarily be generalized to all interpreter services or clinical settings. Our study only examined the performance of 6 interpreters using only 1 non-English language (Spanish). It is possible that outcomes might differ in settings with different staffing, training, languages, or quality management programs. It is also possible that both absolute and relative error rates might be different in real-life, as opposed to the simulated, scripted clinical encounters that we used to "level the playing field" across different interpreting methods.

The United States is home to an increasingly diverse population. Immigrants face multiple barriers to effective health care. Of these, language is key. Efficient, accurate medical interpreting strategies need to be widely disseminated. We found that RSMI provided advantages in terms of speed and accuracy, when compared to other interpreting modalities, and may therefore be a promising option for patients needing language assistance. Future studies are needed, however, to examine the comparative advantages and disadvantages, and the cost-effectiveness of different interpreting modalities across clinical settings and patient populations.

Acknowledgements: This work was generously supported by grants from the Commonwealth Fund and the California Endowment. Dr. Kavitha Prakash was supported as a fellow by the Commonwealth Fund for her work on the study. Thank you to David Abramson, Ph.D. and Tasha Stehling, M.P.H. for their statistical support.

Conflict of Interest: Two of the authors (FG, JC) have ownership in a company that provides technology solutions for remote simultaneous medical interpreting.

Corresponding Author: Francesca Gany, MD, MS; Center for Immigrant Health, New York University School of Medicine, OBV-D-402, 550 First Avenue, New York, NY 10016, USA (e-mail: fg12@nyu.edu).

REFERENCES

- Larsen LJ, U.S. Census Bureau.** The foreign-born population in the United States: 2003. Washington, DC: U.S. Census Bureau; 2004; Current Population Reports, P20-551:1-9. Available at: <http://www.census.gov/prod/2004pubs/p20-551.pdf>. Accessed May 20, 2005.
- Shin BH, Bruno R.** Language use and English speaking ability: census 2000 brief. Washington, DC: US Census Bureau; 2003. Available at: <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf>. Accessed May 20, 2005.
- Kwang K, Collins M, McArthur E.** Participation of adults in English as a second language classes: 1994-1995. Washington, DC: National Center for Education Statistics; 1997. Available at: <http://nces.ed.gov/pub97/97319.pdf>. Accessed August 5, 2000.
- Lobo AP, Salvo JJ; New York Department of City Planning, Population Division.** The newest New Yorkers, 2000 immigrant New York in the new millennium. New York: New York City Department of Planning, Population Division; January 2005.
- Downing B, Roat C.** Models for the provision of language access in health care settings. Albany, NY: National Council on Interpreting in Health Care; 2002. Available at: http://www.ncihc.org/NCIHC_PDF/ModelsfortheProvisionofLanguageAccessinHealthCareSettings.pdf. Accessed March 2006.
- Baker DW, Hayes R, Fortier JP.** Interpreter use and satisfaction with interpersonal aspects of care for Spanish-speaking patients. *Med Care*. 1998;36(10):1461-70.
- Elderkin-Thompson V, Silver RC, Waitzkin H.** When nurses double as interpreters: a study of Spanish-speaking patients in a US primary care setting. *Soc Sci Med*. 2001;52(9):1343-58.
- Ku L, Flores G.** Pay now or pay later: providing interpreter services in health care. *Health Aff (Millwood)* 2005;24:435-44.
- Woloshin S, Bickell N, Schwartz L, Gany F, Welch G.** Language barriers in medicine in the United States. *JAMA*. 1995;273:724-8.
- Timmings CL.** The impact of language barriers on the health care of Latinos in the United States: a review of the literature and guidelines for practice. *J Midwifery Womens Health*. 2002;47:80-96.
- Gandhi TK, Burstin HR, Cook EF, Puopolo AL, Haas JS, Brennan TA, Bates DW.** Drug complications in outpatients. *J Gen Intern Med*. 2000;15:149-54.
- Hampers LC, Cha S, Gutglass DJ, Binns JH, Krug SE.** Language barriers and resource utilization in a pediatric emergency department. *Pediatrics*. 1999;103:1253-6.
- Jacobs EA, Shepard DS, Suaya JA, Stone EL.** Overcoming language barriers in health care: costs and benefits of interpreter services. *Am J Public Health*. 2004;94(5):866-9.
- Karliner LS, Jacobs EA, Chen AH, Mutha S.** Do professional interpreters improve clinical care for patients with limited English proficiency? A systematic review of the literature. *Health Serv Res*. 2007;42(2):727-54.
- U.S. Environmental Protection Agency, Office of Civil Rights.** Title VI of the Civil Rights Act of 1964, as amended (42 U.S.C. §§2000d to 2000d-7)
- Chang PH, Fortier JP.** Language barriers to health care: an overview. *J Health Care Poor Underserved*. 1998;9(supp/1):S5-20.
- Hornberger JC.** Eliminating language barriers for non-English-speaking patients. *Med Care*. 1996;34(8):845-56.
- Mouzourakis P.** Remote interpreting: a technical perspective on recent experiments. *Interpreting*. 2006;8(1):45-66.
- Telephone interpreting: a long-distance success in saving money. Available at <http://www.uscourts.gov/newsroom/tip.htm>. Accessed on May 16, 2007
- Altman J.** Error analysis in the teaching of simultaneous interpreting: a pilot study. In Lambert S, Moser-Mercer B, eds. *Bridging the Gap. Empirical Research in Simultaneous Interpretation*. Philadelphia, PA: John Benjamins, 1994.
- Barik H.** A description of various types of omissions, additions and errors of translation encountered in simultaneous interpretation. In Lambert S, Moser-Mercer B, eds. *Bridging the Gap. Empirical Research in Simultaneous Interpretation*. Philadelphia, PA: John Benjamins, 1994.
- Bugel-Shunra D.** Proofreading translations: a checklist, not a blank check. *ATA Chronicle*. 2000;29(October):12.
- Kussmaul P.** Training the Translator. Chapter 6: Evaluation and errors. Philadelphia, PA: John Benjamins, 1995.
- MacCulloch CE, Searle SR.** Generalized, Linear, and Mixed Models. Hoboken, NJ: Wiley, 2001.