

Gender API App Validation

2/27/2025

Jim Hagberg

Summary: **The error rate for invalid sex identified via the gender-api app validated versus direct internet information for that specific individual using our 80% confidence criterion was only 7 first names out of 488 or 1.4%.**

Three methods were used to identify the sex of authors across the three journals on which we have generated data bases (Journal of Applied Physiology, Medicine and Science in Sports and Medicine, International Journal of Sports Medicine) relative to the sex of the research participants and manuscript authors. The first was a sex-oriented first name (eg – James, Robert, Sarah, Jane, etc). The second was if the author of this paper knew the individual to be a male or female individual. The third method was based on a commercial internet application (gender-api.com) that uses AI to identify the sex of an individual based on their first name. The company indicates that their app covers over 6 million first names across 190 countries.

Two reviewers of our initial HERstory submission to the Journal of Applied Physiology quantifying the inclusion of men and women as research participants and manuscript authors raised an issue about the validity of using a commercial internet application (gender-api.com) to identify the sex of authors of the manuscripts. We queried the company numerous times, about any internal validity data they could share. And not surprisingly, they really provided no useful information relative to this point.

Thus, I undertook a completely independent assessment of the validity of this app. The first names identified for inclusion in this assessment were those from the Journal of Applied Physiology published papers from 2024 and 2025. This time frame was selected to not overlap with the data used from this journal in our database because data collection was stopped at the end of 2023. Five hundred first names were collected using the same criteria as for the original database – not known by myself and not an obviously sex-oriented first name.

The gender-api app returns two pieces of information for each search – first is the % confidence that the app has that their prediction is correct. In our papers we use a criterion that the app must be $\geq 80\%$ confident otherwise that individual author was coded as unknown Gender. The second is the number of samples assessed to determine their prediction and confidence levels. For our validation we recorded both of these pieces of data.

The names selected for inclusion in this analysis were generally not very common first name, or if they were common, they were common in another country. To illustrate this point, if we analyze the top 10 US male and female first names, the gender-api app is 97-99% confident in all of their assessments and the average number of samples viewed to obtain this confidence level was 110,000. For the names included in this validation, the number of samples averaged 5350 (hence – these names were dramatically less common)!

For validation of the sex of the individual, I used their full name and institution to search the web for a picture or some type of text directly related to them using sex-oriented pronouns (his, her). There were individuals for whom I could not obtain such information and they were excluded from the database as they would not have information for the criterion outcome.

I ended up having data relative to 499 first names.

- a) Eleven of these first names (or 2.2%) generated no information when used in the gender-api app for that first name.
- b) 435 of these 488 first names (or 89.1%) had the correct sex identified with >80% confidence
- c) 392 of these 488 first names (or 80.3%) had the correct sex identified with >90% confidence
- d) 359 of these 488 first names (or 73.5%) had the correct sex identified with >95% confidence
- e) 282 of these 488 first names (or 57.8%) had the correct sex identified with >98% confidence

The percent confidence over the 488 identified names averaged $94 \pm 13\%$ and the average number of samples reviewed was $5351 \pm 10,375$.

Overall, 22 first names of the 488 (4.5%) exhibited differences between the gender-api app identified gender and the gender validated via internet on-line information. However, the error rate was reduced substantially when differences identified based on our 80% confidence criterion, as now only 7 first names (1.4%) were misclassified in terms of gender by the gender-api app.