"This Candle Has No Smell": Detecting the Effect of COVID Anosmia on Amazon Reviews Using Bayesian Vector Autoregression

Nick Beauchamp

Deptartment of Political Science & Network Science Institute
Northeastern University
n.beauchamp@northeastern.edu

Abstract

While there have been many efforts to monitor or predict Covid using digital traces such as social media, one of the most distinctive and diagnostically important symptoms of Covid – anosmia, or loss of smell – remains elusive due to the infrequency of discussions of smell online. It was recently hypothesized that an inadvertent indicator of this key symptom may be misplaced complaints in Amazon reviews that scented products such as candles have no smell. This paper presents a novel Bayesian vector autoregression model developed to test this hypothesis, finding that "no smell" reviews do indeed reflect changes in US Covid cases even when controlling for the seasonality of those reviews. A series of robustness checks suggests that this effect is also seen in perfume reviews, but did not hold for the flu prior to Covid. These results suggest that inadvertent digital traces may be an important tool for tracking epidemics.

Introduction

A distinctive and early symptom of COVID-19 is anosmia, or loss of smell (Gane et al. 2020; Heidari et al. 2020; Han et al. 2020; Hannum et al. 2020; Meng et al. 2020). This symptom has become a valuable diagnostic tool for quickly distinguishing mild or early cases of Covid from other, similarly-presenting viral infections (Gerkin et al. 2020). While there have been suggestions that daily tests for anosmia could be a useful and inexpensive method for mass testing and surveillance (Larremore, Toomre, and Parker 2021; Parma et al. 2021; Menni et al. 2020; Weiss et al. 2021), such programs have yet to be implemented at scale. However, although individuals are often unaware of their own anosmia, they may nevertheless reveal this symptom through daily activities such as discussions and comments online. Explicit discussion of smell and odor, however, appears to be much less common than of other senses such as taste and flavor, so detecting traces of anosmia remains a challenge, even in world of abundant online social media data.

One domain where discussion of smell may be relatively common, however, is in online reviews of scented products, such as perfumes or scented candles. Popular scented candles and perfumes on Amazon have tens of thousands of reviews with multiple new reviews posted daily, most of them
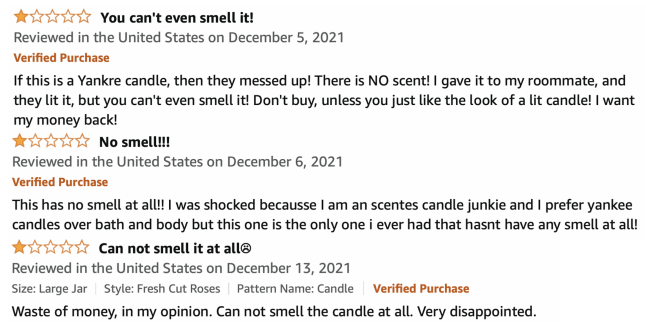
Figure 1: Amazon reviews complaining of a lack of smell.

explicitly discussing the quality of the smell of the product. It has been suggested in a number of popular posts on Twitter that Amazon complaints about scented candles lacking odor may reflect unwitting anosmia due to Covid, and that such complaints may reflect concurrent Covid surges (Nelson 2020; Dee 2021); see Figure 1 for examples.
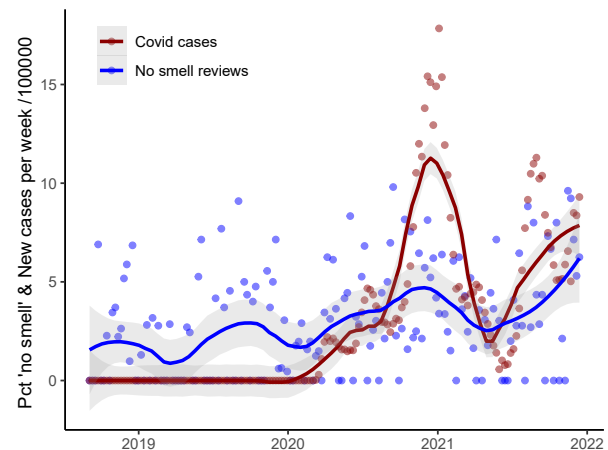

Figure 2: Weekly Covid cases vs "no smell" candle reviews.

But while there has been some preliminary analysis of this hypothesis using Amazon review data (Petrova 2020;

Beauchamp 2021), previous quantitative examinations have been cursory and potentially confounded by seasonal effects. As described in further detail below, Figure 2 shows weekly new US Covid cases (x100,000, red) juxtaposed with the percentage of reviews of the most popular scented candles on Amazon that mention "no smell" or "no scent" (blue). Both series rose during the winters of 2020 and 2021, but when the review series is extended prior to Covid back to 2018, it is evident that "no smell" review percentages rise every winter. Thus the apparent correlation may simply be due to seasonal coincidence.

To better ascertain whether there could be a causal connection between Covid cases and "no smell" reviews, a purpose-built Bayesian vector autoregression model was constructed, where cases and "no smell" reviews are each functions of both lagged cases and reviews. The model additionally controls for review seasonality via monthly effects or a sinusoidal function, and includes a hard prior that cases are 0 before February 2020 to increase the precision of the seasonal control. The core hypothesis is that lagged Covid cases will affect subsequent "no smell" reviews, once controlling for seasonal effects. A secondary hypothesis is that lagged reviews may additionally serve as a leading indicator of rising cases, due to delays in case reports relative to review posting.

After fitting the model, Covid cases were found to significantly affect "no smell" reviews, even when controlling for the seasonality of those reviews. No evidence was found that reviews predict cases, however. A number of robustness checks were performed, showing that (1) "no smell" complaints in reviews of a few popular perfumes are moderately affected by Covid as well; (2) positive, "good smell" reviews are uncorrelated with Covid; and (3) "no smell" complaints do not appear correlated with flu cases in the pre-Covid era. Together, these robustness checks suggest that the "no smell" results are unlikely to be an artifact specific to keywords or products, and do appear to reflect a likely causal relationship between unwitting Covid anosmia and Amazon complaints about lack of odor.

Related Work

Using social media and other digital traces in order to track and predict epidemics and other social phenomena has a long and fraught history. The flaws of early efforts to predict public opinion (O'Connor et al. 2010) or influenza (Ginsberg et al. 2009; Dugas et al. 2013) have by now been well-documented (Gayo-Avello, Metaxas, and Mustafaraj 2011; Lazer et al. 2014; Beauchamp 2017). Common pitfalls of these earlier efforts include using simplistic models that, for instance, merely correlate time series using hand-adjusted lags; ad-hoc or post-hoc natural language processing (NLP) decisions, including keyword or search term selection; limited geographical or temporal variation; and poor or absent out-of-sample validation for predictive models. A number of these drawbacks have been progressively improved upon (Serban et al. 2019; Samaras, García-Barriocanal, and Sicilia 2020), and the start of the Covid-19 pandemic marked a resurgence in interest in these techniques, with more sophisticated models, larger and more real-time data, and more

rigorous testing (Li et al. 2020; Fantazzini 2020; Gencoglu and Gruber 2020; Cuomo et al. 2021; Kogan et al. 2021).

The approach taken here and described below draws on many of these recent developments, particularly by adopting the vector autoregression approach within a Bayesian framework that allows priors that vary pre- and post-pandemic, as well as seasonal effects. Unlike some previous efforts, this approach unfortunately lacks geolocated data; but also unlike many previous "kitchen sink" approaches to detecting social media correlates of epidemics, the approach here is a hypothesis-driven examination of a specific, highly distinctive symptom that is especially elusive in the digital record. Indeed, the public unawareness of this symptom, and unconscious expression of it via misdirected complaints, may make for a particularly clean signal unconfounded by expectations – although as discussed in the Conclusion, the virality of Twitter means that that unawareness may now be lost.

Methods

Data 9837 reviews were scraped from the four best-selling "Yankee Candle" brand scented candles on Amazon dating from 9/1/2018 to 12/20/2021. For each of the 172 weeks in that timespan, the percentage of reviews that mentioned "no smell" or "no scent" was calculated. Manual examination suggests these are almost always associated with complaints about lack of odor; while other n-grams were explored, none meaningfully increased precision or recall. The mean number of reviews per week was 57 (sd 40, max 229), while the mean number of "no smell" reviews per week was 1.7 (sd 2, max 6). Covid cases were similarly aggregated by week, and are presented here in units of 100,000. For robustness checks, 6861 reviews of the four best-selling "Vera Wang" perfumes on Amazon were also collected, as well as weekly flu case counts (in units of 1000) from the CDC. There are no significant ethical issues raised by this dataset collecting public information.

Model The core model is a vector autoregression where:

1. Each week's total new US Covid cases (cov_t) is a function of the previous week's total cases (cov_{t-1}) plus the previous week's percentage of Amazon reviews featuring "no smell" complaints (rev_{t-1}).

2. Each week's percentage of "no smell" reviews is likewise a function of rev_{t-1} and cov_{t-1}, plus a control for the seasonal nature of "no smell" reviews via either monthly effects, or a sinusoidal control with two free parameters. Both controls produced very similar results, and the monthly controls variant is shown hereafter.

Thus we have:

$$cov_t \sim N(\mu_1, \sigma_1^2) \tag{1}$$
$$\mu_1 = \beta_0 + \beta_1 cov_{t-1} + \beta_2 rev_{t-1}$$
$$rev_t \sim N(\mu_2, \sigma_2^2) \tag{2}$$
$$\mu_2 = \beta_{m(t)} + \beta_3 rev_{t-1} + \beta_4 cov_{t-1}$$

Uninformative normal priors were placed on the various β coefficients, and inverse gamma priors on the various σ parameters. But unlike traditional vector autoregression (VAR)

methods, the customized Bayesian framework here allows the easy addition of priors that α, β_1, β_2 and β_4 all be 0 before the pandemic, since there were no Covid cases then. This increases the precision of the estimates of those parameters as well as the monthly effects $\beta_{m(t)}$ (results were weaker but somewhat similar using standard VAR with seasonal controls). The Bayesian model was estimated using Markov chain Monte Carlo sampling via rjags (v 4.12) in R (v 4.1.2), with four chains showing good convergence (Gelman-Rubin diagnostic = 1.01).

Results

The four core parameters of interest are the $\beta_{1:4}$ coefficients in equations (1) and (2). Posterior means and 95% credible intervals for each are shown in Figure 3. Those results show:

1. The effect of last week's Covid cases on the subsequent week's cases (β_1) is nearly 1, unsurprisingly.

2. The effect of last week's reviews on the subsequent week's Covid cases (β_2) is not significant. Reviews of course do not cause Covid, but nor do they appear to be a leading indicator, at least at this level of aggregation.

3. The effect of last week's reviews on the subsequent week's reviews (β_3) is not significant, reflecting both the noisiness of review data, as well as the effect of the seasonal monthly controls.

4. Most notably, the core parameter of interest, the effect of last week's cases on the subsequent week's percentage of "no smell" reviews (β_4), is significantly positive.
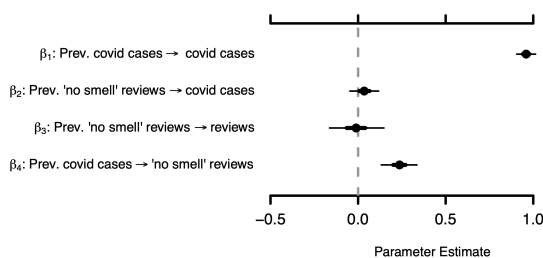

Figure 3: "No smell" candle model, β parameter Bayesian posteriors with 95% credible intervals.

At 0.24, the β_4 effect means that each additional 100,000 new cases per week (or 14000 cases per day) increases "no smell" reviews by a quarter of a percentage point. So for instance, in the week of 12/13/20-12/20/20, there were about a million new cases in the US, generating an estimated increase in "no smell" reviews of 2.4 percentage points in the subsequent week. This is a meaningful increase given that 25% of weeks have zero "no smell" reviews and 75% of weeks have fewer than 5% "no smell" reviews. And as Figure 2 suggests, this relationship appears to persist even through the Omicron wave, where preliminary reports show anosmia may be somewhat reduced relative to previous variants (Public Health England 2022; Karina-Doris et al. 2022).

Robustness Checks

There are a number of conceivable issues with these results that that are worth examining further:

1. It is possible that these results might be specific to "Yankee Candles" and not representative of a general tendency for undetected anosmia to be projected onto scented products. As an additional test of generality, then, 6861 reviews of the popular "Vera Wang" brand perfumes were also scraped from Amazon, and "no smell" percentages were calculated as before. Figure 4 shows a similar result to Figure 3: while the 95% credible interval touches 0 (mean = 0.10, 95% CI = [-0.01,0.21]), the mean effect is positive and similar to that seen for the candles. This suggests that unwitting anosmia may be evident in a wide variety of reviews discussing scented products.
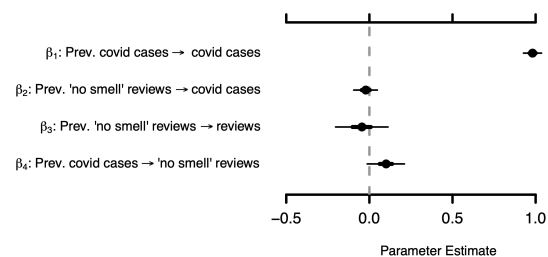

Figure 4: "No smell" perfume model, β parameter Bayesian posteriors with 95% credible intervals.

2. It is possible that, just as review counts in general, and "no smell" percentages in particular, undergo seasonal changes, it could therefore be the case that any similar phrase found in these reviews may undergo a seasonal shift that is correlated with Covid. As one test of this, the percentage of reviews that mentioned "good smell" and related synonyms was calculated instead, and the same model was fitted with this percentage as rev_t. Figure 5 shows the result of fitting this model; as can be seen, there is no relationship between "good smell" and Covid. Exploratory analysis also suggests that other phrases are similarly unrelated.
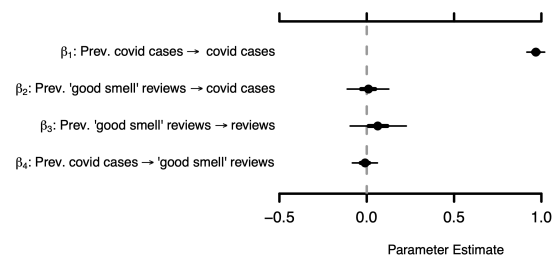

Figure 5: "Good smell" candle model, β parameter Bayesian posteriors with 95% credible intervals.
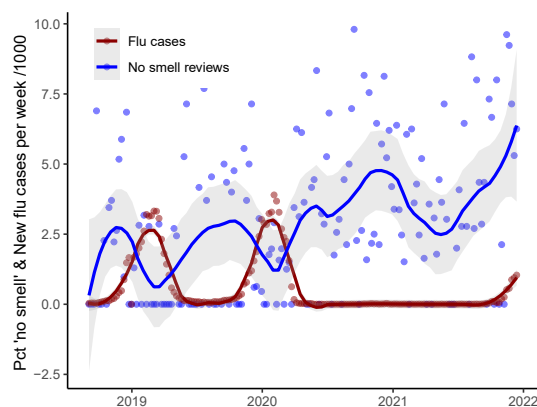
Figure 6: Weekly flu cases vs "no smell" candle reviews.

3. It is possible that these results hold more broadly for seasonal diseases such as influenza, which might affect either smell or disposition, and are not specific to Covid anosmia. To test this, weekly US flu case counts over the same time period were collected, and were used in place of cov_t in the model. The model was fit only on the pre-Covid period because, as Figure 6 shows, flu was basically eradicated during the first 18 months of Covid. As that figure also shows, though, flu waves do not appear synchronized with "no smell" reviews in way they were with Covid, and Figure 7 confirms that that flu cases are mostly unrelated to "no smell" reviews. It is possible that there is a mild negative relationship, but unlike the more complex Covid and "no smell" dynamics after February 2020, both reviews and flu undergo very regular waves prior to Covid that may be responsible for a spurious negative relationship even with seasonal review controls. A longer time-series with seasonal controls for flu may help resolve this; the necessary years of data are unavailable from Amazon, but may be obtainable from other similar sites or products in future work.
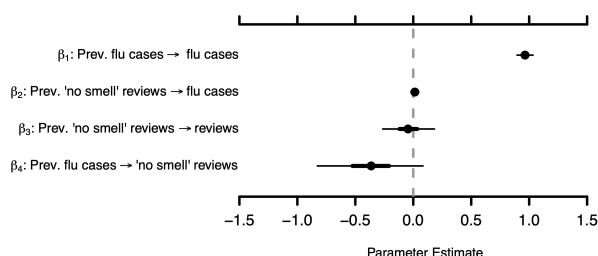

Figure 7: "No smell" candle model vs flu, β parameter Bayesian posteriors with 95% credible intervals.

Conclusion

While this study has found no evidence that reviews are a leading indicator of rising Covid cases, Covid does appear to significantly affect "no smell" review rates. This suggests

that social media sources such as Amazon reviews may be useful for surveillance, for measuring Covid-related behaviors, and potentially for early identification of rising cases when online posts appear more quickly than official statistics.

There are a number of plausible directions in which this particular analysis could be extended: Data sources could be expanded to other products, other online retailers, Google search terms, or social media posts. Symptoms could be expanded to include reviews indicative of ageusia (loss of taste); or complaints about food and other product-induced Covid-like symptoms such as headaches; or complaints about established medicines or homeopathic remedies failing to resolve certain symptoms. Review sentiment and rating scores could also be incorporated, perhaps thereby differentiating between measured criticism and the higher distress associated with illness.

As discussed above, there are many projects that cast a wider net across many media and large quantities of Covid-related symptom or behavioral language. The core idea behind this approach, however, is that unwitting symptoms may be especially revealing, perhaps even more revealing than symptoms that may be confounded by conscious expectations, and that this may particularly be the case when there happens to be a single symptom like anosmia that is both highly distinctive, yet underappreciated and little-discussed. But while revealing, inadvertent signaling is also fragile: although an earlier viral tweet had no apparent effect on reviews, in the aftermath of a more recent viral tweet and the ensuing media attention, "Yankee Candle" reviews have become contaminated with self-aware disavowals of Covid, as seen in Figure 8. These results illustrate both the power and fragility of this narrow, hypothesis-driven approach for finding inadvertent signals in the digital record.

Figure 8: Sample reviews contaminated after the viral event.

Acknowledgments

References

Beauchamp, N. 2017. Predicting and interpolating state-level polls using Twitter textual data. *American Journal of Political Science*, 61(2): 490–503.

Beauchamp, N. 2021. Here's a plot of the "no smell" complaints for the top three Yankee Candles on Amazon. https://twitter.com/nick_beauchamp/status/1473509106187223043. [Accessed 1/8/2022].

Cuomo, R. E.; Purushothaman, V.; Li, J.; Cai, M.; and Mackey, T. K. 2021. A longitudinal and geospatial analysis of COVID-19 tweets during the early outbreak period in the United States. *BMC Public Health*, 21(1): 1–11.

Dee, D. 2021. fresh wave of bad reviews for yankee candles. https://twitter.com/drewtoothpaste/status/1473360059908374539. [Accessed 1/8/2022].

Dugas, A. F.; Jalalpour, M.; Gel, Y.; Levin, S.; Torcaso, F.; Igusa, T.; and Rothman, R. E. 2013. Influenza forecasting with Google flu trends. *PloS one*, 8(2): e56176.

Fantazzini, D. 2020. Short-term forecasting of the COVID-19 pandemic using Google Trends data: Evidence from 158 countries. *Applied Econometrics, Forthcoming*.

Gane, S. B.; Kelly, C.; Hopkins, C.; et al. 2020. Isolated sudden onset anosmia in COVID-19 infection. A novel syndrome. *Rhinology*, 58(3): 299–301.

Gayo-Avello, D.; Metaxas, P. T.; and Mustafaraj, E. 2011. Limits of electoral predictions using twitter. In *Fifth international AAAI conference on weblogs and social media*.

Gencoglu, O.; and Gruber, M. 2020. Causal modeling of twitter activity during covid-19. *Computation*, 8(4): 85.

Gerkin, R. C.; Ohla, K.; Veldhuizen, M. G.; Joseph, P. V.; Kelly, C. E.; Bakke, A. J.; Steele, K. E.; Farruggia, M. C.; Pellegrino, R.; Pepino, M. Y.; et al. 2020. The best COVID-19 predictor is recent smell loss: a cross-sectional study. *MedRxiv*.

Ginsberg, J.; Mohebbi, M. H.; Patel, R. S.; Brammer, L.; Smolinski, M. S.; and Brilliant, L. 2009. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232): 1012–1014.

Han, A. Y.; Mukdad, L.; Long, J. L.; and Lopez, I. A. 2020. Anosmia in COVID-19: mechanisms and significance. *Chemical senses*, 45(6): 423–428.

Hannum, M. E.; Ramirez, V. A.; Lipson, S. J.; Herriman, R. D.; Toskala, A. K.; Lin, C.; Joseph, P. V.; and Reed, D. R. 2020. Objective sensory testing methods reveal a higher prevalence of olfactory loss in COVID-19–positive patients compared to subjective methods: A systematic review and meta-analysis. *Chemical senses*, 45(9): 865–874.

Heidari, F.; Karimi, E.; Firouzifar, M.; Khamushian, P.; Ansari, R.; Ardehali, M. M.; and Heidari, F. 2020. Anosmia as a prominent symptom of COVID-19 infection. *Rhinology*, 58(3): 302–303.

Karina-Doris, V.; Pouwels, K.; Peto, T.; Pritchard, E.; et al. 2022. Omicron-associated changes in SARS-CoV-2 symptoms in the United Kingdom. *Nuffield Department of Medicine, preprint*.

Kogan, N. E.; Clemente, L.; Liautaud, P.; Kaashoek, J.; Link, N. B.; Nguyen, A. T.; Lu, F. S.; Huybers, P.; Resch, B.; Havas, C.; et al. 2021. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. *Science Advances*, 7(10): eabd6989.

Larremore, D. B.; Toomre, D.; and Parker, R. 2021. Modeling the effectiveness of olfactory testing to limit SARS-CoV-2 transmission. *Nature Communications*, 12(1): 1–9.

Lazer, D.; Kennedy, R.; King, G.; and Vespignani, A. 2014. The parable of Google Flu: traps in big data analysis. *Science*, 343(6176): 1203–1205.

Li, C.; Chen, L. J.; Chen, X.; Zhang, M.; Pang, C. P.; and Chen, H. 2020. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance*, 25(10): 2000199.

Meng, X.; Deng, Y.; Dai, Z.; and Meng, Z. 2020. COVID-19 and anosmia: A review based on up-to-date knowledge. *American journal of otolaryngology*, 41(5): 102581.

Menni, C.; Valdes, A. M.; Freidin, M. B.; Sudre, C. H.; Nguyen, L. H.; Drew, D. A.; Ganesh, S.; Varsavsky, T.; Cardoso, M. J.; Moustafa, J. S. E.-S.; et al. 2020. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature medicine*, 26(7): 1037–1040.

Nelson, T. 2020. There are angry ladies all over Yankee Candles site. https://twitter.com/TerriDrawsStuff/status/1331362372179554304. [Accessed 1/8/2022].

O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; and Smith, N. A. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.

Parma, V.; Hannum, M. E.; OLeary, M.; Pellegrino, R.; Rawson, N. E.; Reed, D. R.; and Dalton, P. H. 2021. SCENTinel 1.0: development of a rapid test to screen for smell loss. *Chemical senses*, 46.

Petrova, K. 2020. Scented candles: An unexpected victim of the COVID-19 pandemic. https://github.com/kateptrv/Candles. [Accessed 1/8/2022].

Public Health England. 2022. SARS-CoV-2 variants of concern and variants under investigation in England. *Technical Briefing 34*.

Samaras, L.; García-Barriocanal, E.; and Sicilia, M.-A. 2020. Comparing Social media and Google to detect and predict severe epidemics. *Scientific reports*, 10(1): 1–11.

Serban, O.; Thapen, N.; Maginnis, B.; Hankin, C.; and Foot, V. 2019. Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification. *Information Processing & Management*, 56(3): 1166–1184.

Weiss, J. J.; Attuquayefio, T. N.; White, E. B.; Li, F.; Herz, R. S.; White, T. L.; Campbell, M.; Geng, B.; Datta, R.; Wyllie, A. L.; et al. 2021. Tracking smell loss to identify healthcare workers with SARS-CoV-2 infection. *PloS one*, 16(3): e0248025.