IMAGES OF POST-MASTECTOMY SCARRING ON INSTAGRAM: AN EVALUATION OF CONTENT MODERATION AND THE IMPLICATIONS FOR PUBLIC HEALTH

Alice Witt
Faculty of Law, Digital Media Research Centre, Queensland University of Technology

Yonaira Rivera
School of Communication and Information, Rutgers University

Online platforms play a central role in moderating aspects of our digital lives, including the types of content that we see and how and when we see it.[1] Content moderation refers to the regulatory processes through which platform executives and their moderators set, maintain and enforce the bounds of 'appropriate' user behaviours.[2] Platforms make decisions about the appropriateness of content based on a range of factors, including platform-specific policies, jurisdiction and social and cultural issues.[3] Content is, however, moderated within a 'black box' that obscures the internal workings of platform governance from external scrutiny.[4] Transparency deficits like this matter because online social networks are not only where individual expression through content can be influenced or controlled,[5] but also places that can shape how people represent, share and live different experiences.[6] The direct interventions that platforms make around content can also have far-reaching public health consequences: some can be transformative, such as when vulnerable healthcare consumers have greater access to real-time support and information,[7] while others can have detrimental impacts on health and well-being, including the uptake of potentially harmful behaviours due to health misinformation.[8]

On Instagram, the moderation of women's bodies is particularly controversial, with users and other stakeholders claiming that content (especially nonnormative expression) is arbitrarily removed from the platform.[9] Recent empirical research suggests these claims might not be unfounded: an inconsistent trend of moderation has been identified across

explicitly prohibited and explicitly allowed depictions of apparent female forms,[10] as well as images of women's bodies that do not appear to violate content policies.[2] We aim to build upon this research by empirically investigating the moderation of images depicting post-mastectomy scaring on Instagram, one of few explicitly allowed categories of content under the platform's Community Guidelines,[11] by skin tone darkness and the presence or absence of tattoos. We explore removal by skin tone given the growing number of BIPOC (Black, Indigenous and People of Colour) communities on Instagram.[12] The presence or absence of tattooing is also a focus of our investigation because tattoos may be used to cover post-mastectomy scars and/or recreate nipples that could not be spared. We argue that the moderation of this subject matter is important because images of post-mastectomy scarring are inextricably linked to complex issues among breast cancer survivors: it can, for instance, be educational, help to raise awareness about breast cancer in different communities and connect users who are actively looking to find a shared community.[13] Mastectomies as a result of breast cancer can also have important ramifications for women's self-image and identity worldwide.[14,15] In this paper, we report on our pilot study for this research, which we will undertake on a larger scale in late 2020.

Methodology

Our methodology is based on Witt, Suzor and Huggins' (2019) input/output method of black box analytics that identifies how discrete inputs into a system produce certain outputs.[2] Input refers to individual images, while output pertains to the outcome of content moderation (i.e., whether an image is removed or not). Images for our pilot study were programmatically collected through the Australia-based Digital Observatory at QUT's Digital Media Research Centre.[i] Specifically, automated tools scraped the last 20 images from 16 mastectomy-related hashtags every six hours on an ongoing basis. Our hashtags, which we deliberately selected by searching Instagram and news articles for mastectomy content, are varied: some are mastectomy-specific, including #breastcancersurvivor and #mastectomy, while others pertain to breast tattoos (e.g.,#mastectomytattoo), and some are generally cancer-related (e.g.,#lifeaftercancer).

Once the sample was generated for manual coding (n = 4,410), we undertook content analysis to code images based on the 11-point Massey and Martin Skin Colour Scale (MMSCS) (where zero = albinism and 10 = the darkest possible skin).[16] We also coded for the presence or absence of tattooing (i.e., tattoo = 1, no tattoo = 0), and for content removal (i.e., removed = 1, not removed = 0) based on programmatic testing one month after initial data collection, providing a discrete output for every coded input. We excluded close-ups of women's faces and images in which breasts were covered. This methodology ultimately enables us to identify potential true negatives (explicitly allowed images that were not removed), and false positives (explicitly allowed images that were removed), for analysis by apparent skin tone and by tattooing. Our final coded sample comprises 74 images depicting post-mastectomy scarring with coding agreement ranging from 78% (MMSCS) to 100% (presence/absence of tattoo). Descriptive statistics and quantitative analyses were conducted using STATA v15.

Pilot Study Results and Implications

Of the collected images, 44 depicted post-mastectomy scarring alone and 30 included breast tattooing. Given the small sample size of the pilot data, we were not able to detect statistically significant differences in images removed versus not removed. However, we believe there is a strong case for further largescale investigation. First, despite Instagram explicitly allowing depictions of post-mastectomy scarring, we identified 31 potential false positives (41.9%). Second, preliminary data suggest differences in removal rates by presence or absence of tattoos: 67.7% of images without tattoos were removed, compared to only 32.3% of images with tattoos ($n=10$)($\chi^2=2.03$, $p=0.15$). Lastly, a logistic regression suggests that the odds of an image being removed increase by a factor of 1.21 for every unit increase in skin tone darkness ($p=0.31$). Moving forward, given that most images in our pilot study depicted white women, we will diversify our selected hashtags to attempt to collect more images of women with darker skin tones. We will also extend the data collection timeframe to collect a larger sample of images for manual coding.

Our further in-depth investigation has the potential to provide valuable insights about how images of post-mastectomy scarring are moderated on Instagram in practice. In general, women who undergo mastectomies face unique experiences related to body image, with some expressing feelings of empowerment in light of this procedure.[17] Content moderation that is consistent with Instagram's policies may provide additional support for the effective use of this platform to promote body positivity,[18] which can be leveraged for interventions that share the lived experiences of breast cancer survivors. However, evidence of potentially arbitrary moderation may have a wide range of negative effects on the well-being of women seeking to share their experiences with, or learn from the experiences of, others who have undergone mastectomies. For example, content removal might exacerbate some women's negative body and self-image,[11] or contribute to social media platforms, including those who have been victims of body shaming. While we cannot draw definitive conclusions about whom (i.e., the platform or individual users) removed content and why (e.g., user choice, platform policy reasons) from outside the black box, our research will make valuable contributions to the growing body of literature on the moderation of female forms on Instagram. It will also help to better assess the efficacy and ethics of using black box methods to examine content moderation at scale and contribute to ongoing debates around the transparency and accountability of platform governance more broadly. We look forward to furthering this research in late 2020.

References

1. Leaver T, Highfield T, Abidin C. *Instagram: Visual Social Media Cultures*. John Wiley & Sons; 2020.

2. Witt A, Suzor N, Huggins A. The rule of law on Instagram: An evaluation of the moderation of images depicting women's bodies. *UNSWLJ*. 2019; *42(2)*:557.

3. Gillespie T. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Connecticut, US: Yale University Press; 2018.

4. Roberts ST. *Behind the screen: Content moderation in the shadows of social media*. Connecticut, US: Yale University Press; 2019.

5. Suzor NP. *Lawless: The secret rules that govern our digital lives*. Cambridge, UK: Cambridge University Press; 2019.

6. Locatelli E. Images of breastfeeding on Instagram: self-representation, publicness, and privacy management. *Social Media+ Society*. 2017;*3*(2): 2056305117707190.

7. Gurrieri L, Drenten J. Visual storytelling and vulnerable health care consumers: normalising practices and social support through Instagram. *J Services Market*. 2019; 33(6): 702-720.

8. Rivera YM, Moran M, Thrul J, Smith KC. The role of culture and cognitive heuristics in assessing the credibility of cancer (mis)information on Facebook among U.S. Latinos. *AJPH*. 2020; Under Review.

9. West SM. Facebook's Guide to Being a Lady. OnlineCensorship.org. https://www.onlinecensorship.org/en/news-and-analysis/15 Published 2015.

10. Witt A, *The Rule of Law in Platform Governance: The Moderation of Images Depicting Women's Bodies on Instagram*. Doctoral Thesis: QUT; 2020.

11. Instagram. *Community Guidelines*. https://help.instagram.com/477434105621119 Published 2020.

12. Garcia SE. Where Did BIPOC Come From? The New York Times. https://www.nytimes.com/article/what-is-bipoc.html Published 2020.

13. Koskan, A., Klasko, L., Davis, S. N., Gwede, C. K., Wells, K. J., Kumar, A., ... & Meade, C. D. (2014). Use and taxonomy of social media in cancer-related research: a systematic review. *American journal of public health*, *104*(7), e20-e37

14. Koçan S, Gürsoy A. Body image of women with breast cancer after mastectomy: a qualitative research. *J Breast Heal*. 2016;*12*(4):145.

15. Potter S, Conroy EJ, Cutress RI, Jain A. Short-term safety outcomes of mastectomy and immediate implant-based breast reconstruction. *Lancet Onco*. 2019;*20*(2):254-266.

16. Massey, Douglas S, Martin, Jennifer A. The NIS Skin Color Scale. https://nis.princeton.edu/downloads/nis-skin-color-scale.pdf Published 2003

17. La J, Jackson S, Shaw R. 'Flat and fabulous': women's breast reconstruction refusals post-mastectomy and the negotiation of normative femininity. *J Gend Studies*. 2019;*28*(5):603-616.

18. Cohen R, Irwin L, Newton-John T, Slater A. #bodypositivity: A content analysis of body positive accounts on Instagram. *Body image*. 2019;*29*:47-57.

---

[i] We thank Professor Nicolas Suzor for his assistance with data collection. Any errors are our own.