

# Individual differences in the Movement-Mood Relationship in Digital Life Data

**Glen Coppersmith**

Qntfy

glen.coppersmith@qntfy.com

**Alex B. Fine**

Qntfy

alex.fine@qntfy.com

**Patrick Crutchley**

Qntfy

patrick.crutchley@qntfy.com

**Joshua Carroll**

Qntfy

josh.carroll@qntfy.com

## Abstract

Our increasingly digitized lives generate troves of data that reflect our behavior, beliefs, mood, and wellbeing. Such “digital life data” provides crucial insight into the lives of patients outside the healthcare setting that has long been lacking, from a better understanding of mundane patterns of exercise and sleep routines to harbingers of emotional crisis. Moreover, information about individual differences and personalities is encoded in digital life data. In this paper we examine the relationship between mood and movement using linguistic and biometric data, respectively. Does increased physical activity (movement) have an effect on a person’s mood (or vice-versa)? We find that weak group-level relationships between movement and mood mask interesting and often strong relationships between the two for individuals within the group. We describe these individual differences, and argue that individual variability in the relationship between movement and mood is one of many such factors that ought be taken into account in wellbeing-focused apps and AI systems.

## 1 Introduction

Health and wellbeing research generally seeks to find patterns that hold for all members of a population. A familiar example is the claim that those who exercise more are happier (Stubbe et al., 2007). While this claim has intuitive appeal for most people, there are many individuals for whom this relationship does not seem to hold (e.g., someone who is challenged with chronic pain that is exacerbated by exercise). Where chronic pain is an extreme example, there are many more subtle ways that a person’s individual circumstances might cause them to deviate from expected population norms.

Generally speaking, whether this relationship holds across the population or varies across indi-

viduals is an empirical question, and one with profound implications for delivering effective clinical guidance and for the design of mental health and wellness technology (e.g., Menke, 2018). This may be one of the contributing factors to the difficulty that mental health interventions face with retention and attrition over the course of treatment: what was designed for the population does not necessarily adapt to a particular individual’s life. Preventing attrition is considered a longstanding and core challenge in the design and execution of studies and interventions alike (Eysenbach, 2005; Christensen and Mackinnon, 2006). This is more pronounced in digital mental health apps, many of which are designed to support long term behavior change, yet face significant difficulty retaining users, with a recent study indicating a median retention rate of just 3.3% of users retained after 30 days of usage (Baumel et al., 2019). This strongly suggests a need to understand the individual differences between users that might have an effect on retention and attrition and use that information to augment intervention approaches or suggest novel ones.

Collecting the data necessary to quantify these individual differences has been a challenge historically, especially with traditional behavioral methods (e.g., questionnaires). With the increasing ubiquity of mobile devices, the relevant data can now be captured and recorded to support large-scale, fine-grained analysis and intervention. Recent work shows that indices of mood, mental health, and wellbeing can be estimated from social media behavior (De Choudhury et al., 2013a,b; Coppersmith et al., 2017, 2016, 2015; Schwartz et al., 2016; Resnik et al., 2015; Cohan et al., 2016; Wang et al., 2014; Park et al., 2015; Eichstaedt et al., 2015). Here, we explore the relationships between mood, emotion, and mental health conditions derived from machine classifiers and Fitbit metrics.

## 2 Data

Users come from the [OurDataHelps.org](https://OurDataHelps.org) program, which enables participants to donate social media and wearable data to support mental health research. For each of the users ( $n = 160$ ) included in this analysis, we analyzed historic data from at least one source of language (Twitter or Facebook) and subsequent actigraphic data collected via a Fitbit device. All users had at least 30 days in which their wearable recorded data and in which posted at least once on social media. All data analyzed was from before the COVID-19 pandemic, associated lockdown, and changes in pattern of life that it induced. Users opted-in to data collection via OAuth, which was subjected to deidentification and stored following the ethical protocols of (Benton et al., 2017). Due to differences in models of wearable devices, users had different aspects of their movement recorded, so we analyzed data elements common across at least 20 users.

## 3 Methods

We analyzed language data using previously-trained models of mood, emotion, and mental health. Each model examines the text of social media posts using a simple lexicon or character  $n$ -gram language model (CLM), and produces a score relevant to a psychological variable.

We use models created by Coppersmith et al. to score for ADHD, anxiety, bipolar disorder, borderline personality disorder, depression, eating disorders, PTSD, and schizophrenia (Coppersmith et al., 2015). Briefly, these models estimate the relative likelihood that a given text was generated by a user at risk for a specific condition (e.g., PTSD) or a matched control, with one model created per condition. The data used to compare language was derived from users who made self-statements of diagnosis (e.g., “I was diagnosed with PTSD”) publicly on social media. For each user, we estimated age and gender via a classifier similar in spirit to (Sap et al., 2014). An age- and gender-matched control user was identified from a large English-speaking sample.

For each string of characters (i.e., character  $n$ -gram) the model measured how likely it was to occur in the population with the condition and in the matched controls. This forms the basis of the scoring for the model, optimized to provide a score even from short texts. While many machine learning open vocabulary approaches are tuned to look

at all the language that a person generates to estimate risk, the models used here are tuned to work for small amounts of text, given the present task. We refer the reader to Coppersmith et al. (2015) for further details on the pre-processing steps.

For scoring emotion, we used a CLM trained from messages that contain hashtagged emotions (e.g., #joy), from Coppersmith et al. (2016). For scoring sentiment, we used VADER, a closed-vocabulary and rule-based tool specifically tuned for social media data (Hutto and Gilbert, 2014). We report each individual sentiment separately (positive, neutral, negative) as well as the compound sentiment, meant to give a single overall score of the sentiment expressed in the text. We used DepecheMood to estimate mood, another closed-vocabulary approach, with high-coverage and high-precision (Staiano and Guerini, 2014).

All data of each type recorded from midnight to midnight in each user’s local timezone is collapsed into a single number capturing the value for that day. For language data this is the average score for each model across all messages. For wearables, we use the most straightforwardly interpretable version of the data (e.g., hours of sleep) as retrieved from the API. The movement and physical data recorded from the user’s wearable (steps, average heart rate) is similarly accumulated from midnight to midnight, with the exception of sleep data which, following Fitbit’s reporting feature, is recorded on the morning the user wakes up (e.g., the two hours of sleep from 10pm until midnight is included in the next day’s sleep total). Since we were primarily concerned with the relationship between movement and psychological variables measured by language, we excluded any day for which we did not have both movement and language data. Note that the unit of analysis of language here is the language generated in a single day, models tuned for relatively small amounts of text, like closed-vocabulary lexica and machine learning models trained to predict on short texts were ideal.

We calculated Pearson’s  $r$  for each person between each pair of variables, treating each day as a separate observation. Because this is an exploratory analysis and we wish to focus our discussion on effects that are most likely to hold promise for future work, we artificially set the  $r$  value to 0 for all subsequent analysis for any correlation where the  $p$ -value associated with Pearson’s  $r$  is greater than 0.01. This  $p$ -value was selected such that for

any pair of variables we compare, we would expect 1-2 of the 160 users to be spuriously identified as having a significant when no relationship existed. We opted for a more conservative cutoff here than the traditional  $p < 0.05$  since analyses at that  $p$ -value would allow for an expected 8 spurious correlations to be falsely indicated, which could significantly influence the subsequent analytic step. Furthermore, the exploratory nature of the work obviates the need to address multiple comparisons using a technique such as a Bonferroni correction.

## 4 Results

Figure 1 shows the correlation matrix with Pearson's  $r$  computed across all users. The models described above are shown in the same order on both axes. The color of each cell captures the Pearson's  $r$  between the variables, averaged across users, with white indicating a lack of correlation (an  $r$  near 0 or correlations with high  $p$  values which were treated as  $r = 0$ , as noted). Blue indicates positive correlation and red indicates negative correlation – the solid dark blue diagonal reflects the fact that each variable correlates perfectly with itself. The variables are grouped by the construct measured, separated by black lines: emotion, mental health conditions, mood, sentiment, and movement. While some significant relationship can be seen between various language and movement measures, the vast majority seem to be near  $r = 0$ . The notable exception is sleep onset latency (i.e., the amount of time it takes to fall asleep) which has generally negative relationships with positive emotions and moods and a positive relationship with negative emotions and moods. This finding is in line with other work examining the link between aspects of sleep and wellbeing (Short et al., 2013).

However, this picture shows nuance when we examine correlation matrices computed for each *individual*. Figure 1 shows exemplar correlation matrices for individual users. Note that significant relationships exist for individuals that were not observed for the group. This suggests that the relationships between psychological phenomena and aspects of movement are not uniform in direction or magnitude.

Figure 2 illustrates this point in more detail with a histogram of the distribution of correlations between a few measures of movement and mood. All correlations that were not significant were excluded from these histograms. Note that there are users

for whom there are statistically significant correlations, in both the positive and negative directions, of both large and small magnitudes. Many of the other histograms for other pairwise comparisons, excluded for brevity, show similar patterns. Taken with the previous results, this demonstrates that relationships between movement- and mood-related constructs exhibit sufficient individual-level variability in both direction and magnitude that inferences about these relationships must explicitly and quantitatively account for this variability.

These results highlight the need for personalized approaches to improving mental health and wellbeing through movement- or activity-based interventions.

## 5 Anecdotes

A subset of the users opted in to allow us to discuss the results and data with them in order to allow for validation of the findings.

For one user, many aspects of their sleep are more strongly correlated with emotions than for the population. The amount of time spent in bed was correlated with negative emotions and negatively correlated with joy. Similarly, the number of times they were awakened during the night was positively correlated with posts classified as angry or annoyed the following day. This suggests that this user's mood is particularly sensitive to sleep, relative to the general population. This aligned with the user's subjective impressions of their experience.

For another pair of users, we found significant correlations involving the time spent sedentary throughout the course of the day. For one user, the time spent sedentary during the day was positively correlated with positive mood outcomes, while the second user demonstrated a negative correlation between these two measures. Subjective reports from these users was consistent with these findings: the first indicated that if they were sitting still it meant that their children were being well-behaved, and thus was indicative of a pleasant day. The second reported, by contrast, that if they were sitting still throughout the day, that indicated a long day of meetings, which tended to increase their frustration and negative mood.

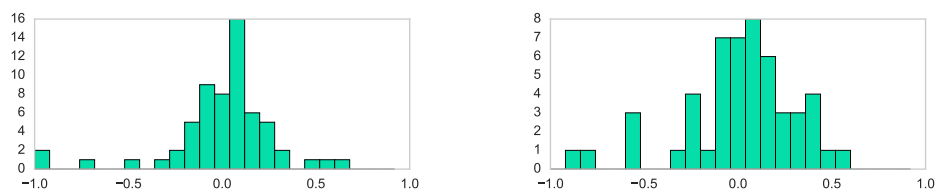
## 6 Discussion

We replicated previous work finding some significant relationships between movement and mood at a population level (i.e., sleep latency's relation-

Figure 1: Pearson's  $r$  values for correlations between model outputs, averaged across the population (Top) and exemplar individuals (Middle). Strength of correlation is indicated by color (Bottom) Note the strong correlations between language and movement measures, and that they are between different measures for different users.



Figure 2: Histogram of Pearson's  $r$  values for users with statistically significant correlations between (left) number of steps per day and posts with positive sentiment and (right) time asleep and posts with the blasé "don't care" mood.



ship to a range of psychological factors), while also demonstrating that significant relationships between movement and mood exist for individuals that do not hold across the population. This supports anecdotal and observational experience where, for population-level findings, there are individuals who seem to defy the expected trend.

The results reported here hold promise for future work, both theoretical and applied. Further study, with a larger subject pool, will allow us to examine *structured variability*, i.e., subpopulations with homogeneous relationships between movement and mood. There are well-established statistical techniques for characterizing and simultaneously modeling individual- and group-level relationships like those under discussion here, including multi-level modeling (Gelman and Hill, 2007), as well as numerous clustering techniques for inferring homogeneous subsets of users in a principled way (e.g., hierarchical clustering; Johnson, 1967). Without a strong *a priori* hypothesis for how many such homogeneous subsets of users exist, techniques with an inherent measure of cluster quality to suggest the number of clusters would be worthwhile. With the inherent relational nature of the data, it may be prudent to approach this clustering problem via techniques that take advantage of this information explicitly in the form of a (dis)similarity matrix (e.g., spectral clustering; Ng et al., 2001). Moreover, for developers of mental health and wellness technology that hinges on providing users with guidance related to movement and sleep, these results point the way forward for user testing that may enhance the quality and efficacy of these tools.

## 7 Caveats

Because the results reported here are based on donated digital life data, we expect this sample is biased in certain ways, assuming that the propensity to (1) share data without compensation, (2) actively contribute to mental health research, and (3) come across the donation opportunity at [OurDataHelps.org](https://OurDataHelps.org) are not uniformly distributed throughout the population. For example, in a project similar to [OurDataHelps.org](https://OurDataHelps.org), we solicited data donation from veterans of the US Armed Forces. To date, 22% of individuals that donated their data to this project identify as female. By contrast, according to the Department of Veterans Affairs, roughly 9% of US veterans are women (of Veterans Affairs et al., 2017). Thus, women are

over-represented in our sample. It is difficult to say exactly why this is, but the bias is most likely due to a confluence of factors, including gender-based differences in the propensity to participate in research that is considered altruistic or pro-social (e.g., Bani and Giussani, 2010) as well as idiosyncrasies in the way the study was promoted. However, we expect that this sort of bias would work *against* the observed pattern (i.e., since the population is more homogeneous than the general population, the relationship between movement and mood is less likely to vary significantly across individuals).

One underlying assumption of this work is that posts on social media have some reflection of the emotional state, mood, or other transient psychological phenomena that a person is experiencing. There is some controversy about the extent and strength of this relationship, with some finding significant reflections of emotion and mood in daily language (e.g., posts on social media Chen et al., 2020) while others fail to find these relations (e.g., in everyday speech Sun et al., 2020).

## 8 Conclusion

We empirically explored the relationship between a variety of movement and mood measures using social media posts and wearable data from 160 users. The relationships uncovered are more nuanced than the population-level conclusions that are generally popularized by the press and highlight the need for individualized approaches to movement-based wellbeing interventions.

Ultimately, understanding the relationship between movement and mood for a particular individual will allow for tailoring of wellbeing and mental health interventions to their specific needs, and thus increase our collective ability to tailor mental health and wellbeing interventions to the user. At minimum, this lays the foundation to provide some predictive ability for how a user may be willing to accept and engage with a suggested exercise-based intervention. The results reported here serve as a particularly strong indication of the promise held in personalized wellbeing interventions, and are consonant with a rich body of recent work highlighting the need for personalized medicine in general.

## Acknowledgments

We would like to thank the people who generously donated their data to [OurDataHelps.org](https://OurDataHelps.org), without whom this research would not be possible.

## References

- Marco Bani and Barbara Giussani. 2010. Gender differences in giving blood: a review of the literature. *Blood Transfusion*, 8(4):278.
- Amit Baumel, Frederick Muench, Stav Edan, and John M Kane. 2019. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *Journal of medical Internet research*, 21(9):e14567.
- Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 94–102.
- Lushi Chen, Walid Magdy, Heather Whalley, and Maria Klara Wolters. 2020. Examining the role of mood patterns in predicting self-reported depressive symptoms. In *12th ACM Conference on Web Science*, pages 164–173.
- Helen Christensen and Andrew Mackinnon. 2006. The law of attrition revisited. *Journal of medical Internet research*, 8(3):e20.
- Arman Cohan, Sydney Young, and Nazli Goharian. 2016. Triaging mental health forum posts. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 143–147.
- Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Denver, Colorado, USA. North American Chapter of the Association for Computational Linguistics.
- Glen Coppersmith, Casey Hilland, Ophir Frieder, and Ryan Leary. 2017. Scalable mental health analysis in the clinical whitespace via natural language processing. In *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*, pages 393–396. IEEE.
- Glen Coppersmith, Kim Ngo, Ryan Leary, and Tony Wood. 2016. Exploratory data analysis of social media prior to a suicide attempt. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, San Diego, California, USA. North American Chapter of the Association for Computational Linguistics.
- Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013a. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)*, pages 3267–3276. ACM.
- Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013b. Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- Johannes C. Eichstaedt, Hansen Andrew Schwartz, Margaret L. Kern, Gregory Park, Darwin R. Labarthe, Raina M. Merchant, Sneha Jha, Megha Agrawal, Lukasz A. Dziurzynski, Maarten Sap, Christopher Weeg, Emily E. Larson, Lyle H. Ungar, and Martin E. P. Seligman. 2015. Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, 26(2):159–169.
- Gunther Eysenbach. 2005. The law of attrition. *Journal of medical Internet research*, 7(1):e11.
- Andrew Gelman and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*, volume Analytical methods for social research. Cambridge University Press, New York.
- Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- Stephen C Johnson. 1967. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254.
- Andreas Menke. 2018. Precision pharmacotherapy: psychiatry’s future direction in preventing, diagnosing, and treating mental disorders. *Pharmacoeconomics and personalized medicine*, 11:211.
- Andrew Ng, Michael Jordan, and Yair Weiss. 2001. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14:849–856.
- Gregory Park, H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Michal Kosinski, David J Stillwell, Lyle H Ungar, and Martin EP Seligman. 2015. Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6):934.
- Philip Resnik, William Armstrong, Leonardo Claudino, Thang Nguyen, Viet-An Nguyen, and Jordan Boyd-Graber. 2015. Beyond lda: exploring supervised topic modeling for depression-related language in twitter. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 99–107.
- Maarten Sap, Greg Park, Johannes C. Eichstaedt, Margaret L. Kern, David J. Stillwell, Michal Kosinski, Lyle H. Ungar, and H. Andrew Schwartz. 2014. Developing age and gender predictive lexica over social media. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1146–1151.

- H Andrew Schwartz, Maarten Sap, Margaret L Kern, Johannes C Eichstaedt, Adam Kapelner, Megha Agrawal, Eduardo Blanco, Lukasz Dziurzynski, Gregory Park, David Stillwell, et al. 2016. Predicting individual well-being through the language of social media. In *Biocomputing 2016: Proceedings of the Pacific Symposium*, pages 516–527. World Scientific.
- Michelle A Short, Michael Gradisar, Leon C Lack, Helen R Wright, and Hayley Dohnt. 2013. The sleep patterns and well-being of australian adolescents. *Journal of adolescence*, 36(1):103–110.
- Jacopo Staiano and Marco Guerini. 2014. Depechemood: a lexicon for emotion analysis from crowd-annotated news. *arXiv preprint arXiv:1405.1605*.
- JH Stubbe, MHM De Moor, DI Boomsma, and Eco J C de Geus. 2007. The association between exercise participation and well-being: a co-twin study. *Preventive medicine*, 44(2):148–152.
- Jessie Sun, H Andrew Schwartz, Youngseo Son, Margaret L Kern, and Simine Vazire. 2020. The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*, 118(2):364.
- Department of Veterans Affairs et al. 2017. Women veterans report: The past, present, and future of women veterans. Retrieved from Washington, DC: [https://www.va.gov/vetdata/docs/specialreports/women\\_veterans\\_2015\\_final.pdf](https://www.va.gov/vetdata/docs/specialreports/women_veterans_2015_final.pdf).
- Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T Campbell. 2014. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 3–14. ACM.