



# Data science, social media, and HPV vaccines: the state-of-the-art and future challenges

Adam G. Dunn

Centre for Health Informatics

Australian Institute of Health Innovation

Macquarie University

@adamgdunn

for the 2018 National HPV Vaccination Roundtable Meeting (Atlanta)



**MACQUARIE**  
University

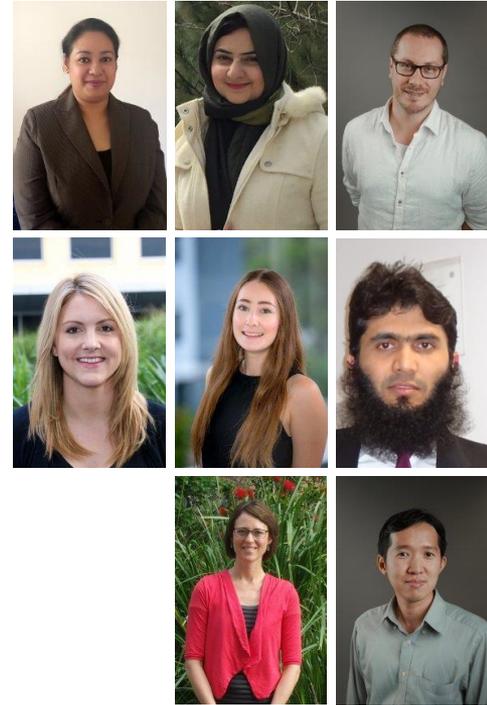
**Competing interests:** No financial conflicts of interest.

**Funding:** **NHMRC Project APP1128968**  
Unrelated funding from NHMRC and AHRQ

**Team:** Samia Amin, Rabia Bashir; **Adam G. Dunn;**  
**Amalie Dyda; Paige Newman; Zubair Shah;**  
**Maryke Steffens; Didi Surian**  
(team includes computer science, data science and  
machine learning, software engineering, epidemiology  
and public health, journalism, and clinical medicine)

**Team alumni:** Xujuan Zhou, Diana Arachi, Smriti Raichand

**Collaborators:** **Julie Leask, Kenneth D. Mandl, Aditi Dey,**  
**Enrico Coiera,** Gilla Shapiro, Margaret Kelaher,  
and many more...



Analyses of HPV vaccine information on social media:

YouTube: Ache 2008

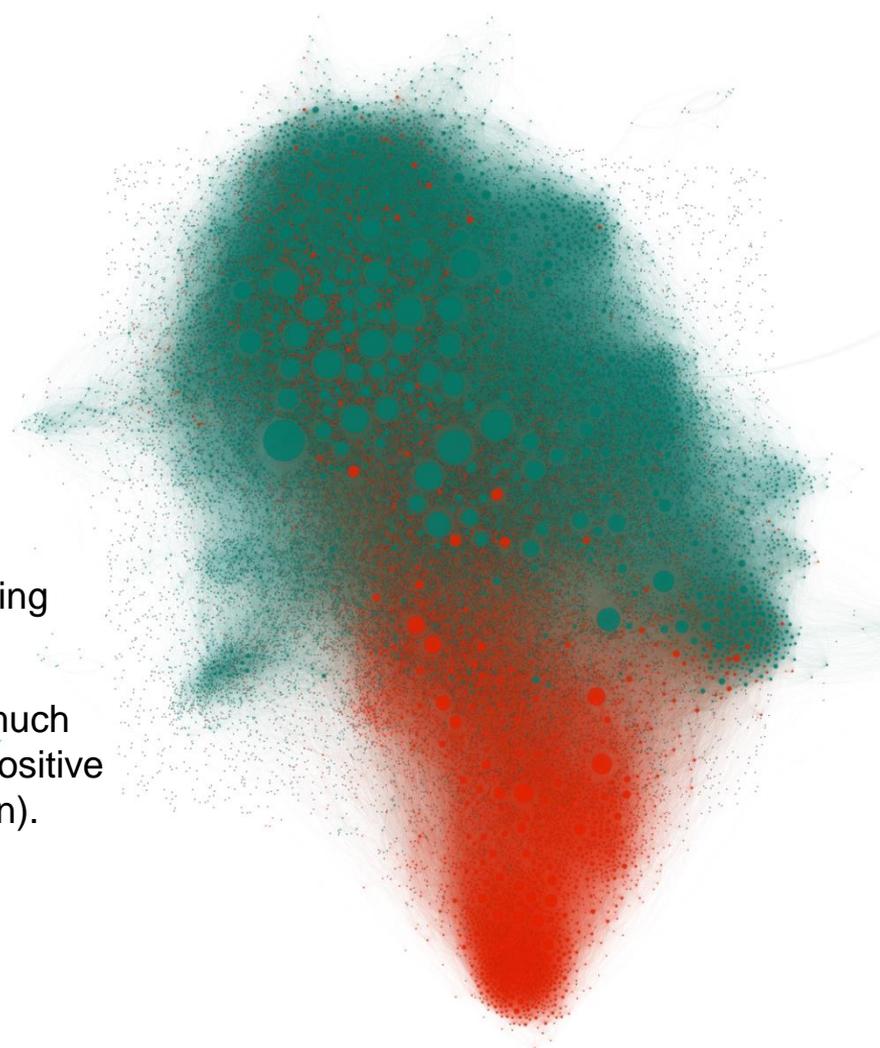
MySpace: Keelan 2010

Analysis of vaccine information on Twitter

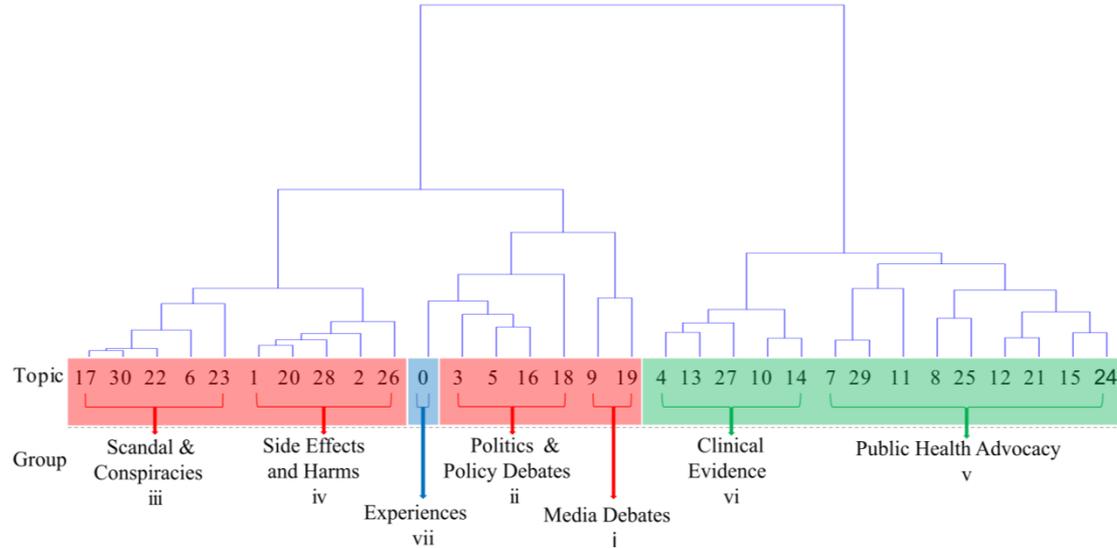
Signorini 2010

Salathe 2011

- We started collecting HPV vaccine tweets in 2013 (764,869 tweets as of 29 Jan 2018).
- Discovered we could predict negative tweets without reading them; just need to know who the users follow.
- Users mostly exposed to negative tweets (orange) were much more likely to go on to post a negative tweet rather than positive tweet about HPV vaccines compared to other users (green).



I wanted to know if Twitter could be a good enough signal of a population's **information diet** that we could *actually use* it to understand whether information was demonstrably associated with attitudes, behaviours, and HPV vaccine coverage.



***“Computer scientists often brilliantly solve the wrong problems.”***

**Problem 1:** flawed studies published with potentially dangerous conclusions.

- Call out bad epidemiology from data science;
- Contribute to pre-publication and post-publication peer review;
- Know and understand the limits and biases in social media data; and
- Real collaboration to teach computer scientists epidemiological methods.

**Problem 2:** robust studies that can't be used to guide policy or practice.

- Ask data scientists the *“why are we doing this?”* question more often; and
- Help translate research findings into real-time news/misinformation reporting and surveillance of attitudes and behaviours.

A poor literature review can be a good indicator of less robust research in this area.



**MACQUARIE**  
University

I wanted to be able to connect the work back to how people and communities *actually engage* with **evidence and misinformation**, because these are where we can provide evidence for how to manipulate the information diets of at-risk populations.





**Twitter:** [@adamgdunn](https://twitter.com/adamgdunn)  
**Email:** adam.dunn@mq.edu.au

These slides, with links, will be made available online at:  
[www.adamgdunn.net/speaking/](http://www.adamgdunn.net/speaking/)  
with thanks to @parryville for video production

1. **AG Dunn**, D Surian, J Leask, A Dey, KD Mandl, E Coiera (2017) Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States, *Vaccine*, 35(23):3033-3040.
2. G Shapiro, D Surian, **AG Dunn**, R Perry, M Kelaher (2017) Comparing human papillomavirus vaccine concerns on Twitter: A cross-sectional study of users in Australia, Canada, and the UK, *BMJ Open*, 7:e016869.
3. D Surian, DQ Nguyen, G Kennedy, M Johnson, E Coiera, **AG Dunn** (2016) Characterizing Twitter discussions about HPV vaccines using topic modelling and community detection. *Journal of Medical Internet Research*, 18(8):e232.
4. **AG Dunn**, J Leask, X Zhou, KD Mandl, E Coiera (2015) Associations between exposure and expression of negative opinions about human papillomavirus vaccines on social media: an observational study. *Journal of Medical Internet Research*, 17(6):e144.
5. X Zhou, E Coiera, G Tsafnat, D Arachi, M-S Ong, **AG Dunn** (2015) Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter, *Studies in Health Technology and Informatics*, 216:761-765.

## Other examples of *data-driven* studies using Twitter data related to HPV vaccines:

1. J Du, J Xu, H-Y Song, **C Tao** (2017) Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data, *BMC Medical Informatics and Decision Making*, 17(Suppl2):69.
2. J Du, J Xu, H-Y Song, **C Tao** (2017) Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets, *Journal of Biomedical Semantics*, 8:9.
3. J Du, J Huang, R Duan, Y Chen, **C Tao** (2017) Comparing the human papillomavirus vaccination opinions trends from different Twitter user groups with a machine learning based system and semiparametric nonlinear regression, *MEDINFO 2017*.
4. P Chakraborty P, **J Colditz**, A Silvestre, M Friedman, K Bogen, B Primack (2017) Observation of public sentiment toward human papillomavirus vaccination on Twitter. *Cogent Medicine*, 4(1):1390853.
5. Wheldon C, **Bian J**, Moser R. Surveillance of HPV-related vaccination discourse on Twitter in the United States: A case study of Kansas and Rhode Island. *2016 International Conference on Social Media & Society*. July 11-13, 2016. London, UK.
6. P Massey, A Leader, **E Yom-Tov**, A Budenz, K Fisher, A Klassen. Applying multiple data collection tools to quantify human papillomavirus vaccine communication on Twitter. *Journal of Medical Internet Research*, 2016;18(12):e318.
7. L Mahoney, T Tang, K Ji, J Ulrich-Schad. The digital distribution of public health news surrounding the human papillomavirus vaccination: a longitudinal infodemiology study. *JMIR Public Health and Surveillance*, 2015;1(1):e2.

Names in bold are good contacts for more information.