

Methodology

To conduct this analysis of tweets mentioning U.S. Olympic athletes at the time of the 2020 Olympic Games, researchers from Pew Research Center collected every athlete Twitter handle listed on the official [Team USA page](#) as of approximately July 26, 2021. (The Games had been postponed due to COVID-19 and were held July 23 to Aug. 8, 2021.) This analysis includes the 598 Team USA athletes who were listed on the site, 438 of whom had their handle listed at the time. This collection includes U.S. Olympic team alternates but not contestants in the Paralympic Games (which occurred later in the month).

Researchers collected all tweets posted from July 21 through Aug. 9, 2021, that mentioned or replied to any of these 438 twitter handles using the [Gnip Historical PowerTrack API](#) (Gnip API), a searchable archive of all publicly available tweets, in two separate GNIP queries: July 21-28 and July 29-Aug. 9. This analysis only includes direct mentions or replies of U.S. Olympians via their Twitter handle (such as @kdtrey5 or @simone_biles) and not more generic references (such as “Kevin Durant,” “KD” or “Biles”).

To group the collected Twitter handles by sport, researchers relied on the TeamUSA URL structure, which uses subdirectories to separate athletes into individual sports. For example, the athlete profile page for Bam Adebayo indicates that he is a member of the USA Basketball team:

<https://www.teamusa.org/usa-basketball/athletes/Bam-Adebayo>

After collecting the full list of athlete profile URLs, researchers parsed each URL to extract the athlete’s sport and categorized it based on that designation. For consistency, the Center defaulted to this structure for classifying athletes by team, even though the TeamUSA URL structure does not include subdirectories for all sports (for instance, athletes who competed in the 3x3 basketball events are listed under the “usa-basketball” subdirectory). Note, at the time when researchers scraped athlete profile handles, the social media handle listed for Jayson Tatum was incorrect, listing Khris Middleton’s handle instead. Researchers were able to update this information when running data collection on the second GNIP query, but tweets that mention Jayson Tatum from July 21-28 are not included in this analysis.

In addition, at the time of data collection, all U.S. Olympians competing in table tennis had @usatabletennis as their twitter handle as opposed to an individual account. For this reason, tweets that mentioned @usatabletennis are not included in this analysis. Also, researchers discovered that at the time of data collection of twitter handles from the Team USA website, the twitter handles for Olympians Jake Mitchell (@jake_mitch1) and Donald Scott (@DonaldScott_II)

were mistakenly listed on Team USA’s website using periods as opposed to underscores in their respective handles (e.g., “@jake.mitch1” vs. “@jake_mitch1”). Although researchers updated this information for the analysis, the resulting tweets that mention either of those accounts in the dataset are likely only those that included their correct handle names along with other Olympic handles. Tweets that only included their handles were not captured in the GNIP queries.

Distinctive terms

Distinctive keywords and phrases used in tweets that mentioned the handles of U.S. Olympians were identified using [pointwise mutual information](#). Researchers then calculated the proportion of party members who mentioned each distinct term (phrase). Terms mentioned in fewer than 100 tweets are excluded. Researchers then used the proportions to calculate a ratio of differences in mentions between tweets that mention @simone_biles and tweets that did not for each term. Researchers removed all handles from this analysis; among the tweets that mentioned @simone_biles, researchers removed “simone” and “biles” from the distinctive-terms analysis. The most distinctive keywords were defined as those terms with the largest ratio difference between the two categories.

As a final step for both keywords analyses, researchers consolidated phrases: removing those that had a word in common with any other phrase that was associated with a larger difference and those that were part of a general speech pattern with no important contextual meaning. Terms have been edited slightly in some cases for clarity. Words that appeared in retweets are included in this analysis, even if the account that retweeted them did not create the original tweet.