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)
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).

PubMed IDs:18675530; 20003922; 21573238; 22022249; 26262154; 26063290
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.
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.
We already have the tools we need to **identify individuals** with opinions and attitudes that put them at risk of unhealthy decisions and behaviours.

With automated personalisation of advertising and chatbots we are now able to deploy **autonomous behavioural interventions**; digital mass persuasion with no consent, and no humans involved in the diagnosis or delivery of the intervention.

If we choose to go down this path as a field of research, it will be vital that we get the ethics and implementation right.
These slides, with links, will be made available online at:
www.adamgdunn.net/speaking/
with thanks to @parryville for video production


All articles are open access or available as pre-prints.
Other examples of data-driven studies using Twitter data related to HPV vaccines:


Names in bold are good contacts for more information.