## Supplemental Material S1.

### 1. Travis et al. (1937)'s results as published

TABLE I

| | English Speaking Only | Bi-lingual* | Total White | Colored | Three Lan-guages† | Foreign Only |
|---|---|---|---|---|---|---|
| Boys.......... | 2.72 | 3.70 | 3.39 | 5.77 | 5.88 | 10.00 |
| Girls.......... | .83 | 1.90 | 1.45 | 1.98 | 0.00 | 5.88 |
| Total......... | 1.80 | 2.80 | 2.43 | 3.76 | 2.38 | 7.41 |

*Figure 1: Table 1 from Travis et al. (1937).* The heading "English Speaking Only" referred to monolingual English-speakers; "Bilingual" applied to children who spoke English and one other language; "Colored" referred to monolingual English speakers who were African American. The group labeled "Total White" combines the English Speaking Only and Bilingual groups. No information is given as to which children were considered to be "white". Legal opinion on that point, e.g. regarding individuals of Mexican or Chinese descent, was in flux at the time. The heading "Three Languages" applied to children who spoke English and two or more other languages (the maximum was five languages including English). The label "Foreign Only" referred to children who did not speak English, regardless of the number of languages they spoke.

## 2. Travis et al. (1937)'s results, reconstructed counts

Table 1: *Percentages and absolute numbers of participants. Figures in bold face were stated in Travis et al. (1937). Figures in plain font were calculated by the method illustrated in Figure 1 in the text and explained below.*

| | | | English Only | Bilingual | Three languages | Foreign only | Total |
|---|---|---|---|---|---|---|---|
| Percentage | Stuttering | Boys | **2.72** | **3.7** | **5.88** | **10** | |
| | | Girls | **0.83** | **1.9** | **0** | **5.88** | |
| | | Total | **1.8** | **2.8** | **2.38** | **7.41** | **2.61** |
| | Fluent | | | **97.20** | | | |
| Counts | Stuttering | Boys | 33 | 43 | 2 | 1 | |
| | | Girls | 10 | 22 | 0 | 1 | |
| | | Total | 43 (or 57) | 65 | 2 | 2 | **126** (or 112) |
| | Fluent | Boys | 1180 | 1119 | 35 | 9 | 2377 |
| | | Girls | 1195 | 1136 | 52 | 16 | 2405 |
| | | Total | 2356 (or 2361) | 2257 | 87 | 25 | 4701 |
| | Combined | Boys | 1213 | 1162 | **37** | **10** | **2405** |
| | | Girls | 1205 | 1158 | **52** | **17** | **2422** |
| | | Total | **2399** (or 1213 + 1205 = 2361 + 57 = 2418) | **2322** (or 2320) | 89 | **27** | **4827** |

### 3. Checking the internal consistency of Travis et al.'s report.

In what follows, the left-hand columns summarize information explicitly stated in the table of results or in the text in Travis et al. (1937). The right-hand columns show calculations based on that information, as well as conclusions that can be drawn from them. Inferences are marked with arrows, to distinguish them from facts of arithmetic. Statements pointing out discrepancies or contradictions are in boldface.

[1] Bilingual children:

| | |
|---|---|
| 1.80% of the 2,399 'English-speaking only' children stuttered. | 1.80% of 2,399 = 43.182, rounding to 43.<br>43 out of 2,399 = 1.792414 %<br>→ There were 43 'English-speaking only' children labeled as stuttering.<br>**43 out of 2399 rounds to 1.79 %, not 1.80 %.** |
| 2.80% of the 2322 bilingual children stuttered. | 2.80% of 2,322 = 65.016, rounding to 65.<br>65 out of 2,322 = 2.799311 %,<br>rounding to 2.80 %<br>→ There were 65 bilingual children (2.80% of 2,322) labeled as stuttering. |

[2] Foreign-only

| | |
|---|---|
| There were 27 children who did not speak English ('Foreign-only'): 10 boys and 17 girls. Of the 10 foreign-only boys, 10% stuttered.<br>Of the 17 foreign-only girls, 5.88% stuttered. | 10% of 10 = 1.<br>5.88% of 17 = 0.9996, rounding to 1.<br>1 out of 17 = 5.882353 %, rounding to 5.88%.<br>→ There were 2 children (1 boy and 1 girl) in the Foreign-only group. |
| 7.41 % of the 27 Foreign-only children stuttered. | 7.41 % of 27 = 2.0007, rounding to 2.<br>2 out of 27 equals 7.407407, rounding to 7.41<br>→ Two Foreign-only children stuttered. |

[3] Trilingual children

| | |
|---|---|
| There were 37 trilingual boys and 52 trilingual girls. | 37 + 52 = 89.<br>→ There were 89 trilingual children who stuttered. |
| Of the 37 trilingual boys, 5.88 % stuttered. | 5.88% of 37 = 2.1756, rounding to 2.<br>→ There were two trilingual boys who stuttered.<br>**2 out of 37 is 5.405 % , not 5.88 %.**<br>**2 is 5.88% out of 34, not 37.** |

| | |
|---|---|
| | **If the stated percentage (5.88%) is accurate, then three fluent trilingual boys were excluded before calculating the stuttering prevalence in this group.** |
| Of the 52 trilingual girls, 0% stuttered. | None of the trilingual girls stuttered. |
| Of all trilingual children, 2.38 % stuttered. | **2 out of 89 is 2.247%, not 2.38%.**<br>**2 is 2.38% of 84, not 89.**<br>**→ If the stated percentage (2.38%) is correct, then 5 fluent children were excluded before the percentage rate for the trilingual group was calculated, inflating the prevalence estimate.** |

[4] Total sample size

| | |
|---|---|
| The total sample size was 4827. | |
| There were 2399 English-only children, 2322 bilingual children, 89 trilingual ones, and 27 foreign-only ones. | 2399 + 2322 + 89 + 27 = **4837.**<br>**The sum of the subgroups is larger than the stated total sample size.** |

[5] Total number of children who stuttered

| | |
|---|---|
| A total of 126 children stuttered. | 126 out of 4827 equals 2.61%. |
| The overall stuttering prevalence in the sample of 4827 was 2.61 %. | **The counts of English-only (43), Bilingual (65), Trilingual (2), and Foreign-only (2) children who stuttered sum to 112, not 126.** |

[6] Bilingual boys vs. girls

| | |
|---|---|
| The group of bilingual children who stuttered was comprised of 66.2% boys and 33.8% girls. | 66.2% and 33.8% of 65 round to 43 and 22, respectively.<br>→ The group of 65 bilingual children contained 43 boys and 22 girls. |
| Of the bilingual boys, 3.70% stuttered.<br>Of the bilingual girls, 1.90% stuttered.<br>There were 2322 bilingual children in the sample. | If the 43 boys who stuttered represent 3.70% of the total number of bilingual boys, and 22 represents 1.90% of the bilingual girls, then there were 1162 bilingual boys and 1158 bilingual girls, for a total of 2320 bilingual children.<br>→ There is a slight difference between the sample size inferred from the percentages (2320) vs. the stated sample size (2322), possibly due to rounding. |

[7] Monolingual English-speaking boys vs. girls

| | |
|---|---|
| The group of English-only children who stuttered was comprised of 77.4% boys and 22.6% girls | There were 43 children in that group. 77.4% of 43 is 33.<br>→ The group of 43 English-only children contained 33 boys and 10 girls. |
| 2.72% of the English-only boys stuttered. | → If 33 represented 2.72%, then the total number of English-only boys (stuttering and non-stuttering combined) was 1213. |
| 0.83% of the English-only girls stuttered. | → If 10 girls represented 0.83%, then the total number of English-only girls was 1205. |
| | **The sum of 1213 and 1205 girls gives us a total group size of 2418 English-only children, 19 more than the reported total of 2399.**<br>**Taken together with item [5] above, this suggests that 14 children who stuttered were excluded from the prevalence calculation for the monolingual group.** |

**4. Robustness of statistical significance to small changes in stuttering classification**

Travis et al. (1937) do not mention what statistical test was used, noting only that “[s]tatistical analysis reveals that there are 98 chances in 100 that [the difference between 1.8 % vs. 2.8 %] is a true one. If we group together all cases speaking English plus one or more foreign languages, we obtain a per cent of stuttering for this combined group of 2.82. Comparing this group with the group speaking English only, we find that there are 99 chances in 100 that there is a true difference" Travis et al. (1937, p. 187). Two possibilities are Pearson’s Chi-square test (Pearson, 1900) and Fisher’s exact test (Fisher, 1922), both of which are both mentioned in contemporary publications by Travis and colleagues. The independence assumption underlying both tests is violated -- age and sex both affect stuttering prevalence and hence skew the distribution – but such violations are widespread in the literature. Applying Pearson’s chi-square test to the reported counts, i.e. 43 stuttering monolingual and 65 stuttering bilingual children, vs. 2257 and 2356 non-stuttering ones per group, yields a $\chi^2(1)$ value of 5.35, $p = .021$. By Fisher’s test, the $p$-value is 0.025. The $p$-value associated with $\chi^2$ comes closer to the reported result (“98 chances in 100” that the effect is “a true one”) than does Fisher’s exact test, so the chi-square test was used in the simulation. (Yates’ continuity correction, introduced in Yates, 1934, was cited in publications aimed at audiences of statisticians, rather than clinical researchers in 1937. In one applies the correction, the $p$-value associated with $\chi^2$ is .027, leaving the statistical significance intact, but differing from Travis et al.’s figure.)

*Table 2: Changes in Pearson's Chi-square and associated p-values in response to incremental changes in the number of children in each fluency group in Travis et al. (1937).*

| Monolingual | | | | Bilingual | | | |
|---|---|---|---|---|---|---|---|
| n (Stuttering) | n (Fluent) | $\chi 2$ | p($\chi^2$) | n (Stuttering) | n (Fluent) | $\chi 2$ | p($\chi^2$) |
| 43 | 2356 | 5.35 | 0.021 | 65 | 2257 | 5.35 | 0.021 |
| 44 | 2355 | 4.87 | 0.027 | 64 | 2258 | 4.95 | 0.026 |
| 45 | 2354 | 4.42 | 0.035 | 63 | 2259 | 4.56 | 0.033 |
| 46 | 2353 | 4.00 | 0.046 | 62 | 2260 | 4.18 | 0.041 |
| 47 | 2352 | 3.60 | 0.058 | 61 | 2261 | 3.82 | 0.051 |
| 48 | 2351 | 3.22 | 0.073 | 60 | 2262 | 3.46 | 0.063 |
| 49 | 2350 | 2.87 | 0.090 | 59 | 2263 | 3.13 | 0.077 |
| 50 | 2349 | 2.54 | 0.111 | 58 | 2264 | 2.80 | 0.094 |
| 51 | 2348 | 2.23 | 0.135 | 57 | 2265 | 2.50 | 0.114 |
| 52 | 2347 | 1.95 | 0.163 | 56 | 2266 | 2.20 | 0.138 |
| 53 | 2346 | 1.69 | 0.194 | 55 | 2267 | 1.93 | 0.165 |

## References

Fisher, R. (1922). On the interpretation of $\chi^2$ from contingency tables, and the calculation of p. *Journal of the Royal Statistical Society*, *85*(1), 87–94.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157–175.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Yates, F. (1934). Contingency tables involving small numbers and the $\chi^2$ test. *Supplement to the Journal of the Royal Statistical Society*, *1*(2), 217–235.