

# Linguistic fairness in the U.S.: The case of multilingual public health information about COVID-19

Damián E. Blasi<sup>a,b,c,d,1</sup>, Vishala Mishra<sup>e,1</sup>, Adolfo M. García<sup>f,g,h,i</sup>, and Joseph P. Dexter<sup>a,b,1</sup>

<sup>a</sup>Data Science Initiative, Harvard University, Cambridge, MA 02138; <sup>b</sup>Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138; <sup>c</sup>Max Planck Institute for the Science of Human History, Jena, Germany; <sup>d</sup>Linguistic Convergence Laboratory, HSE University, Moscow, Russian Federation; <sup>e</sup>Multidisciplinary Research Unit, Madras Medical College, Chennai, India; <sup>f</sup>Cognitive Neuroscience Center (CNC), Universidad de San Andrés, Buenos Aires, Argentina; <sup>g</sup>National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina; <sup>h</sup>Departamento de Lingüística y Literatura, Facultad de Humanidades, Universidad de Santiago de Chile, Santiago, Chile; <sup>i</sup>Global Brain Health Institute, University of California, San Francisco, San Francisco, CA 94143

This manuscript was compiled on September 27, 2021

**Lack of high-quality multilingual resources can contribute to disparities in the availability of medical and public health information. The COVID-19 pandemic has required rapid dissemination of essential guidance to diverse audiences and therefore provides an ideal context in which to study linguistic fairness in the U.S. Here we report a cross-sectional study of official non-English information about COVID-19 from the Centers for Disease Control and Prevention, the Food and Drug Administration, and the health departments of all 50 U.S. states. We find that multilingual information is limited in many states, such that almost half of all individuals not proficient in English or Spanish lack access to state-specific COVID-19 guidance in their primary language. Although Spanish-language information is widely available, we show using automated readability formulas that most materials do not follow standard recommendations for clear communication in medicine and public health. In combination, our results provide a snapshot of linguistic unfairness across the U.S. and highlight an urgent need for the creation of plain language, multilingual resources about COVID-19.**

COVID-19 | health equity | health literacy | linguistic fairness | readability

In culturally and linguistically diverse countries, the availability of medical information in multiple languages is a key determinant of health equity, but efforts to increase linguistic fairness in U.S. healthcare remain suboptimal (1–3). Over the past year, the COVID-19 pandemic has required rapid dissemination of critical public health information across the country. Such communication efforts have been complicated by longstanding gaps in the health literacy of the U.S. population (for instance, only 12% of U.S. adults scored as proficient in the 2003 National Assessment of Health Literacy (4)). Similar issues have been identified for information about COVID-19, and individuals with low health literacy report reduced preparedness for the pandemic (5–9). The situation is even more challenging for non-native English speakers; when combined with low health literacy, limited English proficiency is associated with poor health status in the U.S. (4, 10, 11). Accordingly, public guidance about COVID-19 should be equitable in its language coverage and use accessible communication strategies. Moreover, the extent to which these dual imperatives are being met provides a valuable benchmark for the state of linguistic fairness in the country.

The U.S. population includes 48 million immigrants, a figure that is projected to increase to at least 75 million by 2065 (12). Reflecting the diversity of both its immigrant and indigenous

populations, the U.S. is a deeply multilingual country. Of its more than 330 million residents, 67 million speak a language other than English at home (13), and 25 million have limited English proficiency (an increase of 156% since 1980) (14). Aside from English, by far the most widely spoken language is Spanish (40 million speakers), followed by Chinese (3.3 million), Tagalog (1.7 million), Vietnamese (1.5 million), and French (1.2 million) (13). The quality of many health resources for non-English users, however, is far from ideal (15, 16), contributing to health disparities nationwide and amplifying structural inequalities accentuated by the pandemic (17–19).

U.S. residents receive information about COVID-19 from a diverse selection of official sources, ranging from federal agencies such as the Centers for Disease Control and Prevention (CDC) and Food and Drug Administration (FDA) to state, county, and local health departments. Although federal guidance is subject to the requirements of the Plain Language Act and dissemination of actionable, easy-to-use health information is a major goal of the Healthy People 2030 initiative (3), the accessibility of official COVID-19 information has often been limited (20–24). Compounding these issues is a lack of uniform standards for the provision of multilingual resources, as well as the Trump administration’s rollback of a federal rule requiring that patients be informed of their right to language interpretation services (25).

In addition to its immediate relevance for improving the equity of the pandemic response, a thorough understanding of the availability and accessibility of non-Anglophone COVID-19 guidance is likely to have longer-term implications. The scope of material written about COVID-19 is unprecedented, and the existence of extensive guidance from all U.S. states creates an opportunity for comparative evaluation of linguistic fairness. Moreover, the availability and quality of Spanish-language guidance is of particular importance because of the size of the user base and the disparate impact the pandemic has had on the Latinx community (26, 27). To this end, we undertook a two-part analysis, first of the equity of language coverage in state-level guidance about COVID-19 and second of the readability of Spanish information from both federal

Author contributions: D.E.B., V.M., A.M.G., and J.P.D. designed research; D.E.B., V.M., and J.P.D. performed research; D.E.B. contributed new reagents/analytic tools; D.E.B., V.M., A.M.G., and J.P.D. analyzed data; D.E.B., V.M., A.M.G., and J.P.D. wrote the paper.

The authors declare no competing interests.

<sup>1</sup>To whom correspondence should be addressed. E-mail: [dblas@fas.harvard.edu](mailto:dblas@fas.harvard.edu); [mishra.vishala@gmail.com](mailto:mishra.vishala@gmail.com); [jdexter@fas.harvard.edu](mailto:jdexter@fas.harvard.edu)

Resource or Tool	Number (%) of Websites
Automated translation	30 (59%)
CDC multilingual resources	29 (57%)
American Sign Language videos	34 (67%)

**Table 1. Prevalence of key resources across all states and the District of Columbia.** Resources were tabulated from a review of official health department websites.

and state sources. We find that many states have substantial linguistic communities not served by existing guidance and that most Hispanophone information, regardless of source, is not in compliance with well-known recommendations for clear communication.

## Results

**Prevalence of baseline resources.** Although automated translation services such as Google Translate have substantial limitations (28), providing a machine translation plug-in is a straightforward step to increase the accessibility of a website. We found, however, that only 30 of 51 (59%) health department websites included an option for automated translation of information about COVID-19 (Table 1). Similarly, just 29 (57%) websites referenced or linked to the CDC’s large collection of multilingual resources written for the general public, and 34 (67%) included at least one video in American Sign Language (Table 1).

**Linguistic coverage of COVID-19 guidelines across U.S. states.** We then considered the extent to which state-level information about COVID-19 serves different linguistic communities. In particular, we sought to determine whether the availability of public health guidelines is demographically fair (i.e., if it can be accounted for primarily by the number of users of a given language). For each language and state, we cross-referenced the number of speakers (as reported in the U.S. Census Bureau’s American Community Survey (13)) with the presence of COVID-19 materials for that community (Materials and Methods). Given that Spanish COVID-19 guidelines are available for almost all states (48 of 51 websites), we focused exclusively on speakers of languages other than English and Spanish. We modeled the availability of COVID-19 guidelines for a language  $j$  in state  $i$  following a Bayesian generalized linear model (Materials and Methods):

$$g^{-1}(R_{ij}) \propto \alpha_i + \beta_j + f(n_{ij}), \quad [1]$$

where  $R_{ij}$  is a binary variable (presence or absence of COVID-19 guidelines),  $\alpha$  and  $\beta$  are group-specific effects for state and language, respectively,  $n_{ij}$  is the corresponding number of speakers,  $f(\cdot)$  is a smooth function to be learned from the data, and  $g(\cdot)$  is the logistic function.

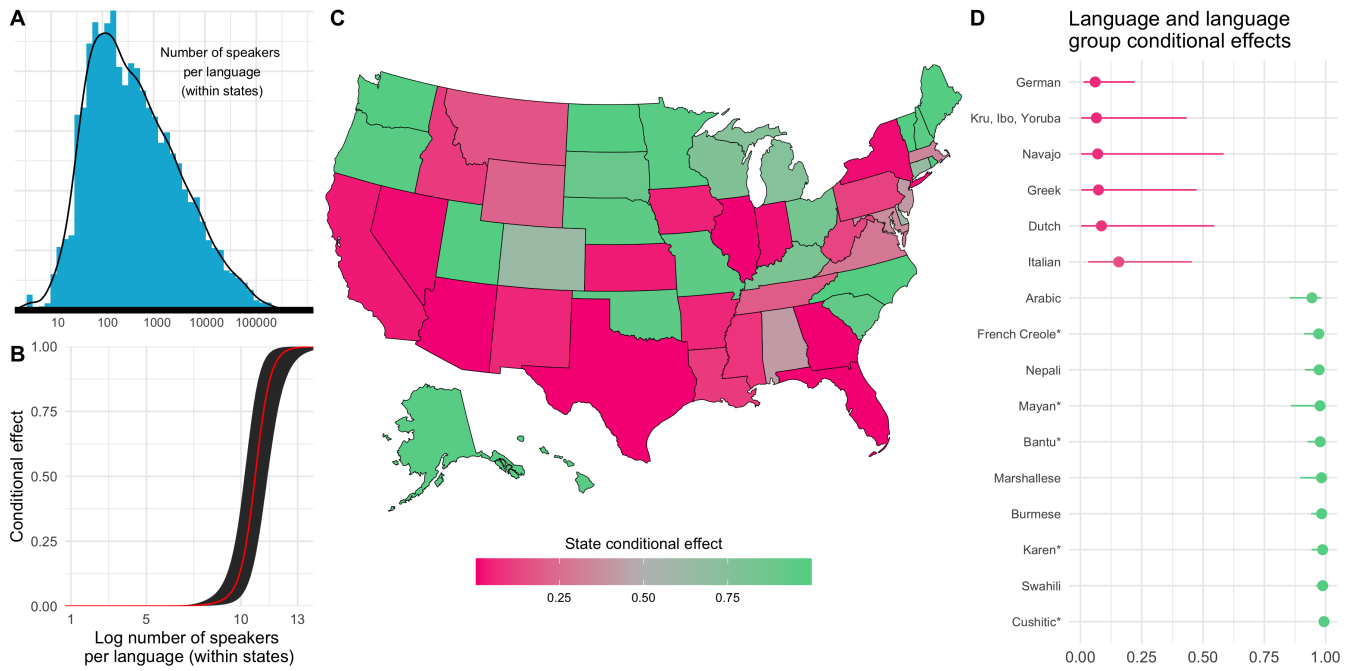
Results of the analysis are summarized in Fig. 1. When considering size alone, only large language communities with 50,000 or more individuals are predicted to be covered. Language and state conditional effects also account for a substantial amount of variance, and large states with many small linguistic communities tend to score worse in the fairness of their coverage. Several East African and Southeast Asian languages are comparatively better represented, whereas some Western European languages (German, Dutch, Greek, and

Italian) and a handful of large West African languages (Igbo, Yoruba, and Kru) and Navajo are less well-covered, other factors being equal. The users of these languages tend to be proficient in English, which might explain why specific health guidelines were not developed for them. We tested this hypothesis on a subset of the Census data for which the number of users who “speak English less than very well” was indicated (67% of the full dataset). Bayesian model selection corroborated that it is the number of users with limited English proficiency - rather than the total number of users - which better predicts the presence of COVID-19 guidelines (Materials and Methods).

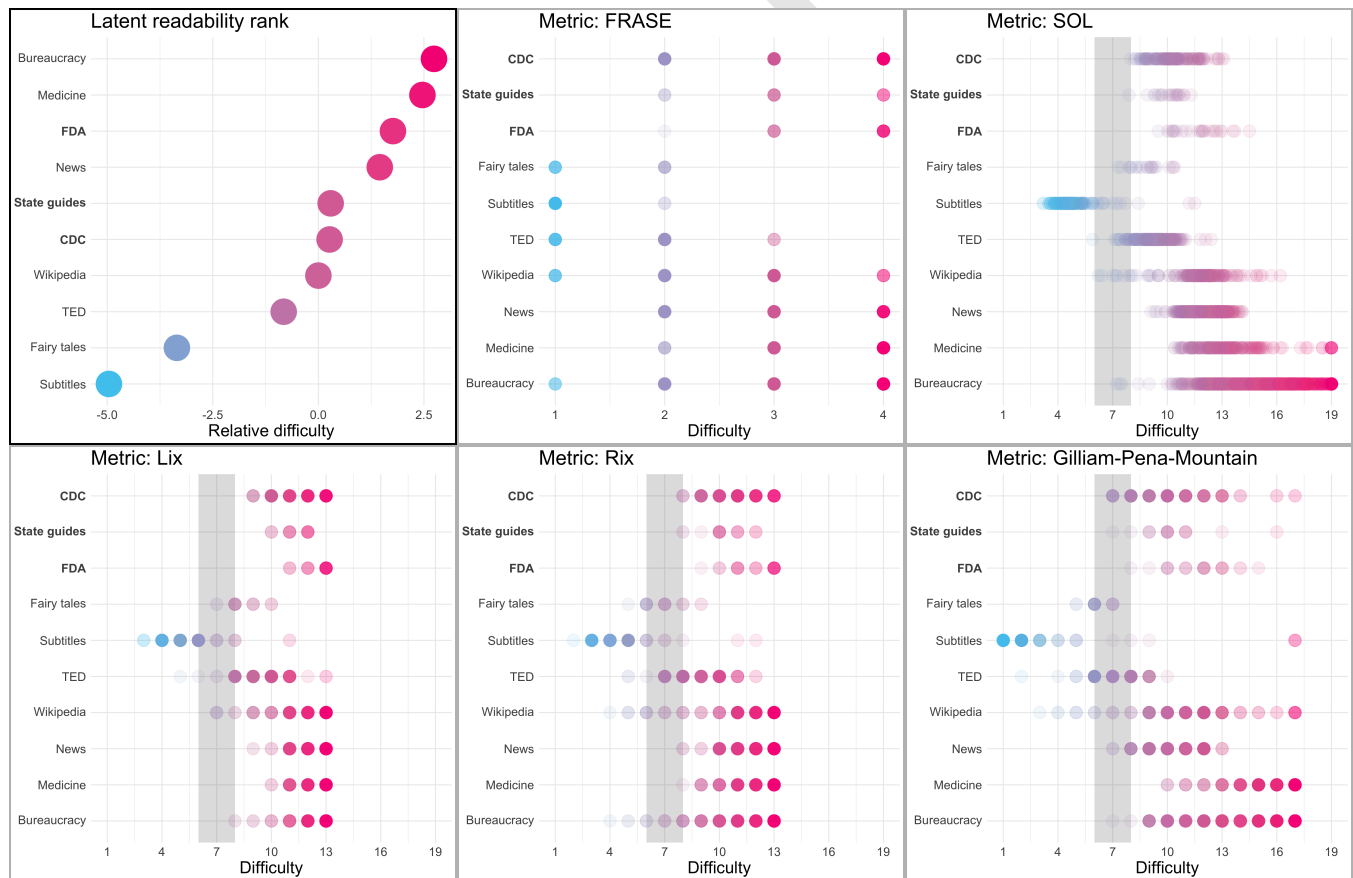
**Readability of Spanish-language COVID-19 materials.** Availability of official guidance in a particular language does not guarantee accessibility and ease of use. Although evaluation of all resources in our multilingual dataset would be challenging, a restricted analysis of the Spanish documents is useful for several reasons. There are more speakers of Spanish in the U.S. than of all other non-English languages combined, and Spanish-language information about COVID-19 is provided by most official sources (Fig. 1). Given the large user base and the wide availability of both manual and automatic English-to-Spanish translation, it is likely that the quality of Spanish resources from the CDC, FDA, and state health departments represents an effective upper bound for the quality of official materials in other languages.

To approximate the text difficulty of Spanish-language COVID-19 information as part of a rapid response to the pandemic, we applied four automated readability formulas to our corpora of public health documents (Materials and Methods). Readability formulas attempt to predict the difficulty of a text, usually expressed as a grade level, and have been used to pinpoint important limitations in Anglophone public health information about COVID-19 (20, 21, 23). These formulas are based on the assumption that, other things being equal, longer linguistic units, such as words and sentences, hinder comprehension. For each of the metrics considered, we found that the vast majority of CDC, FDA, and state documents exceeded an eighth-grade reading level (Fig. 2), which is the CDC, National Institutes of Health, and American Medical Association’s recommended maximum difficulty for health information written for the general public (29–31). Although readability formulas should not be treated as a substitute for reading comprehension data obtained from surveys or focus groups, these results suggest an urgent need for further investigation and development of more accessible communication strategies.

Readability measures are not without problems when used as unbiased estimates of overall comprehension and reading level (32–35), in particular for languages other than English (36). Nevertheless, they may provide partial information about the difficulty of one text relative to another, as word length and sentence length - the basis for most readability formulas - are implicated in reading ease and speed (37, 38). To contextualize the results obtained for the public health documents, we also applied the readability metrics to thirteen other corpora covering a wide range of Spanish prose texts of varying genre and difficulty; these corpora include fairy tales, television and film subtitles, transcripts of TED talks, Wikipedia articles, news articles, and medical and bureaucratic documents. We compared the readability of the CDC, FDA, and state ma-



**Fig. 1. Linguistic diversity and fairness of official COVID-19 information provided by U.S. states.** **A** Distribution of number of users per language within states. **B** Conditional effect of (log) number of users of a language on presence of COVID-19 guidelines. **C** U.S. map depicting state-specific effects on presence of COVID-19 guidelines. **D** Bayesian 95% credible intervals for language conditional effects on presence of COVID-19 guidelines.



**Fig. 2. Readability of official Spanish-language information about COVID-19.** The figure shows difficulty predictions for the public health and benchmark corpora from five readability metrics - FRASE Graph (FRASE), SOL, Läsbarhetsindex (Lix), Rate Index (Rix), and Gilliam-Pena-Mountain Graph (Gilliam-Pena-Mountain) - as well as a latent rank. The gray bars indicate the recommended reading level for medical information written for the general public (grades 6-8).

173 materials to the other corpora using all five measures (Fig. 2),  
174 and we inferred a latent rank from a Plackett-Luce model  
175 (Materials and Methods). As shown in the top left panel in  
176 Fig. 2, CDC and state pages were found to be more difficult  
177 than the simplest corpora (subtitles, fairy tales, and TED talk  
178 transcripts). The FDA documents ranked among the most  
179 complex corpora (news, medicine, and bureaucracy).

## 180 Discussion

181 Here we analyzed the U.S.-wide distribution of COVID-19  
182 guidelines written in languages other than English. The emerg-  
183 ing picture is that official COVID-19 information is available  
184 for languages with many users who have limited English pro-  
185 ficiency but scarce for smaller linguistic communities. While  
186 this trend is modulated by the state and the specific language  
187 under consideration, the sheer numbers are disheartening (21).  
188 Roughly half of all individuals (47%, or 4.2 million people) who  
189 are not proficient in either English or Spanish lack access to  
190 COVID-19 guidelines in their native language (Materials and  
191 Methods). Compounding this inequity in language coverage  
192 is the limited availability even of baseline, easy-to-implement  
193 resources for non-English users; for instance, more than 40% of  
194 state health departments do not offer automated translation of  
195 their online information about COVID-19. Although guidance  
196 in Spanish is widely available (coverage by the CDC, FDA,  
197 and 48 out of 51 state websites), we find that the readability  
198 of this material far exceeds the recommended eighth-grade  
199 level according to five automated metrics (29–31).

200 Despite the well-known limitations of readability formulas,  
201 including focus on shallow linguistic features and often  
202 incomplete validation (32–35), we were able to confirm their  
203 reliability for relative assessment of text difficulty; in partic-  
204 ular, all metrics considered yielded intuitively reasonable  
205 difficulty rankings across a diverse collection of Spanish prose.  
206 While our results identify points of urgent concern for the  
207 ongoing pandemic response, further work is needed to expand  
208 the scope of evaluation to include multimedia resources and  
209 traditional, offline media, as well as guidance issued by mun-  
210 icipal and county health officials (39). Such an expanded  
211 evaluation should seek to integrate statistical and computa-  
212 tional evidence with empirical assessments of comprehension  
213 and usability testing of multilingual resources (9, 40).

214 Despite extensive efforts to address the COVID-19 “info-  
215 demic” (41), proliferation of misinformation, especially about  
216 COVID-19 vaccines, remains a persistent issue (42, 43). Our  
217 results thus point to a concerning combination of structural fac-  
218 tors - incomplete coverage of multilingual resources, restricted  
219 availability of plain language materials, and, as documented  
220 previously, poor media ecology - that make it particularly  
221 challenging for U.S. residents with limited English proficiency  
222 to access trustworthy information. In light of the limitations of  
223 official information, many private groups have worked during  
224 the pandemic to improve health equity through crowdsourced  
225 translation projects, creation of accessible, multilingual pub-  
226 lic service announcements, and other initiatives (39, 44, 45).  
227 These efforts could serve as a blueprint for future efforts that  
228 address the challenges identified by our analysis and increase  
229 the accessibility of health information for diverse audiences.

## 230 Materials and Methods

232 **Annotation of multilingual state resources.** Between March 1, 2021  
233 and March 15, 2021 we reviewed the websites of the health depart-  
234 ments of all 50 states and the District of Columbia for availability  
235 of non-English information about COVID-19. For each state we  
236 recorded the number of languages represented, as well as the pres-  
237 ence or absence of an option for automated translation of the  
238 website, reference to the CDC’s large collection of multilingual re-  
239 sources (<https://www.cdc.gov/pubs/other-languages?Sort=Lang%3A%3Aasc>), and American Sign Language videos. To minimize sub-  
240 jectivity in the annotation process we did not assess content or  
241 accessibility of resources; a state received credit for a language if  
242 its website included any content (whether original or linked directly  
243 from an outside resource) in that language.  
244

245 State-level information about the number of users of each lan-  
246 guage considered was obtained from the 2009-2013 American Com-  
247 munity Survey of the U.S. Census Bureau (<https://www.census.gov/data/tables/2013/demo/2009-2013-lang-tables.html>). This information  
248 was used in our language coverage model and to estimate the total  
249 number of individuals not served by state COVID-19 guidance. The  
250 linguistic annotation in the Census data is of widely varying quality.  
251 While language names often are given directly, a number of coarser  
252 language labels are used as well (e.g., by aggregating all users of  
253 a language family under a single label or by using coarse regional  
254 terms). Whenever possible, we cross-referenced our dataset with  
255 the Census information using the widest language label available,  
256 so as to overestimate the coverage of state guidelines (see Appendix  
257 for details).  
258

259 **Linguistic coverage model.** We implemented the model described in  
260 Eq. (1) in a Bayesian framework, utilizing the No-U-Turn Sampler  
261 as our Markov Chain Monte Carlo algorithm of choice (46). We used  
262 the default weakly informative priors provided by the `brms` package,  
263 version 2.15.0 (47). Posterior predictive checks revealed that the  
264 models succeeded in capturing the distribution of the observed data.

265 We then considered the subset of data for which information  
266 about English proficiency was available. On this dataset we deployed  
267 the model described above, which we contrasted with an equivalent  
268 model in which the predictor  $n_{ij}$  of Eq. (1) was replaced by the  
269 number of users of that language with limited English proficiency  
270 (plus one, as some languages are reported to have no proficient  
271 English users). We performed model selection using the difference  
272 in expected log pointwise predictive density between the two models  
273 (48), which yielded -15.3 (SE=5.3), favoring the second model over  
274 the first one.

275 **Readability and text complexity analysis.** We analyzed the readabil-  
276 ity and text complexity of online public health information about  
277 COVID-19 and 14 benchmark corpora containing a diverse selection  
278 of Spanish prose. The public health corpora were drawn from three  
279 sources, the CDC, FDA, and state health departments. The CDC  
280 corpus contains 153 web pages from the organization’s main Spanish-  
281 language website (all pages under the “Su salud” and “Vacunas” tabs  
282 of <https://espanol.cdc.gov/coronavirus/2019-ncov/index.html>, except for  
283 landing pages and those with primarily non-textual content). The  
284 FDA corpus contains 26 web pages, such as lists of frequently asked  
285 questions and vaccine information sheets, from <https://www.fda.gov/about-fda/fda-en-espanol/enfermedad-del-coronavirus-covid-19>. The  
286 state corpus contains 21 web pages from the 12 states that provided  
287 a list of frequently asked questions or similar material in Spanish  
288 and rank in the top 20 by fraction of Spanish speakers (Texas, Cali-  
289 fornia, New Mexico, Florida, Arizona, New Jersey, Illinois, Rhode  
290 Island, Utah, Oregon, Washington, and Kansas). CDC and FDA  
291 web pages were scraped on March 24, 2021; state web pages were  
292 scraped on April 24, 2021. Readability Studio Professional, version  
293 2020 (Oleander Software) was used for text preprocessing, such as  
294 removal of headers and footers, figure captions, and other extraneous  
295 content.  
296

297 The benchmark corpora include Spanish translations of 21 of  
298 Grimm’s Fairy Tales ([https://www.grimmstories.com/es/grimm\\_cuentos/favorites](https://www.grimmstories.com/es/grimm_cuentos/favorites)), as well as 11 sections of the Spanish Unannotated Cor-  
299 pora (SUA) resource (<https://github.com/josecannete/spanish-corpora>).  
300 These corpora are large (the SUA contains approximately 3 billion  
301 tokens in total) and cover a range of document types, including  
302 Wikipedia and news articles, government documents, and tran-  
303 scripts of speeches. For our analysis we downsampled the corpora  
304

305 by choosing 200 documents at random from each corpus.

306 We applied five standardized readability formulas, the FRASE  
307 Graph, Gilliam-Peña-Mountain Graph, Läsbarhetsindex (Lix), Rate  
308 Index (Rix), and SOL, to the public health documents and bench-  
309 mark corpora. Each formula was developed or adapted specifically  
310 for Spanish and has been applied in previous health literacy stud-  
311 ies (30). The output of the FRASE formula is a categorical estimate  
312 of text difficulty (beginning, intermediate, advanced intermediate,  
313 advanced); the output of the other four formulas is a predicted  
314 grade level. For our analysis, texts scoring as “13+” by Lix or Rix  
315 or “19+” by SOL were reassigned scores of 13 and 19, respectively,  
316 and texts scored as “too difficult to be classified” by Gilliam-Peña-  
317 Mountain were reassigned a score of 17 (the maximum possible  
318 value). Readability scores were calculated using Readability Studio  
319 Professional.

320 Finally, we inferred a latent readability rank for each corpus  
321 class across all five readability measures using the R package  
322 `PlackettLuce` implementation of the Plackett-Luce model, which  
323 represents rankings through a continuous latent variable (49).

324 **ACKNOWLEDGMENTS.** We thank Fernando Alva Manchego and  
325 Pramit Chaudhuri for helpful discussions. D.E.B. was supported  
326 by a Harvard Data Science Fellowship and the HSE University  
327 Basic Research Program, funded by the Russian Academic Excel-  
328 lence Project “5-100.” A.M.G. was supported by CONICET; ANID,  
329 FONDECYT Regular (grant numbers 1210176 and 1210195); and  
330 Programa Interdisciplinario de Investigación Experimental en Comu-  
331 nicación y Cognición (PIIECC), Facultad de Humanidades, USACH.  
332 J.P.D. was supported by a Harvard Data Science Fellowship and a  
333 CoronaVirus Facts Alliance Grant from the Poynter Institute.

334 1. EM Feuerherm, RE Showstack, MG Santos, GA Martinez, HE Jacobson, Language as a  
335 social determinant of health: Partnerships for health equity in *Extending Applied Linguistics*  
336 *for Social Impact: Cross-Disciplinary Collaborations in Diverse Spaces of Public Inquiry*, eds.  
337 DS Warriner, ER Miller. (Bloomsbury Publishing, New York, New York), pp. 125–148 (2021).  
338 2. J Espinoza, S Derrington, How should clinicians respond to language barriers that exacerbate  
339 health inequity? *AMA J. Ethics* **23**, E109–E116 (2021).  
340 3. United States Department of Health and Human Services, Healthy People 2030 (<https://health.gov/healthypeople>) (2020).  
341 4. M Kutner, E Greenburg, Y Jin, C Paulsen, S White, The health literacy of America's adults:  
342 Results from the 2003 National Assessment of Adult Literacy ([https://nces.ed.gov/pubst2006/](https://nces.ed.gov/pubst2006/2006483.pdf)  
343 [2006483.pdf](https://nces.ed.gov/pubst2006/2006483.pdf)) (2006).  
344 5. O Okan, et al., Coronavirus-related health literacy: A cross-sectional study in adults during  
345 the COVID-19 infodemic in Germany. *Int. J. Environ. Res. Public Heal.* **17**, 5503 (2020).  
346 6. E Bilancini, L Boncinelli, V Capraro, T Celadin, R Di Paolo, The effect of norm-based mes-  
347 sages on reading and understanding COVID-19 pandemic response governmental rules. *J.*  
348 *Behav. Econ. for Policy* **4**, 45–55 (2020).  
349 7. I Montagni, et al., Acceptance of a Covid-19 vaccine is associated with ability to detect fake  
350 news and health literacy. *J. Public Heal.* **2021**, 1–8 (2021).  
351 8. MS Wolf, et al., Awareness, attitudes, and actions related to COVID-19 among adults with  
352 chronic conditions at the onset of the US outbreak: a cross-sectional survey. *Annals Intern.*  
353 *Medicine* **173**, 100–109 (2020).  
354 9. V Mishra, JP Dexter, Response of unvaccinated US adults to official information about the  
355 pause in use of the Johnson & Johnson-Janssen COVID-19 vaccine. *medRxiv* **2021** (2021).  
356 10. T Sentell, KL Braun, Low health literacy, limited English proficiency, and health status in  
357 Asians, Latinos, and other racial/ethnic groups in California. *J. Heal. Commun.* **17**, 82–99  
358 (2012).  
359 11. LC Diamond, DS Reuland, Communicating with Spanish-speaking patients—Reply. *J. Am.*  
360 *Med. Assoc.* **301**, 2327–2328 (2009).  
361 12. A Budiman, Key findings about U.S. immigrants ([https://www.pewresearch.org/fact-tank/](https://www.pewresearch.org/fact-tank/2020/08/20/key-findings-about-u-s-immigrants/)  
362 [2020/08/20/key-findings-about-u-s-immigrants/](https://www.pewresearch.org/fact-tank/2020/08/20/key-findings-about-u-s-immigrants/)) (2020).  
363 13. United States Census Bureau, American Community Survey ([https://www.census.gov/](https://www.census.gov/programs-surveys/acs/)  
364 [programs-surveys/acs/](https://www.census.gov/programs-surveys/acs/)) (2015).  
365 14. J Batalova, J Zong, Language diversity and English proficiency in the United States. *Migr. Inf.*  
366 *Source* **2016** (2016).  
367 15. E Jacobs, K Karavolos, P Rathouz, T Ferris, L Powell, Limited English proficiency and breast  
368 and cervical cancer screening in a multiethnic population. *Am. J. Public Heal.* **95**, 1410–1416  
369 (2005).  
370 16. C Honeycutt, et al., Assessment of Spanish translation of websites at top-ranked US hospitals.  
371 *JAMA Netw. Open* **4**, e2037196 (2021).  
372 17. D Khullar, DA Chokshi, Challenges for immigrant health in the USA—the road to crisis. *Lancet*  
373 **393**, 2168–2174 (2019).  
374 18. A A-Oraibi, Migrant health is public health: a call for equitable access to COVID-19 vaccines.  
375 *Lancet Public Heal.* **6**, e144 (2021).  
376 19. R Khazanchi, C Evans, J Marcelin, Racism, not race, drives inequality across the COVID-19  
377 continuum. *JAMA Netw. Open* **3**, e2019933 (2020).  
378 20. V Mishra, J Dexter, Comparison of readability of official public health information about  
379 COVID-19 on websites of international agencies and the governments of 15 countries. *JAMA*  
380 *Netw. Open* **3**, e2018033 (2020).  
381 21. S Khan, A Asif, AE Jaffery, Language in a time of COVID-19: Literacy bias ethnic minorities

382 face during COVID-19 from online information in the UK. *J. Racial Ethn. Heal. Disparities* **8**,  
383 1242–1248 (2021).  
384 22. CM Chen, Public health messages about COVID-19 prevention in multilingual Taiwan. *Multilingua*  
385 **39**, 597–606 (2020).  
386 23. ST Lim, M Kelly, S Johnston, Re: ‘Readability of online patient education material for the  
387 novel coronavirus disease (COVID-19): a cross-sectional health literacy study’. *Public Heal.*  
388 **190**, 145–146 (2021).  
389 24. J Wernimont, Media infrastructure is and has always been a matter of life and death. *Items*  
390 **2021** (2021).  
391 25. K Keith, HHS strips gender identity, sex stereotyping, language access protections from ACA  
392 anti-discrimination rule. *Heal. Aff. Blog* **2020** (2020).  
393 26. MW Hooper, AM Nápoles, EJ Pérez-Stable, COVID-19 and racial/ethnic disparities. *J. Am.*  
394 *Med. Assoc.* **323**, 2466–2467 (2020).  
395 27. M Jacobson, TY Chang, M Shah, R Pramanik, SB Shah, Racial and ethnic disparities in  
396 SARS-CoV-2 testing and COVID-19 outcomes in Medicaid managed care cohort. *Am. J.*  
397 *Prev. Medicine* **000**, 1–8 (2021).  
398 28. X Chen, S Acosta, AE Barry, Evaluating the accuracy of Google translate for diabetes educa-  
399 tion material. *JMIR Diabetes* **1**, e5848 (2016).  
400 29. S Badarudeen, S Sabharwal, Assessing readability of patient education materials: current  
401 role in orthopaedics. *Clin. Orthop. Relat. Res.* **468**, 2572–2580 (2010).  
402 30. SA Novin, EH Huh, MG Bange, FK Hui, PH Yi, Readability of Spanish-language patient  
403 education materials from RadiologyInfo.org. *J. Am. Coll. Radiol.* **16**, 1108–1113 (2019).  
404 31. Centers for Disease Control and Prevention, Simply Put: A guide for creating easy-to-  
405 understand materials ([https://www.cdc.gov/healthliteracy/pdf/simply\\_put.pdf](https://www.cdc.gov/healthliteracy/pdf/simply_put.pdf)) (2009).  
406 32. J Redish, Readability formulas have even more limitations than Klare discusses. *ACM J.*  
407 *Comput. Documentation* **24**, 132–137 (2000).  
408 33. E Pittler, A Nenkova, Revisiting readability: A unified framework for predicting text quality  
409 in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*,  
410 EMNLP '08. (Association for Computational Linguistics, USA), p. 186–195 (2008).  
411 34. M Zamanian, P Heydari, Readability of texts: State of the art. *Theory & Pract. Lang. Stud.* **2**  
412 (2012).  
413 35. SA Crossley, S Skalicky, M Dascalu, Moving beyond classic readability formulas: new meth-  
414 ods and new models. *J. Res. Read.* **42**, 541–561 (2019).  
415 36. T François, C Fairon, An “AI readability” formula for French as a foreign language in *Proceed-*  
416 *ings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and*  
417 *Computational Natural Language Learning*, EMNLP-CoNLL '12. (Association for Computa-  
418 tional Linguistics, USA), p. 466–477 (2012).  
419 37. K Rayner, TJ Slattery, D Drieghe, SP Liversedge, Eye movements and word skipping during  
420 reading: effects of word length and predictability. *J. Exp. Psychol. Hum. Percept. Perform.* **37**,  
421 514 (2011).  
422 38. BB Kadayat, E Eika, Impact of sentence length on the readability of web for screen reader  
423 users in *International Conference on Human-Computer Interaction*. (Springer), pp. 261–271  
424 (2020).  
425 39. I Feinberg, Building a culture of health literacy during COVID-19. *New Horizons Adult Educ.*  
426 *Hum. Resour. Dev.* **33**, 60–64 (2021).  
427 40. J Howe, C Young, C Parau, G Trafton, R Ratwani, Accessibility and usability of state health  
428 department COVID-19 vaccine websites: A qualitative study. *JAMA Netw. Open* **4**, e2114861  
429 (2021).  
430 41. J Zarocostas, How to fight an infodemic. *Lancet* **385**, P676 (2020).  
431 42. G Pennycook, J McPhetres, Y Zhang, J Lu, D Rand, Fighting COVID-19 misinformation on  
432 social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychol.*  
433 *Sci.* **31**, 770–780 (2020).  
434 43. D Scales, J Gorman, K Jamieson, The Covid-19 infodemic—applying the epidemiologic  
435 model to counter misinformation. *New Engl. J. Medicine* **385**, 678–681 (2021).  
436 44. The Community Health Literacy Project (<https://communityhealthliteracyproject.org/>) (2020).  
437 45. Coronavirus Adult Literacy Resources ([https://education.gsu.edu/research-outreach/alrc/](https://education.gsu.edu/research-outreach/alrc/adult-literacy-coronavirus-resource-links/)  
438 [adult-literacy-coronavirus-resource-links/](https://education.gsu.edu/research-outreach/alrc/adult-literacy-coronavirus-resource-links/)) (2020).  
439 46. MD Hoffman, A Gelman, The No-U-Turn sampler: adaptively setting path lengths in Hamil-  
440 tonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).  
441 47. PC Bürkner, brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.*  
442 **80**, 1–28 (2017).  
443 48. J Piironen, A Vehtari, Comparison of Bayesian predictive methods for model selection. *Stat.*  
444 *Comput.* **27**, 711–735 (2017).  
445 49. HL Turner, J van Etten, D Firth, I Kosmidis, Modelling rankings in R: the PlackettLuce package.  
446 *Comput. Stat.* **35**, 1027–1057 (2020).  
447