# Supplementary information: Notes on statistical analysis outcomes

## Comparison of condition and person-first language use

### Supplementary Note 1:

A Welch’s Two Sample t-test of the difference between the frequency of person-first and condition-first language (mean frequency of person-first = 8.23 words per million, mean frequency of condition-first language = 284.74 words per million) in the 10000 instances across the randomly sampled 1000-article subcorpora suggested that the effect is negative, statistically significant, and large (difference = -276.51, 95% CI [-277.13, -275.89], t(10666.13) = -871.18, p < .001; Cohen’s d = -12.32, 95% CI [-12.49, -11.03])

### Supplementary Note 2:

A Welch’s Two Sample t-test of the difference between the number of articles (per 1000) which use person-first and condition-first language (mean number of articles using person-first language across all subcorpora = 4.03, mean number of articles using condition-first language across all subcorpora = 122.54) suggests that the effect is negative, statistically significant, and large (difference = -118.51, 95% CI [-118.71, -118.31], t(10748.87) = -1149.40, p < .001; Cohen’s d = -16.26, 95% CI [-16.47, -16.04])

### Supplementary Note 3:

Goodness of fit tests normalised to total word counts (across all articles from broadsheets and tabloids, for example) or total article counts revealed the following:

- The total number of uses of condition-first language we observe is higher in tabloids and lower in broadsheets than we would expect based on the word count in these subcorpora (p < 0.001).

- The number of articles with condition-first language we observe is also higher in tabloids and lower in broadsheets than we would expect based on the total article count in these subcorpora (p < 0.001).

- The total number of uses of person-first language we observe is somewhat higher in tabloids and lower in broadsheets than we would expect based on the word count in these subcorpora, but this result is not strongly significant (p < 0.05).

- The number of articles with person-first language we observe is, in contrast, lower in tabloids and higher in broadsheets than we would expect based on the total article count in these subcorpora (p < 0.002).

The Pearson’s Chi-squared test with Yates’ continuity correction contrasting articles from tabloids and broadsheets that use only condition-first vs only person-first language indicate a significant link (X-squared = 4.8274, p-value = 0.02801) between type of publication and number of articles using a specific language type. The effect size is quite small (<0.2), indicating that while the result is statistically significant, the fields are weakly associated. Acknowledging the propensity of the Yates correction to overcorrect, we have also provided uncorrected Chi-square results, which result in less stringent p-values, at [LINK].

The Pearson’s Chi-squared test with Yates’ continuity correction contrasting articles from left- and right-leaning publications that use only condition-first vs only person-first language indicate a significant link (X-squared = 4.6405, p-value = 0.03123) between type of publication and number of articles using a specific language type. The effect size is, however, also quite small (<0.2), indicating that while the result is statistically significant, the fields are weakly associated.

### Supplementary Note 4:

A Welch Two Sample t-test testing the difference between the normalised frequency per 1000 words of condition- and person-first language in articles that use either one or both language types (mean of condition-first 4.34 words per 1000, mean of person-first – 2.67 words per 1000) suggests that the effect is positive, statistically significant, and small (difference = 1.66, 95% CI [1.16, 2.17], t(131.59) = 6.49, p < .001; Cohen’s d = 0.44, 95% CI [0.30, 0.58]).

### Supplementary Note 5

Consistent with the above binary comparisons of tabloids and broadsheets, simple linear modelling revealed a difference among sources: relative to *The Advertiser*, a tabloid, all considered broadsheets (the *Age, Australian, Canberra Times* and *Sydney Morning Herald*) had a lower frequency of condition-first language, as did the *Courier Mail*, a tabloid; no difference across the time period could be detected.

More specifically, a linear model of the form log(frequency) ~ scaled\_year + source was fitted and estimated using ML and found to have the lowest AIC among models (see GitHub for additional details of all models considered).

The model’s intercept, corresponding to scaled\_year = 0 and source = Advertiser, is at 1.28 (95% CI [1.21, 1.36], t(3031) = 32.98, p < .001). Within this model:

The effect of scaled year is statistically non-significant and negative (beta = -7.05e-03, 95% CI [-0.02, 2.13e-03], t(3031) = -1.50, p = 0.132; Std. beta = -9.82e-03, 95% CI [-0.02, 3.54e-03])

The effect of source [Age] is statistically significant and negative (beta = -0.45, 95% CI [-0.57, -0.33], t(3031) = -7.43, p < .001; Std. beta = -0.19, 95% CI [-0.25, -0.14])

The effect of source [Australian] is statistically significant and negative (beta = -0.65, 95% CI [-0.79, -0.51], t(3031) = -9.12, p < .001; Std. beta = -0.27, 95% CI [-0.33, -0.21])

The effect of source [CanTimes] is statistically significant and negative (beta = -0.37, 95% CI [-0.50, -0.23], t(3031) = -5.33, p < .001; Std. beta = -0.18, 95% CI [-0.24, -0.12])

The effect of source [CourierMail] is statistically significant and negative (beta = -0.14, 95% CI [-0.25, -0.03], t(3031) = -2.48, p = 0.013; Std. beta = -0.07, 95% CI [-0.12, -0.02])

The effect of source [HeraldSun] is statistically non-significant and negative (beta = -0.06, 95% CI [-0.17, 0.04], t(3031) = -1.18, p = 0.240; Std. beta = -0.05, 95% CI [-0.09, -2.66e-04])

The effect of source [HobMercury] is statistically non-significant and positive (beta = 0.13, 95% CI [-0.02, 0.27], t(3031) = 1.68, p = 0.092; Std. beta = 0.05, 95% CI [-0.01, 0.12])

The effect of source [NorthernT] is statistically non-significant and positive (beta = 0.16, 95% CI [-0.02, 0.35], t(3031) = 1.72, p = 0.086; Std. beta = 0.04, 95% CI [-0.04, 0.12])

The effect of source [SydHerald] is statistically significant and negative (beta = -0.53, 95% CI [-0.64, -0.42], t(3031) = -9.42, p < .001; Std. beta = -0.23, 95% CI [-0.28, -0.19])

The effect of source [WestAus] is statistically non-significant and positive (beta = 0.02, 95% CI [-0.12, 0.15], t(3031) = 0.24, p = 0.808; Std. beta = -0.02, 95% CI [-0.08, 0.04])

Similar results were observed when using more complex mixed effects models: a difference among sources and no effect of year.

## Comparison of use of the adjectives OBESE, OVERWEIGHT and FAT

Summary comparison

Both OBESE (Supplementary note 6) and OVERWEIGHT (Supplementary note 7) were used more frequently in tabloids than in broadsheets, with an average of 5.59 and 5.73 instances per 1000 words in tabloids vs 3.16 and 3.55 instances per 1000 words in broadsheets, respectively; these results were also observed when using linear modelling to consider sources individually. These results are attributable in part to tabloids having shorter article lengths in the corpus (Supplementary note 6), irrespective of adjective use, so a similar number of *instances* of use of the adjectives of interest will result in a lower *frequency* in broadsheets than in tabloids. A very subtle decrease in the use of the adjective OBESE was observed over time when using fixed effects linear modelling (beta = -0.01, 95% CI [-0.02, -0.005]), but this was not replicated when using a random effects linear model; a similar very subtle decrease was observed in the use of OVERWEIGHT over time (beta = -0.01, 95% CI [-0.02, -0.0066]), supported by both fixed and random effects modelling (Supplementary note 7).

Tabloids used FAT as weight-emphasising adjectival label for a person more frequently than broadsheets, as described using both a pairwise comparison between the two classes and considering each source individually, and this effect was, like the case of OVERWEIGHT and OBESE, partially attributable to the longer article lengths in broadsheets than tabloids (Supplementary note 8). No differences by year were observed. Details are provided below.

### Supplementary note 6

Goodness of fit tests normalised to total word counts (across all articles from broadsheets and tabloids, for example) or total article counts revealed the following:

- The total number of uses of "obese" we observe is higher in tabloids, and lower in broadsheets than we would expect based on the word count in these subcorpora (p < 0.001).

- The number of articles with use of "obese" we observe is not different between tabloids and broadsheets.

The mean frequency per 1000 words of articles where OBESE is used is higher in tabloids than in broadsheets. Based on a Welch’s Two Sample, the mean frequency per 1000 words of articles where OBESE is used was 3.16 in broadsheets and 5.59 in tabloids, which suggests that the effect is negative, statistically significant, and small (difference = -2.43 words per 1000, 95% CI [-2.63, -2.23], t(10062.88) = -24.01, p < .001; Cohen’s d = -0.46, 95% CI [-0.50, -0.42].

This is in part related to the fact that articles that use OBESE are longer in broadsheets than in tabloids. Across articles that used the word OBESE, using a Welch Two Sample t-test to compare the mean word count of articles from broadsheets and tabloids that used OBESE (mean of broadsheet = 778.84, mean of tabloid = 485.42) suggested there was a positive, statistically significant effect, with a medium size (difference = 293.42, 95% CI [273.82, 313.02], t(7348.69) = 29.34, p < .001; Cohen’s d = 0.60, 95% CI [0.56, 0.65]).

We next compared the overall length of all articles in tabloids and broadsheets in the corpus and found that articles in broadsheets were *generally* longer than in tabloids, irrespective of whether any of the adjectives of interest were used. A Welch Two Sample t-test testing the difference between wordcounts in broadsheet and tabloid articles (mean of broadsheets = 810.15 words, mean of tabloids = 502.44 words) suggested that the effect is positive, statistically significant, and medium (difference = 307.71, 95% CI [291.01, 324.42], t(16482.54) = 36.11, p < .001; Cohen’s d = 0.56, 95% CI [0.53, 0.59]).

In order to investigate in more depth that the difference in frequency of the word OBESE per 1000 words in tabloids and broadsheets is not fully an artifact of this difference in length, we used resampling without replacement (selecting 1000 articles 10000 times) to compare the raw counts of the word OBESE in articles from tabloids and broadsheets, and found that the counts themselves were higher in broadsheets, which had a bimodal distribution of instances of use of the word OBESE – so much of the difference in *frequency* is observed due to this difference in article lengths. We have provided a discussion of this phenomenon at [GITHUB].

The Welch Two Sample t-test testing the difference between resampled broadsheet counts and tabloid counts (mean of broadsheets = 1.71, mean of tabloids = 1.60) suggested that the effect is positive, statistically significant, and large (difference = 0.108519, 95% CI [0.1073502 , 0.1096892 t(18721.64) = 182.01, p < .001; Cohen’s d = 2.57, 95% CI [2.83, 2.83])

Linear modelling by source and year supported the difference between frequency of OBESE in tabloids and broadsheets, explaining a statistically significant and weak proportion of variance (R2 = 0.10, F(10, 10245) = 109.54, p < .001, adj. R2 = 0.10), with broadsheets considered (the *Age, Australian, Canberra Times* and *Sydney Morning Herald*) having lower frequency of use of the word OBESE than the *Advertiser* (a tabloid), while other tabloids (the *Mercury* and *NT News*) had a higher frequency of use than the *Advertiser*. This fixed effects modelling also suggested a slight decrease in the use of the word OBESE with time (beta = -0.01, 95% CI [-0.02, -5.16e-03], t(10245) = -3.96, p < .001; Std. beta = -0.02, 95% CI [-0.02, -8.05e-03]).

Mixed effects modelling supported the observation of higher frequency of counts in tabloids, but the decrease by year was not found to be significant.

### Supplementary note 7

Goodness of fit tests normalised to total word counts (across all articles from broadsheets and tabloids, for example) or total article counts revealed the following:

- The total number of uses of "overweight" we observe is higher in tabloids and lower in broadsheets than we would expect based on the word count in these subcorpora (p < 0.001).

- The number of articles with use of "overweight" we observe is higher in tabloids and lower in broadsheets than we would expect based on the word count in these subcorpora (p < 0.001).

The word OVERWEIGHT was used more frequently in tabloids than broadsheets. A Welch’s Two Sample t-test of the difference between the frequency per 1000 words in broadsheets and tabloids across articles that use the word OVERWEIGHT (mean of broadsheets = 3.55, mean of tabloids = 5.73 words per 1000) suggests that the effect is negative, statistically significant, and small (difference per 1000 words = -2.18 words, 95% CI [-2.42, -1.95], t(6712.33) = -18.20, p < .001; Cohen’s d = -0.43, 95% CI [-0.48, -0.38]).

Similar to the case of OBESE and the overall corpus statistics (Supplementary note 6), the length of articles that use the word OVERWEIGHT is longer in broadsheets than tabloids. The Welch Two Sample t-test testing the difference between the wordcount of articles that use the word OVERWEIGHT from broadsheets and tabloids (mean of broadsheets = 732.66, mean of tabloids = 484.51) suggests that the difference is positive, statistically significant, and medium (difference = 248.15, 95% CI [226.60, 269.71], t(4626.50) = 22.57, p < .001; Cohen’s d = 0.58, 95% CI [0.53, 0.63]).

In order to investigate in more depth that the difference in frequency of the word OVERWEIGHT per 1000 words in tabloids and broadsheets is not fully an artifact of this difference in length, similar to the case of OBESE, we used resampling without replacement (selecting 1000 articles 10000 times) to compare the raw counts of the word OVERWEIGHT in articles from tabloids and broadsheets, and found that the counts themselves were higher in broadsheets, which had a bimodal distribution of instances of use of the word OVERWEIGHT – so much of the difference in *frequency* is observed due to this difference in article lengths. The Welch Two Sample t-test testing the difference between resampled broadsheet counts and tabloid counts (mean of broadsheets = 1.95, mean of tabloids = 1.80) suggests that the effect is positive, statistically significant, and large (difference = 0.15, 95% CI [0.15, 0.15], t(19959.90) = 242.90, p < .001; Cohen’s d = 3.44, 95% CI [3.78, 3.78])

Similar to the analysis of OBESE, linear modelling suggested that the word OVERWEIGHT was used less in broadsheet sources (the *Age, Australian, Canberra Times* and *Sydney Morning Herald*) than in the *Advertiser*, and more in the *Mercury* and *NT News* (both tabloids); a small effect of year was also observed. More specifically, we fitted a linear model (estimated using OLS) to predict frequency with source and scaled\_year (formula: log(frequency) ~ source + scaled\_year). The model explained a statistically significant and weak proportion of variance (R2 = 0.08, F(10, 6710) = 58.73, p < .001, adj. R2 = 0.08). The model’s intercept, corresponding to source = Advertiser and scaled\_year = 0, is at 1.31 (95% CI [1.26, 1.37], t(6710) = 48.02, p < .001). Within this model:

* The effect of source [Age] is statistically significant and negative (beta = -0.44, 95% CI [-0.53, -0.36], t(6710) = -10.23, p < .001; Std. beta = -0.18, 95% CI [-0.22, -0.14])
* The effect of source [Australian] is statistically significant and negative (beta = -0.52, 95% CI [-0.62, -0.43], t(6710) = -10.41, p < .001; Std. beta = -0.21, 95% CI [-0.25, -0.16])
* The effect of source [CanTimes] is statistically significant and negative (beta = -0.29, 95% CI [-0.37, -0.20], t(6710) = -6.52, p < .001; Std. beta = -0.13, 95% CI [-0.17, -0.09])
* The effect of source [CourierMail] is statistically non-significant and positive (beta = 0.04, 95% CI [-0.04, 0.11], t(6710) = 0.95, p = 0.340; Std. beta = 0.01, 95% CI [-0.02, 0.04])
* The effect of source [HeraldSun] is statistically non-significant and positive (beta = 0.03, 95% CI [-0.05, 0.10], t(6710) = 0.72, p = 0.472; Std. beta = 3.54e-03, 95% CI [-0.03, 0.04])
* The effect of source [HobMercury] is statistically significant and positive (beta = 0.17, 95% CI [0.07, 0.26], t(6710) = 3.34, p < .001; Std. beta = 0.06, 95% CI [0.02, 0.10])
* The effect of source [NorthernT] is statistically significant and positive (beta = 0.43, 95% CI [0.31, 0.56], t(6710) = 6.84, p < .001; Std. beta = 0.18, 95% CI [0.13, 0.24])
* The effect of source [SydHerald] is statistically significant and negative (beta = -0.40, 95% CI [-0.48, -0.33], t(6710) = -10.33, p < .001; Std. beta = -0.17, 95% CI [-0.20, -0.13])
* The effect of source [WestAus] is statistically non-significant and positive (beta = 0.03, 95% CI [-0.06, 0.11], t(6710) = 0.62, p = 0.537; Std. beta = -2.62e-03, 95% CI [-0.04, 0.04])
* The effect of scaled year is statistically significant and negative (beta = -0.01, 95% CI [-0.02, -6.59e-03], t(6710) = -4.13, p < .001; Std. beta = -0.02, 95% CI [-0.03, -0.01])

Unlike in the case of OBESE, mixed effects modelling supported both the differences by source type and the slight decrease by year (p < 0.05) (see GitHub).

### Supplementary note 8

Goodness of fit tests normalised to total word counts (across all articles from broadsheets and tabloids, for example) or total article counts revealed the following:

- The total number of uses of "fat" we observe is higher in tabloids and lower in broadsheets than we would expect based on the word count in these subcorpora (p < 0.001).

- The number of articles with use of "fat" we observe is not different between tabloids and broadsheets.

A Welch Two Sample t-test testing the difference between the frequency of the use of FAT as weight-emphasising adjectival label for a person revealed a difference in the frequency per 1000 words of this in broadsheets vs tabloids (mean of broadsheets = 2.62, mean of tabloids = 4.49) suggesting that the effect is negative, statistically significant, and small (difference = -1.87, 95% CI [-2.18, -1.56], t(2251.24) = -11.73, p < .001; Cohen’s d = -0.47, 95% CI [-0.55, -0.39]).

Similar to the case of OBESE, OVERWEIGHT and the overall corpus statistics (Supplementary note 6), the length of articles that use the word FAT in this way is longer in broadsheets than tabloids. The Welch Two Sample t-test testing the difference between wordcount of articles that use FAT in this way describes a difference between broadsheets and tabloids (mean word count of broadsheets = 922.40, mean of tabloids = 573.96), suggesting that the effect is positive, statistically significant, and medium (difference = 348.45, 95% CI [301.88, 395.01], t(1740.32) = 14.68, p < .001; Cohen's d = 0.70, 95% CI [0.61, 0.80])

Re-sampling of the counts without replacement revealed that, similar to OBESE and OVERWEIGHT raw counts were higher in broadsheets than in tabloids in articles that used FAT as a weight-emphasizing adjectival label for a person, in part due to more articles in broadsheets having 7+ uses vs only tabloids (23 vs 10 articles, respectively). The Welch Two Sample t-test testing the difference between broadsheet counts and tabloid counts (mean of broadsheets = 1.85, mean of tabloids = 1.48) suggests that the effect is positive, statistically significant, and large (difference = 0.37212, 95% CI [0.371, 0.374], t(17733.95) = 491.95, p < .001; Cohen’s d = 6.96, 95% CI [7.65, 7.65]).

Investigating the data by source and year using linear modelling revealed that these variables explained a small amount of variance in the data, with the word fat as weight-emphasising adjectival label for a person being used less frequently in individual broadsheets (*Age*, *Australian*, *Canberra Times* and *Sydney Morning Herald*) than in the *Advertiser* (a tabloid), while in the *NT News*, another tabloid, the word FAT was used somewhat more frequently than in the *Advertiser*. The word FAT was not used differently in the corpus across the time period.

These results were obtained by fitting a linear model (estimated using OLS) to predict frequency with source and scaled\_year (formula: log(frequency) ~ source + scaled\_year). The model explained a statistically significant and weak proportion of variance (R2 = 0.10, F(10, 2289) = 24.23, p < .001, adj. R2 = 0.09). The model’s intercept, corresponding to source = Advertiser and scaled\_year = 0, is at 1.02 (95% CI [0.92, 1.11], t(2289) = 21.55, p < .001). Within this model:

* The effect of source [Age] is statistically significant and negative (beta = -0.39, 95% CI [-0.53, -0.26], t(2289) = -5.62, p < .001; Std. beta = -0.18, 95% CI [-0.24, -0.12])
* The effect of source [Australian] is statistically significant and negative (beta = -0.52, 95% CI [-0.68, -0.37], t(2289) = -6.66, p < .001; Std. beta = -0.22, 95% CI [-0.29, -0.15])
* The effect of source [CanTimes] is statistically significant and negative (beta = -0.25, 95% CI [-0.41, -0.09], t(2289) = -3.07, p = 0.002; Std. beta = -0.13, 95% CI [-0.20, -0.06])
* The effect of source [CourierMail] is statistically non-significant and positive (beta = 0.05, 95% CI [-0.08, 0.17], t(2289) = 0.70, p = 0.483; Std. beta = -6.33e-03, 95% CI [-0.06, 0.05])
* The effect of source [HeraldSun] is statistically non-significant and positive (beta = 0.12, 95% CI [-7.55e-03, 0.24], t(2289) = 1.84, p = 0.066; Std. beta = 0.03, 95% CI [-0.02, 0.09])
* The effect of source [HobMercury] is statistically non-significant and positive (beta = 0.18, 95% CI [-1.69e-03, 0.35], t(2289) = 1.94, p = 0.052; Std. beta = 0.06, 95% CI [-0.02, 0.14])
* The effect of source [NorthernT] is statistically significant and positive (beta = 0.59, 95% CI [0.40, 0.79], t(2289) = 5.88, p < .001; Std. beta = 0.23, 95% CI [0.15, 0.32])
* The effect of source [SydHerald] is statistically significant and negative (beta = -0.35, 95% CI [-0.48, -0.23], t(2289) = -5.46, p < .001; Std. beta = -0.17, 95% CI [-0.22, -0.11])
* The effect of source [WestAus] is statistically non-significant and positive (beta = 0.06, 95% CI [-0.10, 0.22], t(2289) = 0.69, p = 0.489; Std. beta = -8.61e-03, 95% CI [-0.08, 0.06])
* The effect of scaled year is statistically non-significant and negative (beta = -9.25e-03, 95% CI [-0.02, 1.24e-03], t(2289) = -1.73, p = 0.084; Std. beta = -0.02, 95% CI [-0.03, -9.21e-04])