Using Social Media Data
to Understand Changes in Gender Norms

Guide

USAID FROM THE AMERICAN PEOPLE PEPFAR MEASURE Evaluation

Using Social Media Data to Understand Changes in Gender Norms

Guide

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1 919-445-9350
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. MS-18-147
ISBN: 978-1-64232-089-3

ACKNOWLEDGMENTS

We thank the United States Agency for International Development (USAID) and the United States President's Emergency Plan for AIDS Relief (PEPFAR) for their support of this work.

This guidance was based on research conducted by the USAID- and PEPFAR-funded MEASURE Evaluation. A special thank you to John Spencer and Becky Wilkes for their efforts in identifying data sources and appropriate abstraction methods and working with third-party vendors as well as obtaining and cleaning data. Brittany Iskarpatyoti and Heather Biehl were responsible for qualitative coding. Ms. Iskarpatyoti conducted qualitative and quantitative analysis. Thank you to Jessica Fehringer and Carolina Mejia for their formative contributions and review. Thank you to MEASURE Evaluation's knowledge management team for editorial, design, and production services.

Suggested citation:

MEASURE Evaluation. (2018). Using Social Media Data to Understand Changes in Gender Norms: Guidance. Chapel Hill, NC, USA: MEASURE Evaluation.

CONTENTS

TABLES

FIGURES

ABBREVIATIONS

AI	artificial intelligence
API	application programming interface
GBV	gender-based violence
IP	Internet protocol
IRB	institutional review board
NLP	natural language processing

BACKGROUND

Gender and Social Norms

As technology becomes increasingly accessible around the globe, social media has emerged as a tool to help researchers and health workers and administrators understand a broad array of health issues. It has been harnessed to surveil infectious disease, forecast asthma exacerbations, monitor mental health following traumatic events, and gauge attitudes and beliefs about medical conditions, products, and treatments. Gender norms—the socialized expectations about women, men, boys, and girls and the power dynamics between them—constitute one frontier for social media research. Although many national and international programs seek to change harmful gender norms to achieve equities in health, few collect data on societal attitudes. Because social media provides people with an opportunity to share aspects of their lives, it may have the potential to provide insights into attitudinal and behavioral aspects of gender inequality and to capture information that is difficult and costly to obtain through regular surveys.

A MEASURE Evaluation study explored the feasibility of using large social media data sets to track changes in attitudes toward, and gender norms regarding, sexual relationships between younger women and older men and gender-based violence (GBV) against women and girls in sub-Saharan Africa. After reviewing several possible social media platforms, including Facebook, Snapchat, and Instagram, we selected Twitter for its data availability and ease of access. We assessed existing methods for extracting data from Twitter and analyzing metrics to understand the challenges and limitations of using social media data. We also reviewed ethics and data security considerations with each method.

Purpose of the Guidance

The purpose of this document is to provide guidance on collecting, analyzing, and interpreting Twitter data on gender norms. We will discuss when social media can be useful in monitoring, evaluation, and research; what data are available; and methodological challenges including generalization, biases, protecting individual privacy, and considering ethical implications.

Intended Users

This guidance is written for monitoring and evaluation officers or data users with some background in Microsoft Excel.

METHODS

Social Media Platforms

This guidance is based on research conducted on gender norms.[1] We explored a variety of social media platforms to assess which would have available and accessible data on gender norms from 10 sub-Saharan African counties: Kenya, Lesotho, Malawi, Mozambique, South Africa, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe. Certain platforms lacked a strong user base (e.g., Snapchat) or applicable content for analysis (e.g., Instagram).

At the time of this research, Facebook user protections prevented the abstraction of data from large numbers. Because of the reach, applicability, and accessibility of its data, Twitter was selected as our social media source. This guidance looks at methods for obtaining and cleaning Twitter data. Discussion of analysis, reporting, and ethics may be generalizable to other social media platforms.

Getting the Data

Twitter data can be collected either retrospectively or prospectively. In the early days of Twitter, it was possible to search and obtain historical tweets by writing a short script in Python or R (or another language) that queried Twitter servers. However, at the time of this study, historical tweets were available only through third-party vendors.[2] Unfortunately, obtaining multiple months' worth of tweets that use a specific hashtag or are on a specific topic can be expensive. If historical data are needed, it will be necessary to contact a vendor authorized to resell it. Several vendors sell data; we relied on Crimson Hexagon and Union Metrics.

Real-time tweets, however, can be obtained manually, using several different approaches. The one with the most flexibility and power is writing a script in R, Python, or another language, which makes possible obtaining data, analyzing it, and producing data products all in one script. The downside to this approach is that it requires knowledge of and experience with the programming language and obtaining a Twitter developer's account allowing access to Twitter's API. Those who don't know a programming language or don't need to integrate the collection, analysis, and output of the data have a couple of other options, such as Microsoft Flow.

We will discuss three methods, at three price points, for collecting and analyzing social media data. Table 1 provides an overview of the three methods we will review.

---

[1] Iskarpatyoti, B., Biehl, H., & Spencer, J. (2018). Using Social Media Data to Understand Changes in Gender Norms. Chapel Hill, NC, USA: MEASURE Evaluation. https://www.measureevaluation.org/resources/publications/tr-18-295/

[2] As this document went to press, Twitter introduced a developer account that once again makes it possible to obtain some retrospective data. Data access changes regularly; it is important to review current rules and regulations.

Table 1. Comparison of three data collection platforms for Twitter data

	Crimson Hexagon	Union Metrics	Microsoft Flow
Program access cost	$$$$	$$	$
Analysis cost	$$	$$	$$$
Number of tweets	10,000 per query	Unlimited	100 results/check (1 check per 1–5 minutes, depending on package)
Direction	Retrospective	Retrospective + prospective	Prospective

Crimson Hexagon is an AI-powered consumer insights company. Its online data library consists of more than 1 trillion posts and includes documents from social networks such as Twitter and Facebook along with blogs, forums, and news sites. Pew Research uses Crimson Hexagon's social media analysis platform to analyze media coverage and discourse.

Union Metrics, like Crimson Hexagon, is a commercial software platform that extracts social media data. It has fewer built-in analytics, but data can be collected retrospectively or prospectively.

Microsoft Flow is a workflow automation tool that integrates with some versions of Office 365 and online cloud storage. Flow makes it possible to set up a series of steps to regularly query Twitter for tweets that meet specific criteria and then output them to an Excel spreadsheet, a .csv file, or another data file. Although Flow offers powerful automation abilities, it is not available for all versions of Office 365.

Scraping using tailored code is an option we won't spend much time talking about here. It requires the user to be skilled in R or Python and to constantly monitor the scraped data to ensure that updated Twitter privacy policies or procedures don't introduce interruptions of scraping.

Overview of Data Collection and Data Available

It is possible to get metadata about a tweet in addition to the text. This metadata includes the Twitter user's name, the date of the tweet, the number of retweets, the number of likes, how many followers the user has, and how many people he or she follows. Some data vendors include other proprietary information, such as how influential a tweet might have been, the user's gender, and sentiment analysis. Some even provide a customized analytics dashboard.

When collecting real-time tweets, only variables passed by the Twitter API will be available. The specific variables are subject to change by Twitter but will include basic information such as tweet text, Twitter username, date, time, number of retweets, and number of likes. Consult Twitter documentation for the latest list of variables that are available. Table 2 provides an overview of the types of data and analytics provided by each platform or method.

Table 2. Comparison of data and analytics provided by each platform

	Crimson Hexagon	Union Metrics	Microsoft Flow
Metadata			
Tweet ID	✓	✓	✓
Tweet URL	✓	✓	✓
User ID	✓	✓	✓
User location	✓	✓	✓
Country	✓	✓	✓
Tweet text	✓	✓	✓
Time	✓	✓	✓
Retweet ID		✓	✓
Reply to ID		✓	✓
Impressions		✓	
Likes		✓	✓
User number posts	✓		✓
User followers	✓		✓
User following	✓		✓
Quantitative Analytics			
Average tweets/day		✓	
Type of tweet breakdown		✓	
Retweet rate		✓	
User gender	✓		
Tweets per User (contributors)		✓	
Top Contributors		✓	
Potential reach		✓	
Geography	✓	With additional subscription	
Qualitative Analytics			
Sentiment (positive, negative, neutral)	✓	With additional subscription	
Emotion	✓		
Top Hashtags		✓	

Data Cleaning

Depending on your method for collecting data or number of search strings, you may have to merge and clean your data. This is particularly true with data from Microsoft Flow, because it is the rawest.

Merging Data

You may be limited as to how many search terms can be included in a query to Twitter. You may need to break up your search string into two separate searches. These will run in parallel, collecting tweets in separate files (such as a .csv or Excel file). When collection is finished, however, they should be merged and duplicates deleted. To do this, you can use the delete duplicate function in Excel or the query wizards in Access. We were able to get unique values from each query (in other words, values in 1 but not 2, and values in 2 but not 1) and then paste the results back together in Excel.

Cleaning Data

Because data from Flow are taken directly from Twitter and are relatively unfiltered for use, they require much more cleaning than data from Crimson Hexagon or Union Metrics.

1. Location: A user may turn on the "Tweet with a location" setting, and the user's location will be derived from either the app or the IP address. This location (often a latitude and longitude) is stored in the metadata and visible on the tweet. However, not all users have this setting activated, resulting in missing data. Alternatively, location data may be entered by the user as a string variable—but these data are not clean and may be inaccurate. Although we can't account for accuracy, we can at least exclude data that do not fit our inclusion criteria. Of course, that is easier said than done. Combing through locations takes time. People may switch city and province or country (as in Pretoria, South Africa versus South Africa, Pretoria), for example, or use acronyms (USA versus United States). These locations are easily interpretable, but others are less so.

    - Nicknames, such as the Big Apple, the Garden City, the Heart of Africa. Some of these are easily identifiable with specific locations, whereas others may apply to multiple places.
    - Ambiguous, such as does "East London" refer to East London in the UK or East London in South Africa?
    - Two or more locations: (South Africa and USA)
    - Special characters (♦♦✡♦♦)

    Determine how conservative you want to be about including or excluding criteria and clean appropriately. Depending on how much data you have, you may want to do this for all data or for a random sample to understand how relevant your data are or from where they are proportionately coming.

2. Relevancy: The more complex your search string, the more likely you are to end up with irrelevant data. For example, if you're interested in tweets about 16 Days of Activism against Gender-Based Violence, you might search specific hashtags that correspond with online activism campaigns (#16days, #orangetheworld). A search of this type will give you high rates of relevance. If you expand to look more generally at the conversation around GBV by including terms such as "beat," you may find tweets about song beats or a specific brand of headphone. The specificity of your search will determine the depth, breadth, and relevance of the data you get.

Unfortunately, the only way to assess the relevance of your data is often to examine it in detail. It may be useful to examine a subsample and assess what percentage is irrelevant to your research question. If that percentage is unreasonably high, you may want to rework your search terms.

Data Analysis

Quantitative Data Analysis

Qualitative data from Twitter are best used for descriptive analysis. There are too many limitations in who uses and produces Twitter data to recommend greater statistical analysis.

Time Frequency

Both Crimson Hexagon and Union Metrics provide dashboards with frequency of tweets by day. However, after cleaning or if you want to look at a subset, you may need to produce a new frequency analysis. If you're using Microsoft Flow, you will need to do this yourself. It can be difficult to get information regarding frequency without extracting and reformatting as a number, because time and date information are collected as string text in Excel. To analyze frequency by date, you will need to extract the date information from the time element. The same is true if you want to look at frequency by hour of the day. Once you have extracted the correct information, you can create a quick visualization of tweets by time. These can be disaggregated by any other variable, such as location, type of tweet, or gender (if provided by the vendor).

Location

As discussed under "Data Cleaning," location data may be missing or may be entered by the user as a string variable. This increases the "messiness" of the data. Some online data analytic software companies, including Crimson Hexagon and Union Metrics, provide cleaner location variables at levels that are more granular (province versus country) and can filter for data from only those areas specified in your search. If you are scraping from Twitter yourself, you will have to filter out excluded locations by hand and may have to be more generous with your level of analysis (for instance, you may not be able to analyze beyond country level). Once the data are cleaned, you can do a frequency analysis by location.

Impressions

Crimson Hexagon, Union Metrics, and Flow collect information on number of followers for each Twitter user. With this information you can understand the potential impression or reach of a single tweet, a group of tweets, or a conversation over time. You can also assess how often a single tweet has been retweeted.

Tweet Type

The various types of tweets can be looked at to understand how a conversation is occurring or evolving. Regular tweets are messages containing text, photos, a GIF, and/or video. A reply is the response of one user to another. A quote tweet allows a user to tweet someone else's tweet with an added comment. A retweet is a reposting of another user's tweet. Crimson Hexagon can provide any type of tweet or filter for a specific type only. For the purposes of our study, and owing to costs, we excluded retweets from Crimson Hexagon's data collection. With Union Metrics, we could collect all types of tweets to understand how retweeting and quoting existing tweets affected the conversation.

Flow has no specific "type of tweet" indicator, although whether the tweet is original content, a retweet, or quote can be determined by assessing if "original tweet" information is present.

Qualitative Data Analysis

Sentiment

Crimson Hexagon—and Union Metrics with an additional subscription—provide sentiment (positive, negative, or neutral) in their analytics packages. Crimson Hexagon's technology uses natural language processing (NLP) to do sentiment analysis of text-based conversations on a large scale, quickly. However, depending on your research topic, you may find these assignments to be of limited accuracy. For example, a tweet may be negative in language but positive toward your research topic. We encountered this in our research on gender-based violence. A tweet might say, "I can't believe these losers who think they can hit a woman and get away with it." The language might be coded by machine as negative, but we would code the tweet as positive against GBV. See the companion report for greater detail: https://www.measureevaluation.org/resources/publications/tr-18-295/

The most accurate way to measure sentiment is by a human. The manual assignment of sentiment is necessary for data collected through Microsoft Flow. For data that come with precoded sentiment, we recommend that you manually assign sentiment in a subsample of tweets to determine whether the full sample should be hand-coded.

Language Metrics

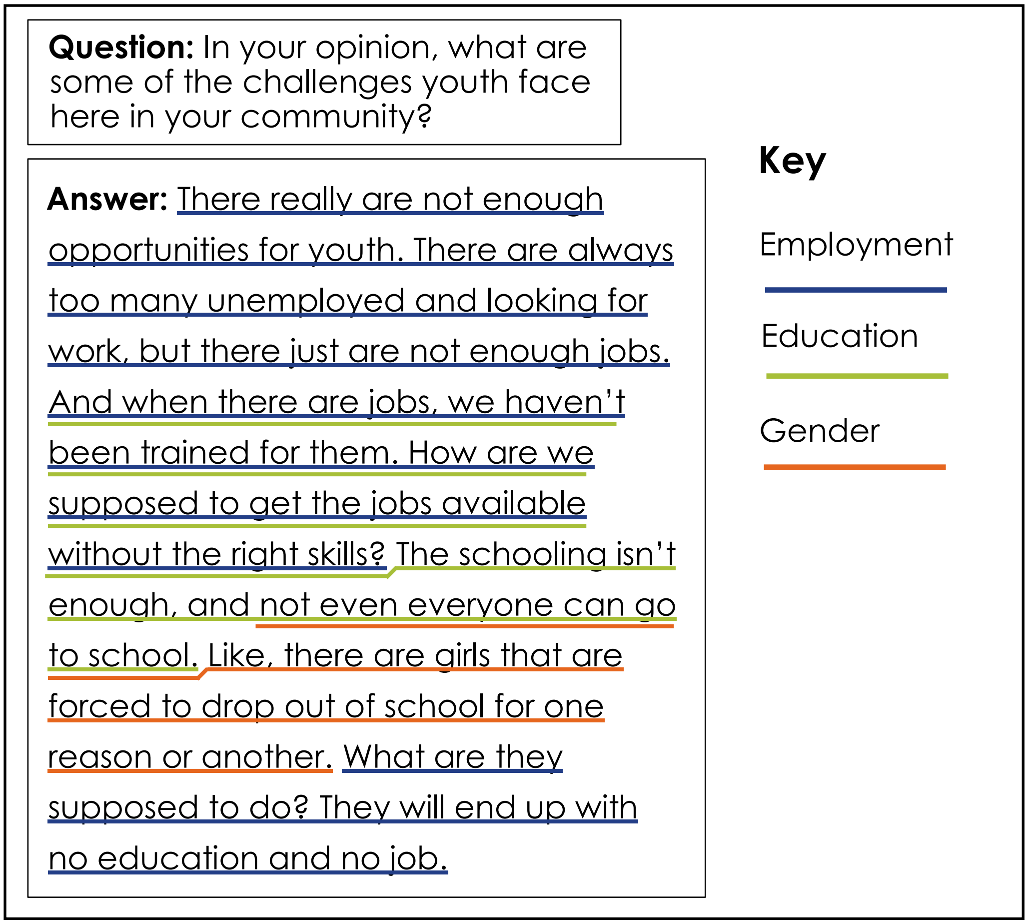
One quick way to analyze or determine emerging themes is to generate a word cloud from the content of your tweets. A quick online search will produce several good free options for doing this. If you want a more quantitative assessment, some of them will also produce a table showing how many times each word is used. Word clouds are great for quick visualizations of possible themes to inform the qualitative component, but they do not necessarily show the drivers and connections behind the themes.

Themes

Qualitative thematic analysis of tweets follows many of the same rules that traditional qualitative analysis does. Themes may be developed inductively (from the tweets themselves) or deductively (from previously determined theory). You may find that themes must be iteratively reassessed as you are going through the data. We will not go into the basics of traditional qualitative thematic analysis here; rather, we will focus on how qualitative analysis of an interview or a focus group discussion differs from qualitative analysis of a tweet.

The greatest difference between the two is the length of the unit of analysis. Analysis of narratives, interviews, and focus group discussions is intended to break up lengthy prose into consecutive, overlapping, and staggered segments associated with various themes. Figure 1 provides a transcript of a fictional interview about the challenges youth face in a community to understand how gender intersects with those challenges. Themes have been highlighted and coded.

Figure 1. Traditional qualitative analysis



Overlapping thematic analysis lets researchers think about the connections between themes even with just a few data collection moments (individual interviews or focus group discussions). However, a maximum tweet length of 280 characters (as of September 2017) leaves little room for segmented analysis. Each tweet essentially acts as a segment and therefore either is or is not coded to a specific theme. Figure 2 illustrates a collection of fictional tweets on the same topic—challenges youth face in a community—and how coding might occur.

Figure 2. Qualitative analysis of tweets

Tweet	Employment	Education	Gender
These days man, there's no jobs. Where'd da jobs go!?	✓		
You see politicians complain that youth are lazy. Then you ask these kids and they tell you there are no jobs. Who's right? #nosense	✓		
Looking for a job—maaaan it's like I didn't even go to school!	✓	✓	
Once a month, many girls are forced to stay home from school. #periodprejudice #genderinequality		✓	✓

HAHAHA! This kid thinks he can just walk in with his basic education and be a boss! #entitled #notraining	✓	✓	
Jobs are being snatched up by only those who can afford fancy education. What's the system doing for the rest of us?	✓	✓	

This type of analysis requires many more tweets to get a good picture of overlapping themes and make connections. Here the researcher used themes developed deductively from previous research to guide the search string and analysis. But if the researcher is interested only in the intersection of gender with these themes, only one tweet is relevant.

"Participants" are not responding to a researcher-developed question or inquiry here. The types of data you analyze will be limited by your search string, which may need to be iteratively developed to fully capture the information you want. Also, you may not be able to answer certain questions because people are not openly discussing them on social media. For example, a Twitter study looking at GBV may be able to describe the conversation around domestic violence but unable to look at individual experiences or personal attitudes.

The data are also less formal; it is not unusual for tweets to include expletives, emojis, jargon, and slang, and researchers have no way to follow up with the tweeter for clarification. Analysis can be improved by having a social media user on the research team who is familiar with the terminology and use norms—essentially, someone who is fluent in social media language.

This type of analysis can easily occur in Excel rather than by using potentially costly qualitative analysis software.

Data Ethics

The ethical implications around the use of social media data are only beginning to be understood. A full discussion of these issues is beyond the scope of this document, but users of this type of data should fully consider ethical and privacy implications.

One issue is whether an activity qualifies as human-subjects research within an institutional review board (IRB), and if so, what type of review is necessary. Social media data are public; information is identifiable; and information gathering requires no interaction with the person who posted it online. Therefore, social media research often does not constitute human-subjects research and may be exempt from IRB review.[3] However, social media website policies and controversy regarding specific studies have raised new questions about whether such research should continue to receive IRB approval.[4] Some IRBs now consider social media privacy concerns, website privacy guidelines, and legal issues. However, this is not universal, and exemption from review presents researchers with difficult decisions about ethical constraints.

Applying ethical principles to online research has provided a challenge for researchers. For example, is consent necessary for certain topics? One possibility is to obtain consent from users whose data are being analyzed. However, this generates a number of concerns. Participants over the age of 18 years may be able to give consent after being directed to a study site, but those under the age of 18 years would require parental

[3] Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking, 16*(9), 708–713.

[4] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks,* 30(4), 330–342.

consent if following traditional ethical procedures. Given the quantity of data involved, consent can make research time-consuming at best and impossible at worst. Furthermore, with the use of big data becoming more popular for research, inundating users with requests for consent may reduce the functionality of social media and change how it is used.

Another question involves reporting social media research. Data themselves can be identifying. A single unique tweet can be entered into a search engine and link to the poster's profile, thus identifying him or her. When asked, individuals who contribute to online discussion boards said they had no problem with aggregated information (e.g., statistics or trend analyses) being used without consent. However, views diverged on the use of more-identifiable information, such as direct quotations.[5] Although there is no gold standard for reporting social media data, one alternative is to use aggregated quotations. Several tweets around the same topic can be brought "together to maintain the essence of meaning but rendering it impossible to identify any individual poster." Aggregate quotations may lose some nuance and introduce bias in analysis and reporting, but when balanced against confidentiality, particularly around sensitive subjects, they are a prudent choice.

There has been some discussion about triangulating social media data with other big data sources. However, the ethical implications of obtaining information beyond what is on the Twitter profile (such as individuals' movements, email addresses, home addresses, etc.) have not been fully explored.

Other implications include the handling of tweets that present criminal situations, such as child abuse and GBV. While researchers may intend to look only at social attitudes and norms around these topics, they may find that some users have disclosed incidents in a tweet. What is their ethical responsibility to follow up or report this information?

Just like social media sites, IRBs and ethical review boards frequently change their policies and procedures. Before undertaking any type of social media research, it is important to consult the current policies of your institution and social media platform.

[5] Bond, C. S., Ahmed, O. H., Hind, M., Thomas, B., & Hewitt-Taylor, J. (2013). The conceptual and practical ethical dilemmas of using health discussion board posts as research data. *Journal of Medical Internet Research, 15*(6).

CONSIDERATIONS AND RECOMMENDATIONS

Policies change frequently, and previous methods of collecting and analyzing data may no longer work. Furthermore, what was once not possible is now possible. Thus, it is impractical to write a step-by-step guide to getting and using Twitter data (or any other social media data) that may well be obsolete the minute they are published. One can, however, consider some guiding principles when undertaking this type of research.

The following types of social media data may be useful in understanding gender norms:

- Targeted social media campaigns. As social and behavior change communication campaigns include more social media components, data from Twitter or other platforms may help in understanding access, use, and reaction to these campaigns.
- Hashtag movements and advocacy. International days of health and social movements are increasingly accompanied by hashtags. These hashtags are easy to search for by any data collection method and will return a high rate of relevance.
- Areas where social media use is high. Because gender norms are time- and place-dependent, they may be shaped by and reflected on social media as the Internet grows even more. For example, we looked at intergenerational transactional relationships (companion report citation when available) in South Africa and found that the term "blesser" referred specifically to such relationships and was a term originated and popularized by Twitter users. Therefore, social media is an ideal source of data for understanding "blesser" culture.
- As a complement to other data.
- When exploring trends in the popularity of topics, events, news, or public figures.

The following types of social media data may **not** be useful in understanding gender norms:

- Broader gender norms. The more specific the norm, the easier and more relevant the search will be. More-complex norms, such as GBV, may be difficult to distill into search terms without falsely excluding or improperly including tweets.
- Areas where social media use is limited. Access to computers, the Internet, and social media may be limited in certain settings. For example, among the 10 African countries we included in our search, most tweets came from South Africa (75%–97%) and few from Kenya (2%–24%). Most of the other countries were not represented, so analysis of social media data was not possible for them. Additionally, South African tweets were produced primarily in affluent provinces. (companion report citation when available).
- Trends in specific groups. Data from social media are not generalizable to the broader society. Certain populations—such as the very old and the very young—may not use social media. Certain groups and classes may lack access to the Internet for economic reasons or be excluded by social norms and expectations. For example, although social media use is somewhat common in certain parts of South Africa, a woman in rural South Africa may share a cell phone with her husband and not have access to it for her own use.
- Incidence of disease or health outcomes. Such data are not commonly disclosed on social media, and when they are, those who disclose are not representative of the whole population. There has been discussion about triangulating social media data with other big data sources to look at these trends, but doing so has logistical and ethical implications.

Some ethical considerations follow:

- Researchers should collect only the data essential for answering the research question and should present those data carefully to avoid participant identification.
- How social media research is reviewed by IRBs and how ethical principles are applied to such research is unclear and differs by institution and state.
- Because some IRBs are slow to adapt, there may be ethical issues not considered in IRB review that researchers should take into account. As researchers, we have a responsibility to apply the highest ethical standards possible.
- Researchers should understand the risks of (and avoid) quoting directly from text in presenting their data.

Other considerations and limitations in social media research follow:

- The more complex your search string, the more likely you are to end up with irrelevant data.
- Data quality varies, and there is no standard for assessing and accounting for it in social media research.
- Sentiment or gender identified by computer algorithms may be inaccurate. Although these technologies are continually improving, sentiment should still be reviewed against the research question. The language of a tweet may appear positive even when the tweet is negative toward the research topic.
- Data are created by social media users and thus not filtered through the researcher's lens. However, their inclusion is limited by researcher-developed search strings. Where one bias is eliminated, another often emerges.
- Social media has a language all its own. It is important that someone on your research team be fluent in this language.
- Not everyone uses Twitter, so it is not generalizable to the general population. Furthermore, some users make their accounts private, which may mask key findings.
- Bots are increasingly used to generate tweets, retweets, and fake accounts. Without an intensive forensic effort, it can be difficult to determine whether followers are real people or tweets are from real people.
- Results can also be skewed by the presence of "trolls," who can post content to generate controversy and sometimes harass other users. This can lead to bias in the results because of an abundance of tweets reflecting inaccurate information about the topic of interest or because people don't want to agitate and thus stay silent.

Social media research truly represents a new frontier—one that is wild and uncultivated. It has not replaced and cannot replace traditional forms of data collection for research, monitoring, and evaluation, but as it evolves, it may provide a data triangulation point for understanding how social communities talk about certain issues and norms.

MEASURE Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1-919-445-9350
measure@unc.edu
www.measureevaluation.org

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. MS-18-147 ISBN: 978-1-64232-089-3