

# Online Social Media Analytics Software as a Tool for Automating Data Collection: Feasibility Study and Concurrent Validity

Red Thaddeus D. Miguel<sup>1\*</sup> and Cara Isabella M. Uy<sup>1</sup>

<sup>1</sup>Ottawa, Canada

\*Corresponding Author: rtdmiguel@gmail.com

Facebook-based research is emerging, and social media analytics software programs may be a tool that could lower research costs. Unfortunately, the quality of the data extracted have not been documented. The aim of the study was to test the accessibility and efficiency of a social media analytics software in a Facebook study, and test concurrent validity in Likes extracted. We conducted a review of accessible online social media analytics software and selected one for a case study comparing it to manual extraction procedures. Eighteen software programs were found, of which, five were free. The selected software was used in the completion of a case study at no cost, but had a longer extraction time. Exact data points were matched in a few pages (33.9%), while mean Likes

between methods were similar ( $p=0.471$ ). The software had perfect ICC for half of the pages with the rest having “almost perfect”. Concurrent validity was high ( $CCC = 0.995$ ) with Bland and Altman plot showing only 5% of measurements outside the 95% agreement level. Social media analyzer software are accessible and can be used at no cost. Facebook Likes extracted through a selected software compared to manual extraction had strong agreement and validity.

*Keywords:* Facebook, Social Media Analyzer, Concurrence Validity, Concordance Correlation Coefficient, Intraclass Correlation Coefficient, Feasibility

---

**A**s of June 30, 2019, Facebook has recorded 2.41 billion active users worldwide (Facebook, 2019), with half of American daily users viewing the site several times in a day (Perrin & Anderson, 2019). Together with the rise in the number of users, Facebook has become a useful tool for conducting research owing to advantages such as reduced costs, ease in recruiting participants, increased representativeness, and improved participant selection (Whitaker et al., 2017). Furthermore, multiple Facebook data points, allows for different research investigations in several disciplines to be conducted. Theme-extraction procedures found five types of

studies that can be conducted with Facebook data: analyses of users, motivations and perceptions in using Facebook, studies focussing on identity presentation, dynamics of relationships between groups of Facebook users, and studies on privacy and information disclosure (Wilson et al., 2012). A review by Walker et al. (2018) examining social media platforms as a method for understanding experiences of youth with disabilities, found that Facebook can be used for observational, interactive, and a combination of online and offline studies. Additionally, the growing popularity of Facebook-based research may be attributed to the platform's ability to reflect social constructs and human behavior. One study found significant depression predicting accuracy using Facebook posts (Eichstaedt et al., 2018), while another found Facebook Likes to be significantly more accurate than humans in judging personality (Youyou et al., 2015). Together, many scientists now turn to Facebook as a source of critical and practical data.

Recently, researchers have been utilizing online data extraction tools to automate the process of data collection on Facebook pages. With the possible advantage of being an efficient data collection tool, these software programs may promote for more studies to be conducted on Facebook's rapidly evolving platforms. As with most innovation, however, the current gap in knowledge lies in whether these online tools can actually be relied on in terms of providing researchers with reliable data. This study attempts to fill this gap.

## **LITERATURE REVIEW**

### **Adaptations of Facebook-Based Research Methods**

Owing to the versatility and number of data points available on Facebook pages, Facebook-based studies have been emerging in different fields, applying a multitude of research designs. One study for example launched a web-based pilot survey through Facebook, on fertility apps as a pregnancy prevention tool, to understand user characteristics, user's perceived efficacy of apps, and design preferences of users (Starling et al., 2018). The vast reach of Facebook allowed the researchers to recruit 1,000 completed surveys from eligible app users in less than two weeks and found that these apps may lack the capabilities as a reliable pregnancy prevention device.

User behaviours in Facebook have likewise been documented and utilized in studies. In a study by Tefertiller (2018), the study's dependent variables were related to social media sharing intention and were measured using four questions representing four

different types of Facebook behaviors. Tefertiller's (2018) study, which aimed to investigate whether social capital or opinion leadership best predicted an individual's desire to produce online social network word of mouth, found that perceptions of opinion leadership play an important role in determining social sharing intentions. In another study, data was collected from Facebook accounts of 115 undergraduate students and found that relationship status and having the respondent's partner in the profile picture was significantly associated with relationship satisfaction (Cole et al., 2018).

Moreover, randomized controlled trials (RCT) are also being conducted with the aid of Facebook. In an RCT by Fiks et al., (2017) of 87 pregnant women, participants were randomized to either the Grow2Gether intervention arm or to a control group that would receive appointment reminders through text messages. The Grow2Gether intervention included a Facebook peer group for pregnant women, which the study found to be engaging to participants leading to improved feeding behaviors post-partum.

Additionally, another way Facebook has been utilized in research is in testing how user perception of Facebook may influence actual behavior, mood, or emotion. In one study by Pittman (2018), the researchers investigated how perceptions of social media might influence psychological well-being. Using the Social Presence Theory, the study found that immediacy and apparent intimacy provided by social media may offer some the psychological benefits that typically accompany traditional forms of relationships.

These are just a few of the many studies adopting Facebook-based research. The vast amount of insightful data available through Facebook pages, have led authorities to suggest its use (Inkster et al., 2016).

### **Employing Facebook 'Likes'**

In a collection of statistics on Facebook activity updated at the beginning of 2019, one will notice that the highest value recorded in any of the variables was the number of times the Facebook Like button was pressed (1.13 trillion times) (Aslam, 2019). The Like button, which was officially launched 10 years ago (Pearlman, 2009), is a "thumbs up" symbol that is utilized by Facebook users to manifest their appreciation or positive sentiment to content posted; making the Like button a valuable tool as an "emotional sensor" (Kessler, 2012). In a study investigating the use of the Like button in online social interaction, Eranti and Lonkila (2015) found that the Like button is used for an

assortment of purposes, which includes dating efforts, conversation regulation, and maintenance of social ties. Furthermore, the study found that networked Facebook audience affects liking behavior of users wherein users reflect their liking based on likes made by other users. Together, these findings by Eranti and Lonkila (2015), manifest that the Facebook Like button, though binary in design, allow for a quantifiable glimpse into an assembly of social interactions.

The aforementioned study could be a reason why researchers have been using the Facebook Like button as a primary predictor in certain behaviours. In a study on the effects of negative campaigning strategies during Israel's 2013 elections, Samuel-Azran et al., (2017) measured the number of Facebook Likes each negative political message received. The study found that posts that contained attacks on another candidate or party, and posts that contained comparisons between candidates did not attract significantly more likes than other messages. Furthermore, the study found messages containing responses to messages of political attacks attracted significantly more likes than other messages. With the use of the Like button as a measure of likeability, the authors were able illustrate the nature and situation of negative campaigning in Israel.

The Like button has likewise been used in longitudinal research studies. A study investigating the associations between marijuana displays on Facebook and self-reported behaviors, for example, looked at participants Likes section, which are composed of businesses and groups participants had Liked on their profile (Moreno et al., 2018). The study was conducted on first year students from two United States universities and had a monthly Facebook coding for 4 years. The study found that marijuana nonusers were more often posting marijuana references in the Likes section compared to marijuana users. The authors suggested that the higher frequencies of posts in Likes section among nonusers may be because these pages often referred to marijuana advocacy groups; signaling an engagement among nonusers of the movement for legalization occurring during this period. Although the Like button was not used to directly measure the level of likeability of a post, the Like button as the basis of Likes section, paved the way for the authors to suggest a possible phenomenon that could have been taking place during this period when marijuana legalization was a well talked about subject.

With its applicability in different study designs, researchers have been able to use the Facebook Like button to shed light on diverse subjects, including issues in marketing (Buchanan et al., 2018), health promotions (Kite et al., 2016), impact of scientific work (Ringelhan et al., 2015), and voting behaviour (Kristen et al., 2017).

### **Online Social Media Analytics**

Despite its many applications, probable deterrents to social media research include the amount of time required to conduct a thorough evaluation of the pages, and the need to extract data rapidly from Facebook pages due to the constant changes that take place (Gough et al., 2017). In aid of this difficulty, researchers are beginning to suggest the use of online social media analytics tools (Gough et al., 2017). Online social media analytics tools are third party or social media platform owned software programs that gathers a list of information from social media pages (Lee, 2019). Some studies have used these tools to extract Facebook site frontpage data points and about page information, activity level, interaction with visitors, and number of Likes and connections (Rubio, 2015; Seidel et al., 2018). Social media analytic tools have the potential to gather information from Facebook pages with minimal time and, for some software programs, at no cost.

### **Research Question**

Though promising, to our knowledge, online social media analytics software programs have not been validated as a tool in research studies. Hence, the aim of this study was to answer the following questions:

**Research Question 1:** Are social media analytics tools easily accessible for researchers?

**Research Question 2:** Can an online social media analytics software program be used in a case study research to provide data from Facebook pages in a shorter amount of time compared to manual extraction procedures?

**Research Question 3:** Will a free online social media analytics software program provide similar number of Likes from Facebook public pages to Likes manually extracted from the same pages?

We therefore conducted a search for online social media analytics software programs, a case study to test possible differences in data collection time, and compared the Likes extracted by the social media software versus manual extraction procedures.

## **METHODS**

### **Design and oversight**

We conducted a feasibility study to determine the attainability of tackling a research investigation using an online social media analytics software program. The feasibility study required the implementation of a comprehensive search of available online media social analytics programs and a case study. Thereafter we analyzed the concurrent validity of the online social media analytics software as a method for extracting Likes from Facebook pages. The nature and design of the study did not require formal ethics committee review. Consent is provided under Section III of the Facebook Data Policy, therein stating that posts and comments published in public pages may be used for research purposes (Facebook, 2018).

### **Feasibility study and measures**

The search for online social media analytics software was conducted from April 9 to 11, 2019. The search used Google as the search engine and search terms were: “social media analyzer”, “online social media analytics software”, “social media analyst”, “Facebook analyzer”, and “Facebook audit software”. From the list identified software programs, one product was selected and compared with manual extraction procedures. The social media analyzer was selected based on not having any acquisition cost and its ability to extract number of Likes.

The next step for the feasibility study entailed conducting data collection procedures for a case study with the selected software. We designed a case study with the objective of determining Facebook Likes among public pages within the same theme. The cross-sectional study was designed to approximate an investigation capable of conducting a comparative analysis of Likes between pages. Though it would have been optimal to define other variables a priori to conduct other analyses in the case study, we decided to collect only Facebook Likes for the following reasons: (1) an initial search revealed that brands of social media analytics software differ in the variables they extract, however, most collect the number of Likes per page as a primary data point, (2) the scoping search also showed that other variables extracted were qualitative in nature and were not extracted in the same manner for all software thus confounding a possible content

analysis, and (3) as earlier shown, Facebook Likes by itself is a predictive measure for private attributes (Kosinski et al., 2013).

### **Participants**

Facebook pages were selected when the content and purpose of the page was in line with the predefined theme and were noted to be active pages (at least one post published within the past year). Pages were included regardless of time since creation and number of followers. As the current study was conducted before May of 2019, the theme used was the Philippine midterm elections, which was scheduled on May 2019. The Facebook pages therefore were public pages of the candidates vying for a seat in the Philippine Senate. The authors selected the theme and consequent pages because the event (midterm elections) and its timing allowed for increased activity throughout the pages. Furthermore, Facebook Likes have been shown to have an association with election outcomes (Zhang, 2016; Kristensen et al., 2017).

### **Procedures**

The list of senatorial candidates was compiled through the Commission on Elections (Rosette, 2019). Thereafter, all candidates were searched on Facebook for their verified public pages. When no verified pages for the senatorial candidate was found, the public page with the highest number of Likes and followers was used as a proxy for the candidate. After pages were screened for inclusion, and the pool of pages selected, the computers used for the search were restarted.

A simultaneous data extraction was conducted by both authors using two different computers using the same internet access point and web browser. While one author run extraction through the social media analysis software, the other author manually extracted data from the public Facebook page. Both authors completed recording their data on Microsoft Excel 2019 on one Facebook page before proceeding to the next. All web browsers were closed between every candidate. Data collection for all pages were completed on 1 day approximately one month prior to the midterm elections.

Data extraction time was measured as the time elapsed from the point the researcher opened the web browser to the point that all needed data was encoded, and the web browser was closed. Duration of the complete data extraction per senatorial candidate was recorded as “< 2 minutes”, “2 to 10 minutes”, and “> 10 minutes”.

### *Statistical Analysis*

All analyses were performed using R version 3.5.3 (R Core Team, 2018). Statistical tests were two-tailed, with P values  $\leq 0.05$  considered statistically significant. Variables were tested for normality by the Shapiro-Francia test. For the feasibility study, descriptive statistics are presented for characteristics of software uncovered and data extraction time.

Facebook pages were categorized into four activity levels, “Low”, “Moderate”, “High” and “Very High” to test whether the number of Likes would affect the accuracy of the social media analytics tool in gathering valid responses. Categories were based on quartile range of the number of Likes a page received, with pages within the 1st quartile reflecting “Low” activity, pages within the 4th quartile considered “Very High” level of activity, and pages with Likes at the 2nd and 3rd quartiles encompassing “Moderate” and “High” level of activity, respectively. Between manually extracted procedures and data extraction with social analytics software tool, descriptive statistics are given by level of activity. Statistical significance of differences between Facebook Likes was analyzed using an independent samples t-test or a Mann-Whitney U test where appropriate.

To analyze agreement between methods of extracting Facebook Likes across different page activity levels, outlier pages was first removed as identified by the extreme studentized deviate method. Thereafter, the Intraclass Correlation Coefficient (ICC) was used. The ICC tested agreement between the Likes recorded manually and the number of Likes extracted by the social media analytics software by measuring the overall data variance due to variability between pages (Shrout & Fleiss, 1979).

Concordance Correlation Coefficient (CCC) was then used to test concurrent validity of the social media analytics software. The CCC was used alongside ICC as it does not assume a common mean for the ratings between two measurements. Thus, CCC is able to assess not only the agreement between two measures but also the level of disagreement (Liu et al., 2016). The CCC is evaluated based on the degree that paired measurements fall on the 45-degree line. The CCC has the form (Lin, 1989):

$$CCC = \frac{2 \rho \sigma_Y \sigma_X}{(\mu_Y - \mu_X)^2 + \sigma_Y^2 + \sigma_X^2}$$

Where:  $\rho$  = correlation coefficient of Facebook page Likes between the two extraction methods  
 $\mu_Y$  and  $\mu_X$  = means of Likes between both measures  
 $\sigma_Y$  and  $\sigma_X$  = corresponding variance



The accepted level of agreement for ICC (0.80: almost perfect agreement) (Portney & Watkins, 2000) and CCC (0.99: near perfection) was defined by previous publications (McBride, 2005). A Bland and Altman plot was also presented to report the agreement between Facebook page Likes extracted by both methods (Bland & Altman, 1986).

## RESULTS

### *Research Question 1: Are social media analytics tool easily accessible for researchers?*

Eighteen unique social media analytics software were identified with the comprehensive search. The full software was available at no cost for five of the social media analytics software, while the cost of acquiring the full software for others were from 15 USD per month to 599 USD per month, with four companies offering customizable monthly plans depending on the services needed. Eleven had free trial versions available with free use ranging from seven to 50 days (mean 22.5 [SD: 12.5]).

Majority (n= 17, 94%) of the identified software were provided by third party companies. Most software (n=15, 83%) mentioned that they extracted Facebook Likes, however, only three (17%) were without cost, from which one was selected for the case study.

### *Research Question 2: Can an online social media analytics software program be used in a case study research to provide data from Facebook pages in a shorter amount of time compared to manual extraction procedures?*

Of the 62 senatorial candidates in the 2019 Philippine senatorial midterm elections, 59 of the candidates had accessible public Facebook pages and consequently were included in the case study. The data extraction time for Facebook pages in manual extraction took no longer than 2 minutes for all pages. Compared to manual extraction, the selected software was longer with 19 Facebook pages (33%) taking longer than 2 minutes but not more than 10 minutes. One Facebook page had a data extraction time longer than 10 minutes with the social media analytics software.

*Research Question 3: Will a free online social media analytics software program provide similar number of Likes from Facebook public pages to Likes manually extracted from the same pages?*

Of all included pages, the number of Likes per page as reported by manual extraction ranged from 22 to 9,797,724 with a mean of 442,511.6 Likes per page (SD:1,365,377.7) and a median of 11,620.0 Likes per page (IQR: 1,204.5-250,628.0). Exact matches were more common in Low activity pages (n=13, 86.7%), followed by Moderate (n=4, 28.6%), Very High (n=2, 13.3%), and High activity pages (n=1, 6.7%).

Shapiro-Francia test found the data was not normally distributed ( $W=0.317$ ;  $p<0.0001$ ). Though no statistically significant difference was observed in Likes extracted between groups by activity level, the greatest difference in Likes extracted between manual extraction method and social analytics software was noted in the Very High activity pages ( $p=0.471$ ), followed by High, Moderate, and Low activity pages (Table 1).

Table 1

*Facebook Likes extracted through manual extraction method and social media analytics tool by level of activity*

Activity Level	Facebook pages (n)	Range, Likes: Manually extracted	Range, Likes: Analytics software	Median Likes (IQR): Manually extracted	Median Likes (IQR): Analytics software	p-value*
<b>Low</b>	15	22-1,085	22 to 1,085	276 (193-526)	276 (192-526)	0.984
<b>Moderate</b>	14	1,324-7,044	1,322-7,039	2,820 (1,735-4,336)	2,815 (1,732-4,332)	0.944
<b>High</b>	14	11,620-234,894	11,614-229,336	78,537 (53,557-146,349)	60,885 (44,071-127,751)	0.497
<b>Very High</b>	15	266,362-9,797,724	266,350-8,399,498	968,340 (552,556-1,356,876)	968,340 (439,451-1,346,399)	0.471

\*results of U Mann-Whitney test.

There was perfect agreement between manually extracted Likes and software extracted Likes for Low and Moderate activity level with an ICC of 1 for both. High activity level and Very High had lower agreement with ICC = 0.97 (95% CI: 0.88-0.99) and

ICC = 0.98 (95% CI: 0.96-1.00), respectively. However, both ICC of High and Very High still fall within “almost perfect” agreement.

In terms of concurrent validity, the social media analytics software had strong agreement with manually extracted data, CCC = 0.995 (95% CI: 0.992-0.997). Using the lower border this equates to “near perfection”. CCC and Bland and Altman plot reflecting the relationship between the number Likes extracted by the software and manually extracted data are presented in Figure 1 and Figure 2, respectively. In the Bland and Altman plot we found that only 5% of all measurements of Likes between the two data extraction methods fell outside the 95% agreement level indicating a strong agreement between the two methods.

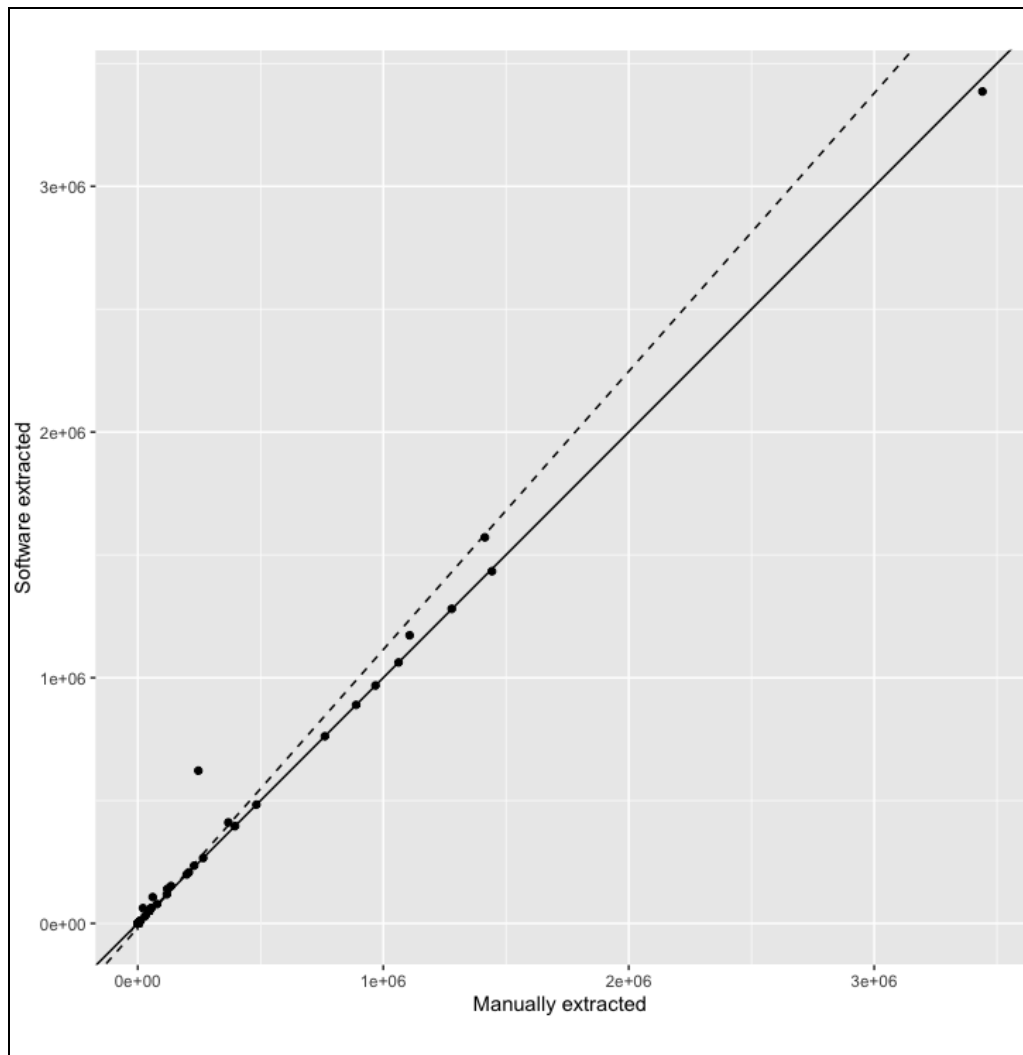


Figure 1. Plot of concordance correlation coefficient for manually extracted and software extracted data

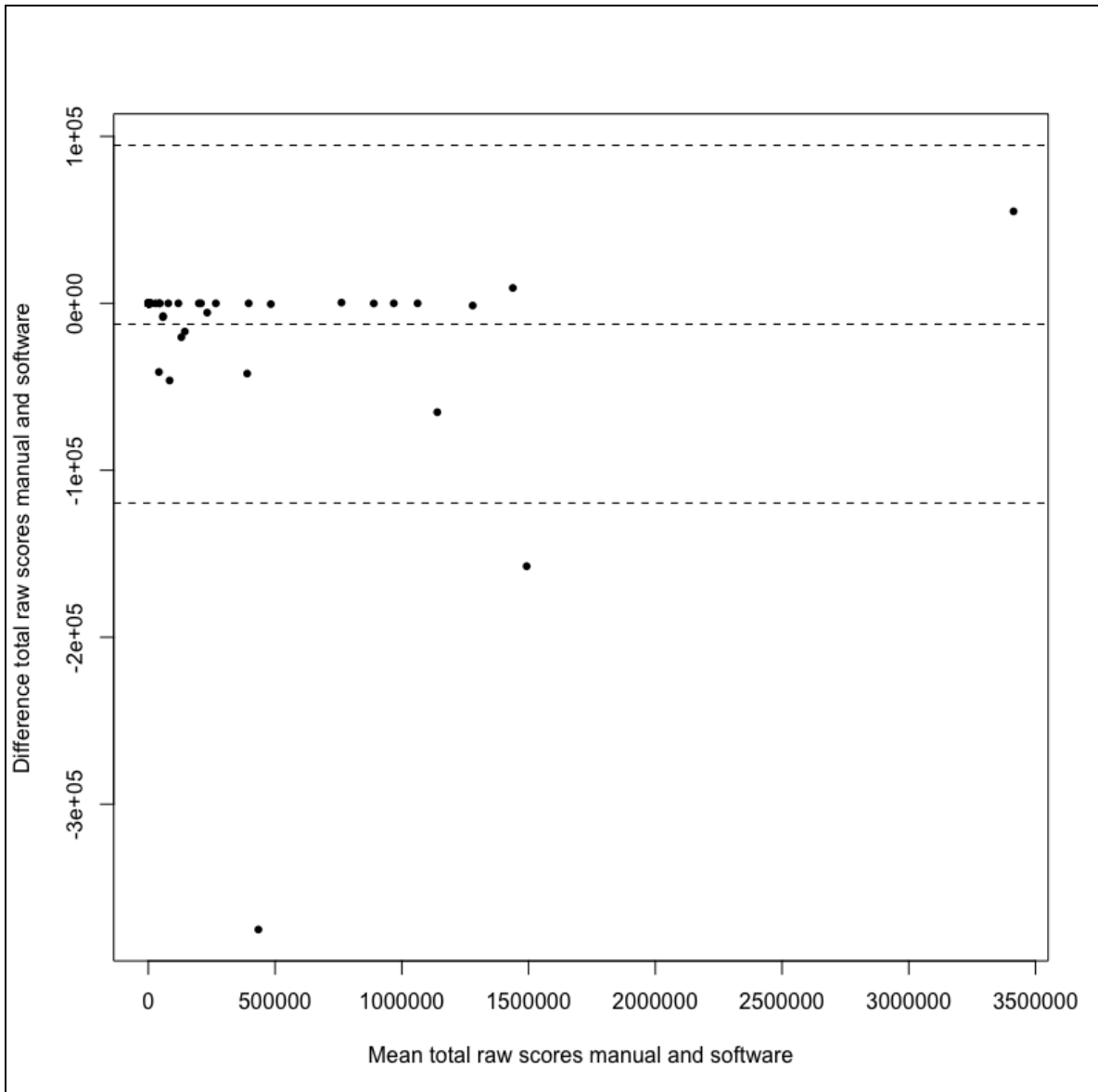


Figure 2. Bland and Altman plot for manually extracted and software extracted data

## DISCUSSION

The selected social media analytics software was accessible and available free of charge for the conduct of the research. However, the utilized social media analytics software we used had a longer data collection time than manual extraction. One probable explanation for the delay in data extraction was that the software was also collecting data not required by the case study. Thus, when comparing only Likes between methods of extraction, manual extraction was faster. We consequently concede to the possibility that if we were to compare other data points, data extraction time could have been faster with a social media analytics software. We therefore suggest other researchers undergoing

extraction with social media analytics tools to first understand each software's capabilities and select the software that is best suited for their objectives. Using a software, which collects unnecessary data, could prolong data extraction procedures.

An observed advantage of the social media analytics software was that it was available at no cost. Although we initially hypothesized that the use of the free software would be hindered by marketing techniques (e.g. software would stop after a few pages and would not proceed unless further services are paid for, or complete data will not be provided for free after a certain time), the social media analyst software used by our case study was not impeded by marketing methods and we were able to complete our case study. This is of particular importance especially for individuals conducting studies with limited or no financial support. In studies where a lack of financial support limits the conduct of research, researchers whose investigations may be conducted around Facebook pages could turn to social media analytics software as a cost-effective method to collect data (Alemayehu et al., 2018; Serugaa et al., 2013).

It is noteworthy, however, that exact matches in extracted Likes between the social media analytics tool and manual extractions were only found for 20 of the 59 pages (33.9%). These discrepancies were unlikely to be a function of delays between extraction times as the Likes extracted by the social media analytics software was noted to be either higher or lower than the Likes extracted manually despite the shorter duration by manual extraction. Furthermore, the size of the discrepancies in some cases were wide enough to make it extremely unlikely to have been caused by Likes fluctuating in the given amount of time. Therefore, the difference in Likes extracted may be explained by the use of algorithms that estimates the Likes based on other factors in the Facebook page.

Despite not having exact number of Likes for most of the pages reviewed, the study did find that the social media analytics software had excellent concurrent validity with a CCC of 0.995. This was further evidenced by the Bland and Altman plot showing majority (95%) of all ratings within the acceptable agreement level. Likewise, the ICC was at least "almost perfect" in agreement between all four quartiles. Thus, if the study being conducted by other researchers are comparing Likes between groups, the data extracted by social media analytics software may provide adequate analysis; this is especially true when the pages being compared have lower number of Likes. In our case, we found that

the Low (median: 276, IQR:193-526;  $p=0.984$ ) and Moderate (median: 2,820, IQR: 1,735-4,336;  $p=0.944$ ) activity pages had the least statistical difference between Likes extracted manually and Likes extracted by social media analytics software.

The study is not without limitations. One, the name of the software used was withdrawn from the text for legal reasons and to avoid advocating for a specific software. In turn, we were not able to show the other characteristics and capabilities, which were specific to the software. This additional information could have been useful to researchers in designing their own research around the tool and could explain why the social media analytics software took a longer time to collect the data than manual extraction. Indeed, if the case-study was more complex, we deduce that the social media analytics software could be more efficient than manually extracting data.

Second, the findings are based on one social media analyzer tool and the findings are therefore not generalizable to other available tools. Other social media analyzer tools may have different outcomes compared to the one we selected for our case study.

Third, we omitted the names of the pages and only public pages were used instead of incorporating any private pages. Though, the study may have been more comprehensive if private accounts were used, the decisions to present the study anonymizing included public pages was made for ethical reasons. Studies and guidelines have suggested that information on Facebook pages can be used for research without consent because of its nature as a document review and its design involves subjects on a public space (Wilkiimson & Thelwall, 2011; Hudson & Bruckman, 2004; American Psychological Association, 2010). However, due to the blurred lines in Facebook research we decided to uphold the highest level of care in dealing with social media data by protecting the privacy of the pages included and only including public pages (Zimmer, 2010; Flicker, Haans, & Skinner, 2004; Eysenbach & Till, 2001).

Despite these limitations, the study was able to evaluate a social media analytics software program as a tool for data collection. To the best of our knowledge, supplemented by a literature review, this is the first study of its kind. Furthermore, the study found and used a software that can be attained by other researchers at no cost. The findings and their implications are therefore useful for researchers who wish to conduct investigation with Facebook pages who are limited by financial and time constraints. Lastly, the

strength of the study rests on the use of Likes in its validity testing. Though future studies may not fully revolve only on Facebook page information, the Like variable could be used in many research to compare agreeability and/or popularity.

Moving forward, further research is warranted in finding the optimal social media analytics software for specific research designs and variations in data types. Further, there is a gap in knowledge when it comes to the algorithm of these software programs and how best to utilize them for research.

## CONCLUSION

Social media analytics software programs are easily accessible to researchers. Despite not collecting the exact number of Likes in most of the pages, extracting Likes through free Facebook analytics software was tested for concurrent validity and was found to have strong agreement with manually extracted data. The software can be a cost-effective method in collecting Facebook based data. Certain limitations hindered the current study from conducting a full review of the software, thus, further research in this field is warranted.

## References

- Adel, R. (2019). List: certified senatorial candidates for 2019 elections. Retrieved from <https://www.philstar.com/headlines/2019/02/06/1891374/list-certified-senatorial-candidates-2019-elections/amp/>
- Alemayehu C, Mitchell G, Nikles J. (2018). Barriers for conducting clinical trials in developing countries- a systematic review. *Int J Equity Health*. 17(1):37. doi:10.1186/s12939-018-0748-6
- American Psychological Association. (2010). Ethical principles of psychologists and code of conduct. Available at: <https://www.apa.org/ethics/code/principles.pdf> (accessed 4 August 2016)
- Aslam, S. (2019). Facebook by the numbers: stats, demographics, & fun facts. Retrieved from <https://www.omnicoreagency.com/facebook-statistics/>
- Bland, J.M., Altman, D.J. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 8:307–310.
- Buchanan, L., Yeatman, H., Kelly, B., Kariippanon, K. (2018), A thematic content analysis of how marketers promote energy drinks on digital platforms to young Australians. *Australian and New Zealand Journal of Public Health*, 42: 530-531. doi:10.1111/1753-6405.12840

- Cole, A., Leonard, M.T., McAuslan, P. (2018) Social media and couples: what are the important factors for understanding relationship satisfaction? *Cyberpsychology, Behavior, and Social Networking*, 21:9. <http://doi.org/10.1089/cyber.2017.0425>
- Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preotiuc-Pietro, D., Asch, D.A., Schwartz, H.A. (2018) Facebook language predicts depression in medical records. *Proc Natl Acad Sci USA*, 115(44): 11203–11208. doi:10.1073/pnas.1802331115
- Eranti, V., Lonkila, M. (2015). The social significance of the Facebook Like button. *First Monday*, 20(6). doi: <http://dx.doi.org/10.5210/fm.v20i6.5505>
- Eysenbach, G., Till, J.E. (2001). Ethical issues in qualitative research on Internet communities. *BMJ*. 323(7321):1103–1105.
- Facebook (2019). Stats. Retrieved from <https://newsroom.fb.com/company-info/>
- Facebook. Data Policy. Retrieved from [https://www.facebook.com/full\\_data\\_use\\_policy](https://www.facebook.com/full_data_use_policy)
- Fiks, A. G., Gruver, R. S., Bishop-Gilyard, C. T., Shults, J., Virudachalam, S., Suh, A. W., Gerdes, M., Kalra, G.K., DeRusso, P.A., Lieberman, A., Weng, D., Elovitz, M.A., Berkowitz, R.I., Power, T.J. (2017). A social media peer group for mothers to prevent obesity from infancy: the Grow2Gether randomized trial. *Childhood Obesity*, 13(5): 356–368. doi:10.1089/chi.2017.0042
- Flicker, S., Haans, D., Skinner, H. (2004). Ethical dilemmas in research on Internet communities. *Qual Health Res*. 14(1):124–134. doi: 10.1177/1049732303259842.
- Gough, A., Hunter, R. F., Ajao, O., Jurek, A., McKeown, G., Hong, J., Barrett, E., Ferguson, M., McElwee, G., McCarthy, M., Kee, F. (2017). Tweet for Behavior Change: Using Social Media for the Dissemination of Public Health Messages. *JMIR public health and surveillance*, 3(1), e14. doi:10.2196/publichealth.6313
- Hudson, J.M., Bruckman, A. (2004). ‘Go away’: Participant objections to being studied and the ethics of chatroom research. *The Information Society*. 20(2): 127–139.
- Inkster, B., Stillwell, D., Kosinski, M., Jones, P. (2016). A decade into Facebook: where is psychiatry in the digital age? *Lancet Psychiatry*, 3 (11): 1087–1090. [https://doi.org/10.1016/S2215-0366\(16\)30041-4](https://doi.org/10.1016/S2215-0366(16)30041-4)
- Kessler, A. (2012). The button that made Facebook billions. Retrieved from <https://www.wsj.com/articles/SB10001424052970204652904577196992203069570>
- Kite, J., Foley, B.C., Grunseit, A.C., Freeman, B. (2016). Please Like me: Facebook and public health communication. *PLoS ONE*, 11(9): e0162765. doi:10.1371/journal.pone.0162765
- Kosinski, M., Stillwell, D., Graepel, T. (2013) Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci USA*. 110(15):5802-5. doi: 10.1073/pnas.1218772110.
- Kristensen, J.B., Albrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M., Bornakke, T. (2017). Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PLoS ONE*, 12(9): e0184562. <https://doi.org/10.1371/journal.pone.0184562>
- Lee, K. (2019). Know what's working on social media: 27 paid and free social media analytics tool. Retrieved from <https://buffer.com/library/social-media-analytics-tools>
- Lin, L. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 45: 255–268.



- Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., Tu, X.M. (2016). Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry*. 28(2):115–120. doi:10.11919/j.issn.1002-0829.216045
- McBride, G.B. (2005). A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient. [PDF file] NIWA Client Report: HAM2005-062; p. 6. Retrieve from <https://www.medcalc.org/download/pdf/McBride2005.pdf>
- Moreno, M. A., Kerr, B., Lowry, S. J. (2018). A longitudinal investigation of associations between marijuana displays on Facebook and self-reported behaviors among college students, *Journal of Adolescent Health*. doi:10.1016/j.jadohealth.2018.03.017
- Pearlman, L. (2009). "I like this". Retrieved from <https://www.facebook.com/notes/facebook/i-like-this/53024537130>
- Perrin, A., & Anderson, M. (2019). Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Retrieved from <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>
- Pittman, M. (2018). Happiness, loneliness, and social media: perceived intimacy mediates the emotional benefits of platform use. *The Journal of Social Media in Society*, 7 (2): 164-176.
- Portney, L.G., Watkins, M.P. (2000). Foundations of clinical research Applications to practice. Prentice Hall Inc. New Jersey ISBN 0-8385-2695-0. 560-567.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Ringelhan, S., Wollersheim, J., Welpel, I.M. (2015). I Like, I cite? Do Facebook Likes predict the impact of scientific work? *PLoS ONE*, 10(8): e0134389. doi:10.1371/journal.pone.0134389
- Rubio AB. (2015) The Colombian radio in front of the digital challenge. In Mateos-Perez J, ed. Comunicación y medios No 31. Chile: Instituto de la Comunicacion e Imagen, ICEI, Unibersidad de Chile. Page 120.
- Samuel-Azran, T., Yarchi, M., Wolfsfeld, G. (2017). Engagement and likeability of negative messages on Facebook during Israel's 2013 elections. *The Journal of Social Media in Society*, 6(1): 42-68.
- Seidel RL, Jalilvand A, Kunjummen J, Gilliland L, Duszak R Jr. (2018) Radiologists and Social Media: Do Not Forget About Facebook. *J Am Coll Radiol*. 15(1 Pt B):224-228. doi: 10.1016/j.jacr.2017.09.013. Epub 2017 Nov 11.
- Serugaa, B., Sadikovb, A., Cazapc, E.L., Delgadod, L.B., Digumartie, R., Leighlf, N.B., Meshrefg, M.M., Minamih, H., Robinsoni, E., Yamaguchij, N.H., Pylek, D., Cufperl, T. (2013). Barriers and challenges to global clinical cancer research. *Oncologist*. 19(1):61–67. doi:10.1634/theoncologist.2013-0290.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlation: uses in assessing rater reliability. *Physiological Bulletin*. 86:420–428.
- Starling, M.S., Kandel, Z., Haile, L., Simmons, R.G. (2018). User profile and preferences in fertility apps for preventing pregnancy: an exploratory pilot study. *Mhealth*, 4:21. doi:10.21037/mhealth.2018.06.02

- Tefertiller, A. (2017). Like us on Facebook: social capital, opinion leadership, and social media word-of-mouth for promoting cultural goods. *The Journal of Social Media in Society*, 7(2): 274-296.
- Walker, M., King, G., Hartman, L. (2018). Exploring the potential of social media platforms as data collection methods for accessing and understanding experiences of youth with disabilities: a narrative review. *The Journal of Social Media in Society*, 7(2): 43-68.
- Whitaker, C., Stevelink, S., & Fear, N. (2017). The use of Facebook in recruiting participants for health research purposes: a systematic review. *J Med Internet Res*, 19(8): e290. Published 2017 Aug 28. doi:10.2196/jmir.7071
- Wilkinson, D., Thelwall, M. (2011). Researching personal information on the public web: methods and ethics. *Social Science Computer Review*. 29(4): 387–401
- Wilson, R.E., Gosling, S.D., & Graham, L.T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science*, 7(3): 203-20.
- Youyou, W., Kosinski, M., Stillwell, D. (2015) Computer-based personality judgments are more accurate than those made by humans. *Proc Natl Acad Sci USA*, 112: 1036–40.
- Zhang, X. (2018). Social media popularity and election results: a study of the 2016 Taiwanese general election. *PLoS ONE*. 13(11): e0208190. doi: 10.1371/journal.pone.0208190
- Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. *Ethics Inf Technol*. 12(4):313–325. doi: 10.1007/s10676-010-9227-5.

### **Funding and Acknowledgements**

The authors declare no funding sources or conflicts of interest.