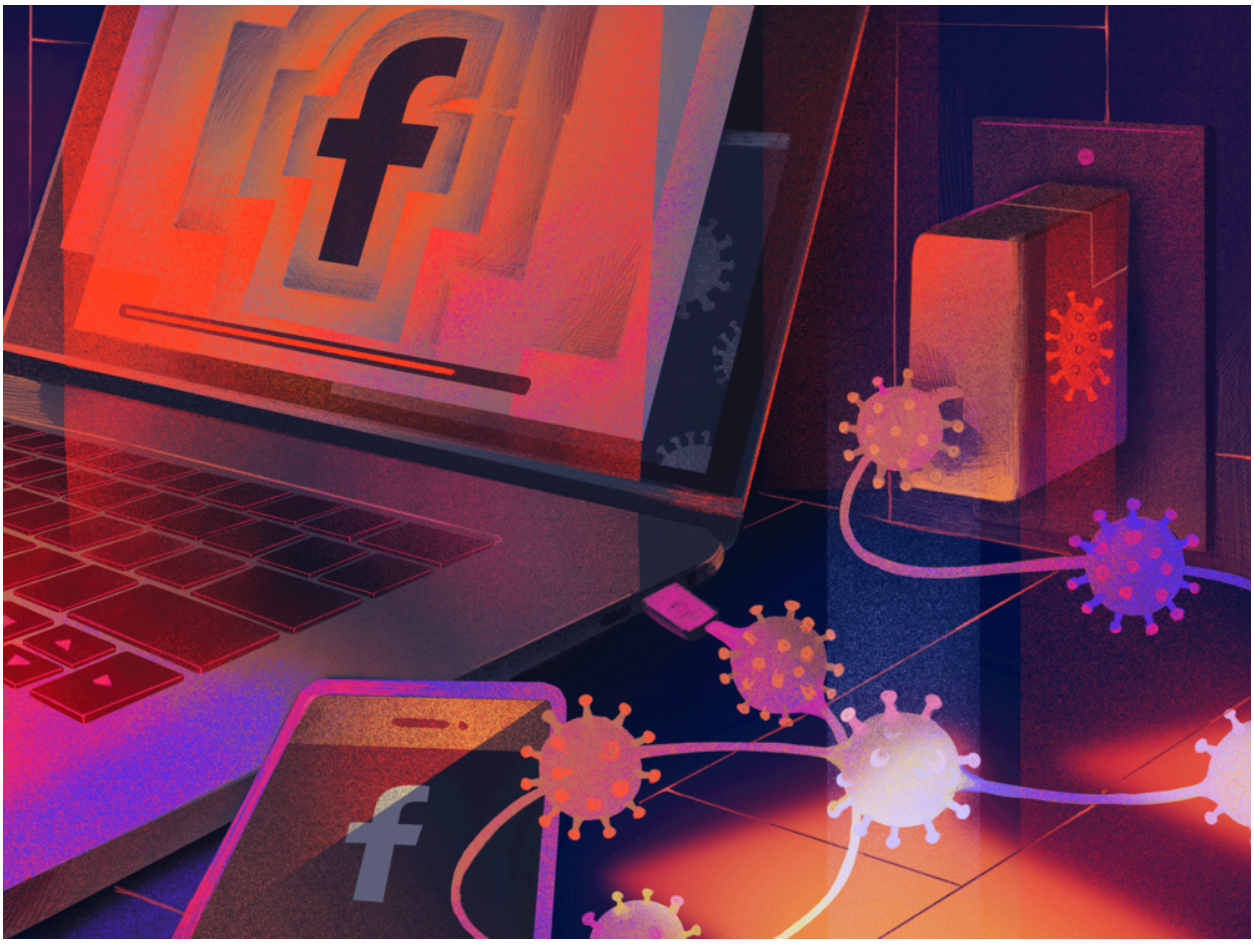
Protecting Public Health via Performance-Based Regulation of Online Health Misinformation



Amy Shim, Alejandro Jimenez Jaramillo, Aaron Matthew Schwartz
STS.487/6.805

Last updated January 2021
ILLUSTRATION BY DEBORAH LEE

Executive Summary

Online health misinformation has come to the center of national attention due to the ongoing COVID-19 pandemic. Although not a new issue, social media and other internet platforms amplify the harms of this content, which has already been traced to perverse economic, social, physical and mental health outcomes. Internet companies are calling on regulators to take action and develop regulatory frameworks to address this problem and that of misinformation more broadly.

The information age has accelerated the spread of online health misinformation. This is made apparent by the patterns of distribution associated with two distinct scenarios: COVID-19 and the modern anti-vaccination movement. Based on these scenarios, managing the spread of health misinformation by regulating the flow of information between users of social media on internet platforms not formally affiliated with medical institutions is a promising pathway to mitigating the harm caused by online health misinformation.

Effective regulation would need to operate within First Amendment jurisprudence. A review of relevant legal precedents reveals a general opposition to content-based speech regulation in the United States. It also highlights the opportunity provided by Section 230 of the Communications Decency Act to incentivize internet platforms towards effective content moderation.

This proposal introduces performance-based regulation (PBR), which ties performance to desired outcomes rather than mandating specific actions, as a framework for internet platforms to align their content moderation practices with the goal of reducing the spread of online health misinformation. PBR would allow internet companies to leverage the liability shield already provided by Section 230. Furthermore, PBR is an effective solution for this particular issue because it can accommodate different platform designs, stay abreast of technological development, and adapt to unforeseen consequences of regulation. This proposal analyzes Facebook's prevalence-based content moderation (PBCM) as a sample measurement system that could inform metric selection and design of national PBR. Existing gaps in Facebook's PCBM, namely lack of transparency and a lack of authoritative standards for what constitutes harmful content, would need to be addressed at the federal level to ensure robust, effective, and transparent regulation.

The FDA is the best agency to regulate health misinformation and implement a comprehensive PBR framework, due to its public-health focused missions, awareness of the issues of online health information, experience implementing PBR for medical

devices, expertise in health-related issues, and recent expansion into digital health. In addition, other federal agencies, such as the FTC, FCC, and CDC, can still play a role in the broader effort to address harmful online health misinformation.

A federal implementation of PBR for online health misinformation could take a variety of forms. One possible version could use the prevalence of federally-defined misinformation as a performance assessment metric and reward or penalize internet companies based on their content moderation performance relative to an agreed-upon prevalence threshold. Two case studies applying this framework to acute (the COVID-19 pandemic) and non-acute (the anti-vaccination movement) health crises show that PBR designed with flexibility, transparency, and accountability in mind can be an effective response to the promulgation of online health misinformation.

Recommendations

1) Regulators must take a proactive, ongoing role in defining what constitutes online health misinformation by partnering with medical experts and professionals.
2) Regulators should focus on moderating the dissemination of health misinformation among internet users.
3) Regulators should target social media platforms specifically to mitigate the spread of health misinformation.
4) Regulators should incentivize internet platforms to leverage their Section 230 immunity to mitigate the spread of health misinformation.
5) Regulators should implement performance-based regulation to address the harms of online health misinformation.
6) Regulators should recognize differences among internet platforms in size, design, and company policy, and implement performance metrics which keep those differences in mind. PBR is not a one-size-fits-all solution.
7) Regulators should minimize the “dangers” of performance-based regulation by incorporating relevant stakeholders such as companies, technical experts, and the public into the PBR design process.
8) Regulators should expect companies to act in good faith, and companies should work with regulators to develop effective standards for audit and verification. Trust and transparency are key.
9) Regulators should establish standards defining harmful content so that internet platforms have a structure within which to develop appropriate tracking and reporting mechanisms of health misinformation.
10) The FDA should implement performance-based regulation as part of a deliberate effort across federal agencies to reduce harmful online health misinformation.

Table of Contents

1. Introduction

In 2018, in his testimony before the United States Senate, Mark Zuckerberg said, “It’s not enough to give people a voice. We have to make sure people aren’t using it to harm people or spread disinformation”[1]. Two years later, Zuckerberg’s words ring hollow. Digital misinformation on Facebook remains a major problem, and one particular form has become particularly prevalent and dangerous – health misinformation. On December 2, 2020, for example, a member of the British parliament stated that “one of the greatest risks to the success of the [COVID-19 vaccine] program is anti-vaccine disinformation warning people not to take it”[2]. Around the world, health misinformation surrounding not only COVID-19, but other key public health issues, such as vaccinations, has led to perverse economic, physical, mental and social outcomes.

Beyond just health content, digital misinformation in general remains an issue in spite of efforts by internet companies to address it. Although Section 230 allows internet companies to self-moderate content on their platforms, regulators have been slow to take meaningful action that would directly mandate or incentivize internet companies to address harmful misinformation. Regulating online content is particularly challenging in the United States given the constitutional protection of free speech. This has internet companies concerned. They recognize the increased public awareness of the harms of online misinformation and are worried that regulation may be impending. As such, they are taking proactive steps to shape the conversation and to ensure that eventual regulation will be, at least in part, on their terms. In his op-ed in the Washington Post, published March 2019, Zuckerberg stated that he “believe[s] we need a more active role for governments and regulators” on internet issues including the proliferation of harmful content[3]. Following up on his comments, his company released specific recommendations for how regulators may address harmful digital content in its 2020 white paper “Charting a Way Forward: Online Content Moderation”[4].

One of the recommendations in Facebook’s white paper suggests that internet companies meet “performance targets” set by regulators that motivate action on digital content moderation. Whether the company knew it or not, they were offering what may be one of the first formal suggestions that internet companies be subject to a regulatory scheme known as “performance-based regulation”.

---

[1] Rocha et al. 2018

[2] O’Sullivan et al. 2020

[3] Zuckerberg 2020

[4] Bickert 2020

Performance-based regulation, or PBR, is a type of regulation that has been used across business sectors such as environmental regulation, building safety, and electric power for decades, but to the authors' best knowledge, never before been applied to digital content moderation. PBR requires or incentivizes companies to meet certain performance targets[5]. Regulators may set penalties such as fines for companies that fail to meet performance targets, or offer financial subsidies or other rewards to companies that exceed them. Under PBR, regulators don't specify how companies meet these targets, but only mandate that they be met. This gives companies flexibility in how they choose to achieve performance targets. When regulators expect companies to self-report performance outcomes, they may choose to establish some type of audit or assessment mechanism to ensure that reported metrics align with actual outcomes.

This proposal will examine what effective performance-based regulation of digital health misinformation may look like on internet platforms. Internet platforms are defined as internet services designed around users' ability to upload, share, and view content online. First, it will outline the dangers of online health misinformation, including misinformation related to acute public health crises, such as the ongoing COVID-19 pandemic, and to health misinformation outside of such crises, such as misinformation surrounding routine vaccinations. Next, it will provide an overview and analysis of the relevant jurisprudence and law relevant to regulating health misinformation. Third, it will describe performance-based regulation, and evaluate how such a regulatory framework could be applied to internet platforms to reduce the negative impacts of online health misinformation. It will then discuss how Facebook is already using a particular metric, prevalence, in its ongoing efforts to minimize harm on its platform through prevalence-based content moderation, as an example of a performance metric that regulators could use in PBR. Fifth, It will argue for the FDA to administer a comprehensive PBR program to reduce health misinformation, and suggest the roles that other federal agencies may play. Finally, it will use two case studies, the COVID-19 pandemic and anti-vaxx misinformation, to examine what a specific implementation of performance-based regulation of internet platforms using a prevalence metric might look like.

---

[5] Coglianese et al. 2003

2. Online Health Misinformation

To establish context for the issue and begin to frame performance-based regulation as a potential solution, this section will describe the modern state of online health misinformation and its impact on consumers via major internet platforms. Although somes sources in the field differentiate between misinformation and disinformation[6] (misinformation spread with deliberate intent to cause harm), this proposal will focus on the impact of health misinformation and not on the intention of the user.

2.1 Defining Online Health Misinformation

The Journal of the American Medical Association has defined health misinformation as "a health-related claim of fact that is currently false due to a lack of scientific evidence"[7]. The dynamism inherent in this definition is worth nothing. For example, some medical claims that are backed by current evidence might be disproved at a future date as additional research is conducted while others with limited support may be embraced only once additional work has been done. As such, any attempt to regulate health misinformation will require ongoing collaboration between medical experts and regulatory agencies to update guidelines and standards for what constitutes "health misinformation" as expertise grows and changes over time.

Recommendation 1: Regulators must take a proactive role in defining what constitutes online health misinformation by partnering with medical experts and professionals.

2.1.1 Two Types of Medical Knowledge and Health Misinformation

Although the contours of medical knowledge may change over time, this proposal will focus on two distinct areas to illustrate the range of impacts that health misinformation can have:

1) Medical knowledge that is **rapidly developing** in the course of a acute health emergency, and
2) **Well-established** medical knowledge that has been consistently supported by decades of research.

This proposal will rely on COVID-19 misinformation and campaigns seeking to challenge the safety of vaccine treatments ('anti-vaccination campaigns'), respectively,

---

[6] Wardle and Derakhshan 2017, Wang et al. 2019

[7] Chou et al. 2018

as examples for each of the domains listed above. Although these are not the only potential examples, these cases provide ample evidence documenting the existence, spread, and harm caused by health misinformation. Furthermore, a meta-analysis of health misinformation found that pandemic content and vaccination content comprise the largest and most-extensively studied proportion of health misinformation to date[8].

2.2 The Rise of Online Health Misinformation

Health misinformation existed long before the information age[9], but access to, promulgation of, and reliance upon, online health information from unverified experts has increased exponentially since the arrival of the Internet. The extent and nature of the issue suggests a distinct approach to regulation focused on limiting the distribution of online health misinformation between users outside of established medical institutions.

2.2.1 The Internet has Decentralized Access to Health Information

Generally, the emergence of the Internet has supported the 'pulling' of information by consumers from the Internet as opposed to the 'pushing' of information to consumers from established medical sources[10]. As early as 2013, over 30% of the general U.S. population and 70% of U.S. internet users reported using the Internet for online health information[11]. It also has been shown that individuals without insurance and individuals located further away from their general source of care were more likely to search for information online, suggesting that financial and time costs might incentivize using the Internet as a source of information rather than a healthcare provider[12].

Although verified, institutionally-affiliated sources of information exist online (for example, hospital websites and government health organizations), there are just as many sources that might be independently managed or not specifically attached to a verified expert. In addition, many user-generated health claims are made and shared on internet platforms, like YouTube, Facebook, and Twitter[13]. YouTube, in particular, has

---

[8] Wang et al. 2019

[9] Gandhi 2013

[10] Madden and Fox 2006

[11] Madden and Fox 2013

[12] Bundorf et al. 2006

[13] Li et al. 2020

been the focus of many research studies identifying and classifying the accuracy of health information[14].

2.2.2 Online Health Information Sharing is Highly Participatory

Historically, the medical establishment has held a monopoly on medical accuracy. However, the Internet has augmented the role that third-party and non-expert individuals play in the health dialogue between a given patient and their healthcare provider. This has effectively lead to a crowdsourced understanding of health[15]. A distinct manifestation of this trend is the proliferation of of online health forums, websites where groups of individuals who may or may not have any formal qualifications where information and advice about specific diseases are shared[16].

2.2.3 Internet Users have Limited Ability to Accurately Evaluate Online Information

In addition to a significant expansion in access and an increase in unverified claims, the potential risk posed by online health misinformation is further exacerbated by consumers' limited information evaluation skills. Studies have shown that internet users are often unable to distinguish between valid and invalid sources of authority and prefer anecdotal patient narratives over scientifically accurate information[17]. This can be exacerbated by the complexity of content from official sources such as hospitals and other health organizations, making it difficult for internet users to decipher accurate information from trusted sources[18].

2.2.4 Online Health Misinformation is a Sociotechnical Exigency

In summary, the increase in online access to information, the participatory nature of the Internet, and limited consumer ability to assess and evaluate online sources has increased the number of people relying on medical claims from unverified online sources. This particular socio-technical matrix led the World Health Organization to

---

[14] Garg et al. 2015, Strychowsky et al. 2013

[15] Silber 2009, Witteman and Zikmund-Fisher 2012, Stewart Loane and D'Alessandro 2014

[16] Kinsora et al. 2017

[17] Fritch and Cromwell 2001, Garg et al. 2015

[18] Dahl and Eagle 2016

declare an “Infodemic” earlier this year that developed in parallel to the general spread of COVID-19[19].

The trends highlighted above suggest three potential “Levers” that can be used to manage the spread of online health misinformation:

1) Restrict or moderate access to websites containing health misinformation,
2) Moderate the sharing of unverified information between internet users, or
3) Improve internet users’ evaluation skills with regard to online content.

Although any of these three levers could be the basis for an online health misinformation management policy, this proposal will focus on the second lever. Lever 1 would risk censoring entire websites, and Lever 3 would require long-term investments in education to yield improvements. As this proposal will show, Lever 2 can be implemented with a more targeted proposal that utilizes the network effects of major internet companies.

Recommendation 2: Regulators should focus on moderating the dissemination of health misinformation among internet users.

2.3 How Health Misinformation Causes Harm

Comparing the impacts of COVID-19 misinformation and anti-vaccination campaigns illustrates a range of types of online health misinformation with ample documentation of demonstrated harms. However, the centrality of social media (and internet platforms generally) to the distribution of online health misinformation suggests that these firms might be best positioned to restrict the flow of misinformation and mitigate subsequent harm.

2.3.1 Types of Misinformation

COVID-19

The types of health misinformation that have been observed in the wake of COVID-19’s spread include unproven cures (i.e. cures with insufficient clinical evidence to be deemed effective or safe), inaccurate descriptions of actions taken by public health authorities, false claims about the extent of the infection (both in terms of geographic area as well as specific ethnic origins), and the mechanism of infection[20]. In the

---

[19] World Health Organization 2020

[20] Brennen et al. 2020

particular case of COVID-19, misinformation was more often a “spun, twisted, recontextualised, or reworked” version of the truth than a completely fabricated lie[21].

Anti-Vaccination Campaign

In the case of the now global anti-vaccination movement[22], the growth of denialists and skeptics was initially fueled by a (now discredited) medical study that purported to show an association between the Measles, Mumps, Rubella (MMR) vaccine and autism[23]. Although initially catalyzed by a medical study in a professional journal, anti-vaccination myths and sentiments are now sustained around the world on social media[24]. The vast majority of misinformation on this topic questions the safety of vaccine treatment, although some more extreme forms invoke conspiracy theories and other rumors designed to question the purpose of vaccines or the integrity of vaccine makers[25].

2.3.2 Types of Documented Harm

COVID-19

COVID-19 misinformation has been shown to cause physical, mental, and social harm. With regard to physical harm, unproven cures for the disease (some of which were endorsed without evidence by the President of the United States) have led to deaths around the globe[26]. At a more general level, misinformation about the severity of the disease and what constitutes effective preventative measures has undoubtedly amplified the spread of COVID-19, particularly amongst younger adults[27]. Although it is likely that COVID-19 misinformation has also caused economic harm, it is difficult to separate the economic impacts due to misinformation specifically from the general impact of the pandemic. As such, this proposal avoids proposing estimates for this category of impact.

COVID-19 misinformation also has negative impacts on mental health[28]. To the extent that it delays the abatement of the crisis, health misinformation has exerted a mental toll

---

[21] ibid.

[22] Roberts 2019

[23] Wakefield 1998

[24] BBC Monitoring 2019, McKee and Diethelm 2010

[25] Goodman and Carmichael 2020

[26] Spring 2020, Love et al. 2020, Waldrop et al. 2020

[27] Wilson et al. 2020

[28] Logie and Turan 2020

on everyone who is in quarantine without social contact, a measure that researchers knew in February 2020 would require active investments in timely mental health care[29] and that a recent literature review has confirmed via elevated levels of stress, anxiety, and depression[30].

From a social perspective, COVID-19 misinformation has amplified xenophobia, sometimes violently[31]. In addition, the supply chain disruptions due to hoarding and panic buying of unproven cures, such as hydroxychloroquine, have caused medical supply shortages for individuals suffering from rheumatoid arthritis and lupus, for whom hydroxychloroquine is an essential and verified treatment[32]. Furthermore, COVID-19 misinformation can erode trust in public health agencies, as was the case with regard to mask-wearing guidance from the Center for Disease Control and Prevention[33]. This lack of faith made it harder for the agency to engage the public and incentivize adherence to mask-wearing and other COVID-19 mitigation tactics, likely leading to a further weakening of containment protocols.

Anti-Vaccination Campaign

In the context of anti-vaccination, harm can be most accurately quantified in physical and economic terms by analyzing outbreaks in areas that had previously eliminated cases of a given disease via vaccine treatments. As an example, the United States had effectively eliminated measles in 2000 only to see thousands of cases arise in the past decade, most of which were reported amongst individuals who had not been vaccinated[34].

Measles outbreaks due to vaccine hesitancy and subsequent low vaccination rates have been documented domestically in Minnesota[35] and Ohio[36]. Although recovery can be swift with prompt medical care, historical data show that 20% of infected individuals

---

[29] Xiang et al. 2020, Zhang et al. 2020, Bao et al. 2020

[30] Rajkumar 2020

[31] Wong 2020, BBC News 2020, Spring 2020

[32] Kim et al. 2020

[33] Wetsman 2020

[34] Belluz 2017, CDC 2020

[35] Myers 2017, Belluz 2017

[36] Belluz 2015

require hospitalization, 5% contract pneumonia, and 0.1-0.3% die due to complications arising from the disease[37].

Given that measles is almost entirely preventable via low-cost infant vaccinations and the United States essentially eliminated the disease in 2000[38], public health expenditures arising from low vaccination rates due to anti-vaccination campaigns constitute unnecessary economic harm. As a result of necessary intensive follow-up care for patients stricken with measles, the 2017 Minnesota outbreak of 79 total cases cost approximately $1.3 million[39]. Cost estimates of 16 outbreaks comprising 107 cases across the US in 2011 range from $2.7-5.3 million[40]. Additional modeling suggests that even 5% reduction in vaccination rates due to vaccine hesitancy can lead to over $2 million in consequent public health costs[41].

Finally, social harm manifested in Minnesota via inequitable health outcomes; the anti-vaccination campaign targeted the Minneapolis-based Somali population beginning in 2008, leading to a vaccination rate that was slightly higher than the non-Somali Minnesotan population (92% in 2004) to plummet by more than half across the next ten years (42% in 2014)[42]. Not only does a targeted anti-vaccination campaign cause lower vaccination rates and a higher risk of contracting the disease for the population in question, but it also sows doubt amongst the community in public health institutions, making it more difficult to manage future outbreaks.

2.3.3 Patterns of Misinformation Dissemination

COVID-19

Many sources suggest that social media platforms contributed to the ‘Infodemic’ that accompanied the COVID-19 pandemic[43]. One particular pattern observed by Reuters

---

[37] CDC 2020

[38] Belluz 2017

[39] Minnesota Department of Health 2017

[40] Ortega-Sanchez et al. 2014

[41] Lo and Hotez 2017

[42] Belluz 2017

[43] Richtel 2020, Li et al. 2020, Miller et al. 2017, Spring 2020, Frenkel et al. 2020, Wang et al. 2019, Burki 2020

was that key nodes in social networks (politicians, celebrities, etc.) were responsible for a disproportionate share of social media engagement with health misinformation[44].

Social media firms have made efforts to mitigate the spread of COVID-19 misinformation, including identifying and correcting disputed information[45] and working with public health agencies to proactively share correct information[46]. However, it is difficult to independently verify how effective these actions have been as platforms like Twitter and Facebook do not share data with external auditors for verification purposes.

Anti-Vaccination Campaign

Anti-vaccination campaigns rely heavily on social media both to share content and establish group identity. With regards to content sharing, the anti-vaccination community has consistently used social media to propagate health misinformation[47]. Furthermore, social media platforms have also created the conditions for public health denialists to find one another, allowing communities of like-minded individuals to form that antagonize public health goals[48]. Although general skepticism or disbelief in the efficacy or safety of vaccination is shared amongst these groups, the arguments justifying their beliefs draw on pseudo-science, religion, and conspiracy-driven political views[49]. In fact, recent research has also found that smaller, well-connected groups on social media have been disproportionately responsible for sowing doubt about the safety and efficacy of COVID-19 vaccines, enough so to prompt action from Facebook by banning high-profile anti-vaccination accounts[50].

Recommendation 3: Regulators should target social media platforms specifically to mitigate the spread of health misinformation.

---

[44] Brennen et al. 2020

[45] Oyeyemi et al. 2014, Spring 2020

[46] Zuckerberg 2020

[47] Wang, Y et al. 2019, Strychowsky et al. 2013, Donzelli et al. 2018, Basch et al. 2017, Tustin et al. 2018

[48] McKee and Diethelm 2010

[49] BBC Monitoring 2019, Wolfe and Sharp 2002

[50] Zadrozny 2020

3. Legal Precedents for Regulating Health Misinformation

Having established how online health misinformation causes harm, this proposal now turns to a brief review of relevant legal precedents with the intent of identifying a defensible regulatory frame that addresses dissemination of health misinformation between users on internet platforms. Key cases include those that address false speech regulation, the protection of public health via speech regulation, and tort jurisprudence for false information.

3.1 False Speech Jurisprudence

The defining court case for this issue is United States v. Alvarez[51], in which the United States declared unconstitutional a statute that criminalized the act of falsely claiming the receipt of military medals, declaring there is no “general exception to the First Amendment for false statements”. One of the key points of reasoning was that false speech must be tied to some type of ‘cognizable harm’ to merit regulation, and even in such cases the court affirmed the general principle of the least restrictive means. Although this proposal has demonstrated clear harm resulting from online health misinformation, the allocation of fault and liability can be difficult to prove in a dynamic social media environment involving many actors. Efforts to restrict false speech altogether would seem to overshoot the mark, even if only for a single health domain. Ultimately, regulation that bans speech on the basis of its veracity alone is not likely to survive judicial review.

3.2 Public Health as a Compelling Government Interest

First Amendment jurisprudence has acknowledged that free speech can be regulated to advance government interests. In United States v. O’Brien[52], the court affirmed that: 1) not all forms of speech have First Amendment protections and 2) that speech can be regulated on the basis of content if a specific, legitimate, and compelling government interest can be demonstrated. In this particular case, the speech in question was the burning of a selective service card. Although no actual words were exchanged, the act in itself was clearly undertaken to communicate an idea, a frame of reference that allowed the court to consider the action in the context of First Amendment jurisprudence.

---

[51] U.S. Supreme Court *United States v. Alvarez* 2012

[52] U.S. Supreme Court *United States v. O’Brien* 1968

There is evidence to suggest that public health is a compelling government interest. In Jacobson v. Massachusetts[53], the Supreme Court upheld the state's right to subject individuals who refused to receive a smallpox vaccination to criminal penalties. While the major focus of the case was on the application of police power to this particular issue, it was undergirded by logic that validated "such reasonable regulations relating to matters completely within its territory, and not affecting the people of other States, established directly by legislative enactment, as will protect the public health and safety"[54].

Although the aforementioned cases suggest that public health can be cited as a clear and compelling government interest to justify regulating certain kinds of speech, 'public health' in general may lack the specificity required to survive judicial review.

3.3 User-Generated Health Misinformation

Given that free speech can be regulated with appropriately tailored regulation for a specific, compelling government interest and that public health and safety has been recognized as a compelling government interest, it stands to reason that legislation regulating online health misinformation on such grounds might be accepted by First Amendment jurisprudence. However, recent legal analysis and events suggest otherwise.

An analysis of tort jurisprudence suggests that First Amendment protections might overpower the right of an individual to seek damages for harm caused by online health misinformation. Borrowing from a taxonomy of tort typologies developed by Diamond and Primm[55], an extensive law review by Rubinstein and Diamond distills previous tort jurisprudence (specifically section 311) and applies it to the case study of harm caused by anti-vaccination activists on Somali migrants in Minnesota to assess the legal existence of liability for anti-vaccine information leading to demonstrated harm[56]. The authors find that the ability to successfully sue using tort law is heavily contingent on the plaintiff being able to assert a legal duty of care, concluding that "in some circumstances, where there is a consultation-like relationship, such a suit is possible. However, a suit is likely unsuccessful when the information is only shared online or on

---

[53] U.S. Supreme Court *Jacobson v. Massachusetts* 1905

[54] ibid.

[55] Diamond and Primm 1988

[56] Reiss and Diamond 2019

forums aimed at the general public”[57]. As such, tort-based lawsuits against individuals who share health misinformation online are unlikely to be successful.

In addition, medical claims and political speech can sometimes be so closely connected that regulating the former effectively constrains the latter, which the judicial system has demonstrated opposition to since the founding of the United States. This was exemplified by a lawsuit earlier this summer launched by the Washington League for Increased Transparency and Ethics (WashLITE) against Fox News to prevent “publishing further and false and deceptive content”[58]. The content produced by Fox News combined skepticism of the harm caused by and method of spread for COVID-19 with political allegations against left-leaning politicians that the virus and subsequent media coverage was a “political weapon against the president”[59]. Although WashLITE cited staff deaths due to COVID-19 as evidence of the harm caused by Fox News as a broadcaster, a state judge in Washington ultimately threw out the case because ”the speech in this case involves matters of public concern that is at the heart of the First Amendment’s protection”[60].

3.4 Section 230 as an Opportunity for Regulation

Ultimately, the preceding analysis suggests that statutory regulations directly addressing the dissemination of health misinformation between internet users is unlikely to survive a First Amendment challenge. However, Section 230 allows internet companies to self-regulate in ways that might otherwise be unconstitutional. The legal shield provided by Section 230 gives internet platforms the flexibility to implement content moderation policies and practices that might not pass judicial review in another context[61]. For example, Section 230 has enabled Facebook[62], Twitter[63], Reddit[64], and other social media platforms to unilaterally establish internal speech standards that they can enforce with limited oversight from regulatory bodies. Furthermore, since social media platforms

---

[57] ibid.

[58] Steinberg 2020

[59] Swoyer 2020

[60] ibid.

[61] U.S.C. §§ 230 1996

[62] Facebook Community Standards 2020

[63] Twitter 2020

[64] Reddit Content Policy 2020

are a major source of online health misinformation, incentivizing internet platforms to self-regulate the dissemination of online health misinformation would bypass the aforementioned legal obstacles to such regulation and directly address the spread of online health misinformation through a medium central to its distribution.

Recommendation 4: Regulators should incentivize internet platforms to leverage their Section 230 immunity to mitigate the spread of health misinformation.

4. Performance-Based Regulation of Health Misinformation

The preceding legal analysis assessed the landscape within which effective regulation of online health misinformation would need to operate. It found that Section 230 provides an opportunity for federal agencies to structure an incentive scheme for internet platforms to self-regulate the spread of online health misinformation. Performance-based regulation (PBR) offers a regulatory framework that could operate under Section 230's "Good Samaritan" clause to minimize the dissemination of online health information and give internet platforms flexibility in their method of implementation. This section will contextualize PBR as an expansion of and improvement upon systems that internet platforms already use to manage harmful content, taking Facebook's prevalence-based content moderation (PBCM) system as an example. An analysis of Facebook's system will highlight the types of issues that a federal regulatory system would have to proactively address.

4.1 What is Performance-Based Regulation?

Performance-based regulation is a type of regulatory model that "specifies required outcomes or objectives, rather than the means by which they must be achieved"[65]. Regulations that set and enforce water quality standards, electric utility energy efficiency targets, mandated air purity levels within certain classes of buildings – all of these are examples of PBR. Peter J. May identifies three "dimensions" of this kind of regulatory framework: characterization of outcomes, characterization of desired level of achievement, and performance assessment[66]. The characterization of outcomes describes the intended effect of the regulation, which can be comprehensive (high level) or less comprehensive (low level). For example, a high level characterization of outcomes could be improved air quality in a major metropolitan area, while the low level characterization of outcomes could be reduced NOx emissions from automobiles. The characterization of desired level of achievement is a specific target against which "compliance" is measured – for example, a specific NOx threshold that must be achieved in order to comply with the regulation. Finally, a performance assessment is required to identify whether the desired level of achievement is being met – for example, a measurement plan to track changes in NOx emissions.

---

[65] OECD 2002

[66] May 2003

Performance metrics are a crucial element in PBR, and are central to the performance assessment process. Performance metrics may directly quantify the desired outcome or measure proxies for desired outcomes. Continuing the example of NOx emissions, a metric that measures and quantifies NOx levels in the air provides a direct indicator of whether the desired outcome is reached. Another performance metric may be total reported incidences of respiratory diseases associated with NOx emissions. This metric is only a proxy for the ultimate value of interest (NOx levels), but can still be used by regulators in the performance assessment process. Proxies are especially useful when it is difficult, expensive, or impossible to directly measure desired outcomes.

In the Michigan Law Review article “The Limits of Performance-Based Regulation,” Cary Coglianese offers a framework for determining when PBR may be a reasonable choice of regulation compared to other options. A necessary condition for PBR is the ability for regulators to “measure and monitor” outcomes. Performance assessments must go beyond one-time measurements of outcomes. They should include active monitoring, and where applicable, verification of reported outcomes. Additionally, PBR offers two kinds of flexibility over other kinds of regulation – ”cross-sectional” flexibility and “longitudinal” flexibility. Cross-sectional flexibility means that PBR allows different firms to meet performance standards in whichever ways work for them, as long as the performance target is achieved. For example, performance-based regulation meant to foster fuel efficiency in automobiles could allow auto manufacturers, which have a diverse range of proprietary designs and technologies, to meet efficiency targets in their own way. This means that PBR has an advantage over “means based” regulation, or regulation that states specific and proscriptive actions a company must take, when companies subject to regulation are heterogeneous. Longitudinal flexibility means that PBR allows firms to more easily change over time, which supports innovation. For example, as auto manufacturers continue to develop new technologies and more efficient vehicle designs, they are free to change their approach to meeting fuel efficiency standards rather than being locked in to a proscribed set of technologies, designs, or materials.

Coglianese cites “global enthusiasm” for PBR, and notes that PBR has been promoted as a preferred method of regulation by both Democratic and Republican federal administrations for decades[67]. New applications of PBR continue to be proposed or enacted. For example, Stephen D. Sugarman proposed applying PBR for reducing death, injury and disease caused by five consumer products – cigarettes, alcohol, guns, junk food, and motor vehicles. Although PBR for the electricity sector was first proposed

[67] Coglianese 2013

in the late 1980s and early 1990s[68], it has seen a resurgence in recent decades, with a number of states implementing PBR of electric utilities.

Despite PBR's popularity, it is not without faults. Coglianese's framework also describes seven "dangers" of PBR (listed below) to help relevant stakeholders, including regulators, industry, and the public, evaluate potential issues that may arise from PBR. Any proposal for a new application of PBR should consider these "dangers" alongside PBR's potential benefits.

The Seven Dangers of PBR:

1) PBR may not lead to less conflict between regulators and regulated entities than other forms of regulation,
2) PBR may lead to unintended consequences from unrecognized tradeoffs,
3) PBR does not necessarily reduce or eliminate regulatory red tape,
4) PBR can limit discretion or flexibility in accomplishing performance standards,
5) PBR can create the potential for fraud or evasive behavior,
6) PBR may incentivize "Teaching to the Test," where companies find ways to meet performance standards that fail to accomplish the regulatory goals,
7) PBR may rely upon incorrect proxies or causal factors (i.e. the selected performance standards are ineffective at accomplishing the regulatory goals).

4.2 Applying Performance-Based Regulation to Internet Companies

Facebook's aforementioned white paper, "Charting a Way Forward: Online Content Moderation", poses the following question: "Should regulation require internet companies to meet certain performance targets?"[69]. This frames content moderation as a potential application of PBR for internet companies. Indeed, the proposed regulation can be mapped neatly to May's three dimensions of PBR:

---

[68] Bowman and McKay 2020

[69] Bickert 2020

Table 1

How Facebook's content moderation proposal fits into the dimensions of PBR

Dimension	**Facebook's Proposal**
Characterization of Outcomes	Hold internet companies accountable for reducing harmful content on their platforms.
Characterization of Desired Level of Achievement	Regulators would set specific thresholds based on metrics such as: **Prevalence** – how much violating content is viewed on the platform. **Time-to-Action** – how long violating content exists on the platform before action.
Performance Assessment	Although not specified in the proposal, this would involve the specific methods by which internet platforms measure outcomes (i.e. measuring prevalence or time-to-action), such as algorithmic detection or human content moderation. In addition, it may involve audits, verification, or assessments of performance by the regulator.

To the best of the authors' knowledge, PBR has never been applied in practice to internet platforms. By recognizing Facebook's proposal as an example of PBR, it can be assessed through the lens of existing literature on PBR. This offers opportunities to identify shortcomings and potential issues alongside potential benefits.

As noted in section 4.1, regulators must be able to "measure and monitor" outcomes for PBR to be effective. It seems unlikely, although not impossible, that the regulator would measure outcomes on behalf of internet companies, due to limited technical capacity, data privacy, difference between internet platforms, and other such constraints. In comparison, internet companies already have sophisticated software and methods for tracking various metrics on their platforms, including ones which may be of interest to regulators of misinformation. In fact, the business models of many internet platforms depend on being able to accurately quantify and track various metrics on their platform relevant to advertisers, and to keep their platforms appealing and appropriate for most users through content moderation. For this reason, rather than measure outcomes themselves, regulators would likely hand measurement responsibilities over to internet

companies. Regulators would be responsible for developing protocols to “monitor” internet companies, ensuring that their performance is measured and reported honestly and accurately.

In Coglianese’s framework, PBR is well suited for industries where firms are different from one another. There are a variety of different types of internet platforms that may incidentally host health misinformation, which range from “feed” style services, like Facebook and Twitter, to “message-board” style services like Reddit, to video-sharing platforms like YouTube. It will be important to define the range of internet platforms that would fall under the scope of PBR, since some companies may not have the requisite level of sophistication to identify misinformation and accurately track moderation metrics. Furthermore, different metrics may be better suited for different internet platforms. While some firms may have developed sophisticated procedures for measuring prevalence, other companies may base their own content moderation on other metrics, which they have robustly developed and integrated into their content moderation practices. Regulators should be mindful of the strengths and weaknesses of different metrics, and the existing sophistication of internet companies’ technologies and strategies for computing them.

Finally, internet companies may be particularly well-suited to benefit from the “longitudinal flexibility” offered by PBR. Due to the historically rapid pace of digital technological innovation, means-based regulation may quickly become ineffective or irrelevant. A PBR framework could offer greater flexibility for regulators to adapt to these changes, and give internet companies more room for innovation. Just as a robust PBR framework should accommodate different metrics to reflect the differences between existing internet platforms today, so should it allow flexibility for internet companies to evolve and adapt their own metrics to be most effective at identifying and addressing misinformation on their platforms in the future.

Facebook’s proposal specifically addresses some of Coglianese’s “dangers” of PBR. It acknowledges the potential for performance-based regulation to cause companies to “teach to the test,” and describes the potential for “perverse incentives” to emerge, whereby internet companies would prioritize meeting performance targets while dedicating less time to other important moderation or content-safety tasks not specifically covered by the performance metric. It also recognizes the potential for unrecognized tradeoffs, noting the “significant trade-offs regulators must consider when identifying metrics and thresholds.”

Recommendation 5: Regulators should implement performance-based regulation to address the harms of online health misinformation.

Recommendation 6: Regulators should recognize differences among internet platforms in size, design, and company policy, and implement performance metrics which keep those differences in mind. PBR is not a one-size-fits-all solution.

Recommendation 7: Regulators should minimize the "dangers" of performance-based regulation by incorporating relevant stakeholders such as companies, technical experts, and the public into the PBR design process.

4.3 Prevalence-Based Content Moderation at Facebook

Having introduced PBR and assessed its general appropriateness for regulating internet companies, this section will describe how metrics are used to moderate content at Facebook to illustrate a metric that could be used in PBR to manage the spread of online health misinformation. Their process, which this proposal will refer to as 'prevalence-based content moderation' (PBCM), relies on quantitative indicators that describe the amount of exposure harmful content receives to assess the efficacy of their algorithms in identifying and reducing the distribution of such content.

4.3.1 Who is Responsible?

In general, Facebook's content moderation system combines human and algorithmic review to combine the benefits of contextual discernment with scale[70]. Human reviewers are able to analyze content in the context of a given local community or language and adjust to dynamic ways of sharing information or ideas[71]. In contrast, algorithms build off of images, video, and other content that human reviewers have classified to proactively identify and remove content before someone reports the content as harmful[72].

This combination also minimizes costs. The sheer volume of content produced on Facebook would make it extremely costly (and potentially impossible) to hire a sufficient number of people to review all content.

---

[70] Koebler and Cox 2018, Facebook "Understanding the Community Standards Enforcement Report" 2020

[71] Bickert 2020

[72] Lyons 2018

4.3.2 Prevalence: A Measure of Content Exposure

Although specific calculations may vary, prevalence metrics seek to measure the amount of exposure a given piece of content has. Content-moderation decisions that utilize this indicator rely on the assumption that the amount of exposure (literally "number of views") a piece of content has is directly proportional to the impact that content can have[73]. Managing harmful content on the basis of prevalence seeks to reduce harm by reducing the amount of exposure harmful content has to users. At Facebook, prevalence is defined as the relative proportion of views that harmful content has relative to all content viewed[74]. The relative nature of this metric is worth noting. In an independent report developed by the Justice Collaboratory, experts in measurement recommended that harmful content relative to all views both relative and absolute numbers of views be used to gauge content moderation efficacy[75].

Facebook uses prevalence metrics to prioritize the content they review and manage exposure of harmful content to other parties ('prevalence-based content moderation'). As mentioned previously, the sheer volume of content makes it infeasible for human reviewers to analyze all user-generated content. Instead, a certain percentage is sampled for review. Content that garners more views, shares, likes, etc. is more likely to be sampled for review. Based on the content it samples, Facebook then derives estimates of prevalence[76]. These values are published on a quarterly basis and quantify the proportion of users that were exposed to content that violates Facebook's self-defined Community Standards[77].

In addition to publishing records for transparency's sake, Facebook uses this system to take action on violating content. As an extension of the liability shield provided by Section 230, individual pieces of content may be removed entirely, labeled with a warning questioning the veracity of the post, used to justify the disabling of an account, or shared with a relevant external agency in extreme cases[78]. Furthermore, employees can use trends in prevalence to train Facebook's algorithm to identify different kinds of content and subsequently modify the degree to which different pieces of content are

---

[73] Lyons 2018, Facebook "Understanding the Community Standards Enforcement Report" 2020

[74] Bradford et al. 2019

[75] ibid.

[76] Facebook Community Standards Enforcement Report 2020

[77] Facebook Community Standards 2020

[78] Facebook "Understanding the Community Standards Enforcement Report" 2020

shared via its News Feed, reducing (but not eliminating) the likelihood that harmful content goes viral[79].

4.4 An Analysis of Facebook's Prevalence-Based Content Moderation

Facebook's use of prevalence metrics to moderate content offers a point of departure for federal PBR. However, Facebook's system was optimized for a single firm rather than an entire industry. This section will briefly review key strengths and weaknesses of Facebook's system of PBCM, focusing towards gaps that would need to be addressed for industry-wide federal PBR to be effective.

4.4.1 Strengths of PBCM at Facebook

The following strengths of PBCM at Facebook are qualities that federal PBR should seek to replicate or leverage at a national level.

4.4.1.1 Quantifiable

Although content regulation is inherently a qualitative task, tracking the spread of content is clearly quantitative[80]. Facebook (and likely many other tech firms who traffic in information) are highly effective at tracking and understanding how content spreads on their platform.

4.4.1.2 Leverages Data that Internet Platforms Already Use

The business models of Facebook and other user-generated content platforms funded by advertisers rely heavily on infrastructure that can accurately and robustly measure content exposure at vast scales. Using metrics that are firmly embedded in internet platforms' economic models reduces the need to develop novel metrics and allows regulators to affect existing organizational decision making.

4.4.1.3 Side Steps First Amendment Jurisprudence

Under Section 230, internet companies are broadly protected from liability based on third-party content on their platforms as well as any decisions or actions made to restrict content that firms' feel is "obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable"[81]. This type of regulation avoids the challenges

---

[79] Bickert 2020, Facebook 2019, Facebook "Understanding the Community Standards Enforcement Report" 2020

[80] Facebook "Understanding the Community Standards Enforcement Report" 2020

[81] U.S.C. §§ 230 1996

of a direct restriction on health misinformation by leveraging a platform's own legal ability to moderate content on their sites under Section 230[82].

4.4.2 Weaknesses of PBCM at Facebook

Generally speaking, the weaknesses of PBCM at Facebook arise from the fact that the firm retains exclusive control over implementation. Effective federal PBR would need to build in significant steps to ensure honest reporting and implementation of whatever metrics are ultimately chosen.

4.4.2.1 Selective Transparency of Results and Implementation Verification

The authority to oversee Facebook's content moderation system and report on its efficacy is exclusively internal. Facebook has developed an elaborate and extensive method of developing review rules and triaging novel corner cases[83]. Although Facebook clearly documents their process and is willing to share these general rules, the firm does not share data or allow external auditors to observe the process, let alone reverse decisions[84]. Even Facebook's publicly independent oversight board does not actually have the power to make new review rules or overturn previous decisions[85]. Perhaps most illustrative of this secrecy, a former analysis by the Justice Collaboratory at Yale Law School noted that the efficacy of the system was dependent upon faith that Facebook implemented their system in the manner they described, as none of the experts were permitted to observe the process and Facebook did not provide data to said experts to verify the veracity of their metrics[86]. Without the ability to corroborate the metrics or organizational algorithm, the efficacy and consistency of the process must be trusted on faith alone. Notably, Facebook's proposal for federal PBR offers some elements of transparency, but they are mainly focused on changing content standards, not the fidelity of implementation[87].

Recommendation 8: Regulators should expect companies to act in good faith, and companies should work with regulators to develop effective standards for audit and verification. Trust and transparency are key.

---

[82] Bickert 2020

[83] Koebler and Cox 2018

[84] Bradford et al. 2019

[85] Collins 2020

[86] Bradford et al. 2019

[87] Bickert 2020, Collins 2020

4.4.2.2 Regulation Only Applies to Content Facebook Deems Harmful

Regardless of their efficacy in managing the spread of misinformation, Facebook's prevalence metrics are only used to manage content that the firm deems harmful[88]. Although the firm's current Community Standards do not include health misinformation[89], Facebook has indicated openness to government defined standards, provided that they are clearly defined and can be readily applied to novel situations in a short time frame relative to the traditional judicial process[90].

Expanding out from a single firm, there may not be clear consensus about what constitutes harmful content, either in general (i.e. specific categories) or in particular instances of user-generated content. To implement effective PBR, it is incumbent on the regulating agency to proactively and clearly define what constitutes harmful content in the context of health misinformation; this point is echoed in Facebook's proposal[91].

Recommendation 9: Regulators should establish standards defining harmful content so that internet platforms have a structure within which to develop appropriate tracking and reporting mechanisms of health misinformation.

---

[88] Lyons 2018

[89] Facebook Community Standards Enforcement Report 2020

[90] Bickert 2020

[91] Bickert 2020

5. Who should regulate?

A central question in considering performance-based regulation of health misinformation for internet platforms is who should manage the process. After surveying the history and function of relevant federal agencies, this proposal finds the Food and Drug Administration (FDA) to be most appropriate for the task. Although regulating internet companies would be a new responsibility for the FDA, its mission of advancing public health, demonstrated awareness of the problem of online health misinformation, and a number of additional factors make it the best choice among existing federal agencies.

Although the FDA is best suited for administering our proposed PBR program, other federal agencies have a role to play in managing online health misinformation as well. This section will describe in detail the reasons why the FDA is the best choice for the job of implementing PBR for online health misinformation, and then briefly review the role of other federal agencies – the FCC, FTC, and CDC – in their relation to this effort.

5.1 The FDA Should Implement PBR for Public Health Misinformation

The Food and Drug Administration (FDA) has broad regulatory authority over a number of different areas, including food and drug safety, tobacco, cosmetics, medical devices and more[92]. Part of the FDA's mission is to "[advance] the public health by...helping the public get the accurate, science-based information they need to use medical products and foods to maintain and improve their health"[93]. Across federal agencies, this mission statement is uniquely aligned with the objectives of reducing health misinformation via PBR of internet companies.

Furthermore, the FDA is already aware of the dangers of online health misinformation and actively working to mitigate subsequent harm. In 2014, the FDA published draft guidance for industry entitled "Internet/Social Media Platforms: Correcting Independent Third-Party Misinformation About Prescription Drugs and Medical Devices"[94]. This draft guidance outlines the FDA's current view on how relevant industry stakeholders "should respond, if they choose to respond, to misinformation related to a firm's own FDA-approved or -cleared products when that information is created or disseminated by independent third parties on the Internet or through social media or other technological

---

[92] Office of the FDA Commissioner "What Does FDA Regulate?" 2020

[93] Office of the FDA Commissioner 2018

[94] Center for Drug Evaluation and Research 2019

venues…" The guidance acknowledges the potential for online user-generated content to misrepresent products regulated by the FDA, and advises firms how they may proceed in correcting that content, if they so choose. In particular the FDA offers four pathways that the firm may take for correcting online misinformation:

> "If a firm chooses to correct misinformation, it may do so by correcting misinformation directly on the forum. Alternatively, the firm may provide the corrective information to the independent author for the author to incorporate; the firm may request that the author remove the misinformation or allow comments to be posted; or the firm may request that the site administrator remove the misinformation or allow comments to be posted."[95]

Notably, the FDA offers protection for firms who attempt to correct misinformation "in a truthful and non misleading manner pursuant to [its] recommendations" and "does not intend to object if these voluntary corrections do not satisfy otherwise applicable regulatory requirements, if any." It is encouraging that the FDA is aware of the risks of online health misinformation, and giving companies a green-light to attempt to address the problem without fear of violating FDA standards.

There are a number of other reasons that make the FDA well positioned to administer PBR of online health misinformation.

First, the FDA already has experience administering performance standards, which are codified in the Code of Federal Regulations[96]. Although these standards are for medical devices and intended to "provide reasonable assurance of the safety and effectiveness of the device," there is nonetheless precedent for the FDA administering PBR. Second, the FDA already carries out inspections of companies who manufacture the products that it regulates[97]. Although physical inspections of manufacturing plants, labs, and food processing facilities are vastly different from audits of digital records, they are committed to monitoring and verification. Third, the FDA employs top scientists and medical experts, who it can draw upon to credibly define and create identification metrics for health misinformation. Finally, the FDA is no stranger to technological innovation and digital technology. The Digital Health Center of Excellence within the FDA "empowers digital health stakeholders to advance health care by fostering responsible and high-quality digital health innovation" through areas including digital health policy and technology support and training, artificial intelligence and machine learning, regulatory

---

[95] Center for Drug Evaluation and Research 2019

[96] US CFC Title 21 Part 1010

[97] Office of the FDA Commissioner "What Does FDA Inspect?" 2020

science advancement, and more[98]. While the FDA's current scope of "digital health" does not include health information on internet platforms, the Center demonstrates the technological savvy and forward-mindedness of the FDA when it comes to issues at the intersection of health and technology.

5.2 The Role of Other Regulatory Agencies

Although the FDA is the best choice for implementing PBR for health misinformation, other regulatory agencies may play a role influencing online health misinformation, and directly or indirectly impact the FDA's implementation efforts.

5.2.1 FTC

The Federal Trade Commission (FTC) has a long history of regulatory action dealing with false or deceptive advertising which harms consumers, and its website states that it looks "especially closely" at advertising claims related to consumer health[99]. The truth-in-advertising laws enforced by the FTC include online advertising and what may be considered online health misinformation. For example, in 2011, the FTC requested federal courts take action against companies operating "fake news sites" advertising acai berry diets using misleading advertising and false claims[100]. However, the scope of the FTCs historic regulation of online content is restricted to commercial content. It is unclear what capacity, if any, the FTC has in regulating non-commercial content. If the FDA is responsible for broadly regulating online health misinformation, the FTC will still play an important role in regulating health misinformation related to commercial products, and taking enforcement actions against those responsible for it.

5.2.2 FCC

The Federal Communications Commission (FCC) "regulates interstate and international communications," and "is the United States' primary authority for communications law, regulation and technological innovation"[101]. Although the FCC enforces certain restrictions on broadcasted speech, the FCC does not appear to have taken regulatory actions regarding online speech or misinformation. However, actions by the FCC

---

[98] Center for Devices and Radiological Health 2020

[99] Federal Trade Commission 2013

[100] Federal Trade Commission 2011

[101] Federal Communications Commission 2010

suggest that it may open rulemaking which could affect Section 230[102]. This could have substantial implications for PRB of online health information, which depend on Section 230 liability protection to allow internet companies to self-moderate content. In this way, the FCC may indirectly influence PBR for online health information administered by the FDA, as its implementation would depend on Section 230 and its associated rules.

5.2.3 CDC

The Center for Disease Control and Prevention (CDC), along with the FDA, is a component of the Department of Health and Human Services. Unlike the FDA, FTC, and FCC, however, the CDC is not a regulatory agency and has no regulatory authority. As such, it is not an appropriate candidate to administer PRR of public health misinformation. However, its mission – “collaborating to create the expertise, information, and tools that people and communities need to protect their health through health promotion, prevention of disease, injury and disability, and preparedness for new health threats"[103] – is highly relevant to the PBR’s goals, and it could still play an important role providing expert guidance, knowledge, and fact-finding capacity to the FDA. This collaborative capacity is already institutionalized. A 2017 memorandum of understanding between the CDC and FDA states that “each agency will coordinate and collaborate with the other agency to protect and improve the public health” and outlines specific procedures for how the two agencies should collaborate[104].

Recommendation 10: The FDA should implement performance-based regulation as part of a deliberate effort across federal agencies to reduce harmful online health misinformation.

---

[102] Federal Communications Commission 2020

[103] Office of the FDA Commissioner 2019

[104] ibid.

6. Demonstrating a Working Model

Having established that the FDA is the appropriate regulatory agency, this section will outline a model for PBR as applied to health misinformation. Prevalence will serve as the primary performance metrics, since platforms like Facebook have already demonstrated the capacity to measure prevalence, as noted in Section 4.4.
Building off the examples described in Section 2, the anti-vaccination movement and COVID-19 crisis will be examined as case studies to demonstrate how this model would handle specific “infodemics”. This example will illustrate how the specific mechanisms of PBR as applied to health misinformation on social media platforms might function, with the understanding that multiple effective implementations exist.

6.1 A Hypothetical Model of PBR for Online Health Misinformation

As described in Section 4.1, political scientist Peter May coined the three necessary dimensions of PBR: characterization of outcomes, characterization of desired level of achievement, and performance assessment. This section will follow this framework.

6.1.1 Characterizing Outcomes

For our model, the high-level goal is to minimize the harm caused by health misinformation, thereby improving public health outcomes. At a lower level, the goal is to reduce the number of interactions a user has with online posts containing health misinformation. These goals apply regardless of the particular health issue in question.

Yet, even if regulation is successful, the outcomes for specific public health scenarios may feel qualitatively different. As noted in Section 2.1.1, there are two types of health misinformation: rapidly developing (as is the case with COVID-19) and well-established (as is the case with vaccines). With COVID-19, the end outcome may simply be a more stable information environment. Less misinformation goes viral, so people are more receptive to true information when it becomes available. With the anti-vaxx movement, regulators should aim to reduce the number of interactions conspiracy theories get but also make sure the correct information is communicated. (In this case, the correct information is that the benefits of vaccination far outweigh the harms in the vast majority of cases.) It should be noted that our model tackles the former, but the latter must be addressed in other ways.

6.1.2 Characterizing Desired Level of Achievements

6.1.2.1 Deciding What Constitutes Misinformation

The first step in regulating misinformation is to separate fact from fiction. To achieve this goal, the FDA could hire a team of medical professionals and collaborate with the CDC (as noted in Section 5.2.3) to publish documents communicating the accepted medical consensus on commonly misunderstood health topics. The documents would also directly refute misinformation claims, like the notion that vaccines cause children to have autism or that COVID-19 vaccines contain microchips. Importantly, the documents would be written with the goal of fact-checking in mind, as opposed to general advice for the public. There would be a high degree of scientific certainty, backed by appropriate references to scientific literature, for any information that is labelled as a "fact."

Platforms can then use these documents in their content moderation process as a clear standard of what constitutes health misinformation. They can also send emerging, dubious health claims to the FDA team to specifically fact check, particularly if the platform notices an unaddressed claim beginning to gain significant traction on their site.

These documents can be routinely updated to incorporate new information as research continues to develop. When it comes to well-researched and understood medical issues, such as the benefit of vaccines, these documents can be updated with monthly frequency. But with ongoing, emergency-level crises like COVID-19, the documents should be updated with a minimum of weekly frequency, as the World Health Organization currently does on their COVID-19 webpage[105]. This creates flexibility based on the situation and allows the agency to take a contextual approach that is informed by the amount of misinformation circulating.

Fact determination may seem like a laborious process, and the FDA will likely have to hire a substantial team of reviewers. Still, it will likely end up being the cost-effective approach in terms of public health. For example, if platforms can stifle the circulation of COVID-19 misinformation such that 70% of Americans get vaccinated, the U.S. will save millions on healthcare costs otherwise incurred by the pandemic continuing. In June 2020, the Congressional Budget Office estimated that the COVID-19 pandemic will end up costing the U.S. economy 8 trillion dollars over the next ten years[106]. Researchers estimated in 2016 that Americans who refuse to get vaccinated cost the

---

[105] World Health Organization "Advice for the Public on COVID-19" 2020

[106] Ziv 2020

U.S. economy more than seven billion dollars per year[107]. Enabling the FDA to perform this vital task would be a prudent investment. Moreover, as algorithmic misinformation detection and prevention continues to improve, especially as more data is collected in these areas, new technologies are expected to reduce the workload of human reviewers in coming years.

6.1.2.2 Quantifying Desired Outcomes

To quantify the goal, platforms can use a prevalence threshold, as described in Section 4.3.2. While the exact method of calculation differs for each platform, prevalence measures the amount of overall user interaction with an individual post.

For example:

- For Facebook, this would include the number of views, clicks, likes, and/or reactions.
- For YouTube, this would include the number of comments, likes/dislikes, and views, as well as the duration of viewing.
- For Twitter, this would include the number of retweets, likes, and replies.
- For Reddit, this would include the number of upvotes and replies.
- For Instagram, this would include the number of likes and views.

Prevalence can be calculated as a relative percentage, indicating the proportion of user interaction with health misinformation with respect to the amount of user interaction with all content on a platform This number can be calculated for a certain range of content. For instance, a daily prevalence threshold would only take into account content posted on a specific day for a specific site. Since platforms are already keeping track of prevalence metrics internally for advertising purposes, platforms can likely leverage existing software infrastructure to generate these numbers.

To illustrate, let's consider a sample of 100 Facebook posts from December 8, 2020. Suppose 10 posts were deemed to be health misinformation. The group of posts containing health misinformation received a total of 45,000 interactions. Meanwhile, the 90 other posts received a total of 55,000 interactions. That would mean that, for this sample, the prevalence of health misinformation was 45%.

A prevalence threshold would be a clear, quantified goal set by regulators. For example, if health misinformation on Facebook in 2020 was estimated to have a prevalence of 45%, regulators can set a performance target of 35% prevalence for Facebook in 2021.

---

[107] Keefe 2016

Platforms that meet their prevalence threshold will be rewarded with incentives, while platforms that fail will be subject to penalties.

Some possible incentives could be tax breaks, which would be coordinated with the IRS, as well as grants for platforms to conduct research relevant to prevalence or misinformation research. Governments could also audit platforms less frequently, given a history of good performance.

Potential penalties could include a fine tied to the prevalence misinformation above the prevalence threshold. If health misinformation on Twitter was calculated to have a prevalence of 23%, and the prevalence threshold was 20%, they would be fined only for the 3% of user interactions above the 20% threshold.

Unlike with fact determination, defining performance outcomes follows the same process regardless of the particular health issue. In this step of the model, anti-vaxx posts and COVID-19 conspiracy posts would follow the same procedure.

6.1.3 Assessing Performance

As this approach depends upon the reliability of the metrics used to calculate prevalence, steps must also be taken to verify the engagement numbers. At least two methods can certify the veracity of internet platforms' metrics and reporting systems. First, platforms should provide any source code that is relevant to the calculation of metrics. The FDA will review the code and validate that the method of calculating engagement numbers is fair and accurate. Second, the source code used to generate metrics for prevalence calculations must be the same as the source code used to generate metrics provided to advertisers. This ensures that social media platforms do not categorically reduce their interaction metrics across the board in an effort to meet the prevalence threshold, as doing so would mean losing money from their advertisers.

A similar source code auditing process is currently used by the U.S. military, when reviewing software supplied under defense contracts[108]. Platforms may be understandably wary of opening their source code, which is high-value. To address these concerns, regulators can only allow the code reviews to take place in secure facilities that can prevent code from being copied or altered; typically, this means a room where reviewers can inspect the code, but do little else.

A code audit would be a new responsibility for the FDA, requiring an expansion of the agency's technical skillset. But, as medical devices (which are regulated by the FDA)

---

[108] MITRE 2013

continues to grow as an industry, the FDA will have to hire more technologists regardless. As mentioned in Section 5.1, the FDA has already acknowledged this need, with the creation of the Digital Health Center of Excellence in 2020[109]. The team of computer scientists necessary for the code audit align with the goals of the Center.

In this step of the model, anti-vaxx posts and COVID-19 conspiracy posts would follow the same procedure.

6.2 Addressing Potential Weaknesses of PBR

The preceding sections provided a model of how PBR relying upon prevalence metrics might function in practice. The following section will assess the aforementioned model in the context of the Coglianese's seven dangers of PBR[110].

6.2.1 Avoiding Conflict

PBR depends on cooperative efforts by platforms and governing agencies at each step of the process. By giving platforms a seat at the table at each stage – defining health misinformation, declaring desired performance outcomes, and assessing performance – regulators and platforms have the opportunity to work out definitions, goals, and methods of assessment that everyone can support. Afterall, the problem of health misinformation is accepted widely in industry to be a serious problem that is too large and complicated for platforms to tackle alone. Cooperative efforts gives platforms the opportunity to leverage governmental resources and authority on a problem that very much affects their bottom line.

6.2.2 Unintended Consequences from Unrecognized Tradeoffs

Regulating platforms may lead to unintended consequences, such as users leaving established platforms and finding other avenues for sharing health misinformation. However, the flexible nature of PBR can incorporate these new avenues in their regulatory scope. As fringe sites gain popularity among health misinformation consumers, the FDA can easily step in and set a prevalence threshold or some other type of regulation. As long as regulators remain vigilant in their investigative efforts when identifying health misinformation, as described in Section 6.1.2.1, they can find the areas where health misinformation is gaining the most attention and respond appropriately.

---

[109] Center for Devices and Radiological Health 2020

[110] Coglianese 2017

Of course, there remains risk of unforeseen consequences that this paper does not specifically address, by definition of being unforeseen. It is important for regulators to take advantage of the flexibility of PBR to recover from these setbacks. While we are suggesting a prevalence-based approach for this model, future events may suggest that a different metric or combination of metrics are more suited for combatting health misinformation. Regulators must be prepared to redefine performance targets in this case, while allowing enough time for platforms to adjust their approaches to realign with the new goals. As technology continues to develop, and as health misinformation continues to proliferate in novel ways, it is important that regulators also continue to adapt their approaches to maximize efficacy.

6.2.3 Regulatory Red Tape

Industry often resists regulation due to the perceived bureaucratic constraints regulations can introduce. In this case, though, platforms may actually improve their internal processes and benefit from regulatory efforts. First, by taking advantage of the regulator's team of medical fact checkers, platforms can reduce the amount of time, money, and labor they currently spend fact-checking on their own. Second, in working with regulators, platforms can distribute some of the responsibility for reducing misinformation, improving their standing from a public relations perspective. When an outrageous health misinformation claim inevitably slips through the cracks, platforms often face massive scrutiny[111]. However, with regulation in place, the scrutiny can be focused on regulators' failure to successfully incentivize ideal outcomes instead of solely falling on the shoulders of the platforms.

6.2.4 Limited Discretion and Flexibility

Some may be concerned that it will not be technically feasible to meet the performance goals established. This is unlikely, given that platforms have already implemented processes for reducing health misinformation and, in some cases, have successfully tested the relevant features. The lack of progress in reducing health misinformation is primarily a result of prioritizing profits rather than technical infeasibility. For instance, most platforms have a search function and can remove misinformation from high-ranking search results and search bar auto-completions. In addition, platforms can also remove misinformation from recommended viewing sections of their sites, such as the "Up Next" function on YouTube or the top of the Newsfeed on Facebook. This would decrease the exposure of users to misinformation without outright banning such content, but also results in users spending less time on platforms, as studies have

[111] Loiaconi 2020

shown[112]. Since user attention (particularly on advertisements) is an important moneymaker, platforms have decided against implementing misinformation-targeted features on a long-term basis.[113] Regulatory compellment will incentivize platforms to reconsider those decisions.

6.2.5 Potential for Fraud or Evasive Behavior

In this model, the risk of fraud primarily lies in the prevalence metrics. If metrics were to be inaccurate or falsified, the model would greatly suffer. However, the validation methods mentioned previously – comparing the numbers to those given to advertisers and examining the source code – provide a sufficient level of security. Moreover, as the regulations are based on incentives rather than criminal prosecution, it is unlikely that platforms would engage in fraudulent activity (an actual crime) to avoid regulations with no criminal penalties.

6.2.6 "Teaching to the Test"

It may be possible to "game" the assessment system by solely targeting health topics that are easier to detect. If this is sufficient to meet the prevalence threshold, there is no incentive for platforms to reduce misinformation on health topics that may be more difficult to detect but still causes serious harm. To avoid this, regulators must be ready to adopt prevalence thresholds for subcategories within the realm of health misinformation. For example, platforms may need incentives not just to reduce health misinformation overall, but also to reduce anti-vaxx content specifically and COVID-19 misinformation specifically.

6.2.7 Incorrect Proxies or Causal Factors

There is a concern that prevalence is not actually the best metric for better public health outcomes. Does reducing prevalence of misinformation content actually stagnate its spread? Studies suggest that it does. In November of 2020, Facebook tested a system that demoted content that was algorithmically determined to have a high probability of containing misinformation. This effectively reduced the prevalence of those posts. As a result, the top links shared on Facebook contained more reputable news stories and less misinformation claims[114]. As such, prevalence of misinformation can serve as a reasonably good proxy for better public health outcomes. However, in the case that future studies reveal the opposite, regulators can adopt new performance targets based

---

[112] Ortutay 2020

[113] Roose et al. 2020

[114] Roose et al. 2020

on other metrics. The flexibility of PBR is key here, as it allows regulators to update their directives without significantly disrupting regulatory pathways. Regardless of the specific performance target, the FDA will still be responsible for defining, monitoring, and auditing the proliferation of health misinformation. Additionally, the model this proposal advocates for may also be implemented in tandem with other interventions, such as a greater investment in digital media literacy efforts, improved health education, and greater access to healthcare.

Table 2

The seven dangers of PBR and the associated risks present in our model

Danger	**Risk level**	**Key Points**
Failure to Avoid Conflict	Medium	● Risk of conflict between platforms and regulators. ● Can be mitigated by giving platforms seat at the table during each stage of PBR process
Unintended Consequences	High	● Impossible to predict every consequence of PBR, potential for unforeseen consequences resulting from unrecognized tradeoffs ● Regulators must be prepared to adapt performance targets according to issues that arise
Regulatory Red Tape	Low	● Industry may resist regulations because of perceived bureaucracy ● In this case, platforms are actually reducing their workload by offloading some fact-checking responsibilities to regulators
Limited Discretion and Flexibility	Low	● Regulators may impose performance standards or reporting requirements that are not technically feasible ● Platforms already have most of necessary infrastructure in place but lack incentive to treat the problem as high priority
Potential for fraud or evasive behavior	Medium	● Platforms may provide fraudulent numbers ● Validation of numbers within audit provides security, and lack of criminal prosecution makes fraudulent behavior unlikely to be worth the risk
“Teaching to the Test”	High	● Companies may focus their content moderation efforts on certain types of health misinformation in order to meet performance targets ● May reduce incentives to deal with more difficult types of health misinformation to identify ● Regulators must be ready to adopt prevalence thresholds within subcategories of health misinformation

Incorrect Proxies or Causal Factors	Low	● Possible that performance standards on internet platforms may be an incorrect proxy for better public health outcomes ● Current research refutes this ● PBR just one leg of many regulations aiming to improve public health

Despite known theoretical weaknesses to PBR, our model is well-equipped to handle each potential danger and respond appropriately. Still, regulators will ideally anticipate shortcomings in advance. To aid in this preparation, Table 2 summarizes each of the dangers, as well the relative risk level.

7. Conclusion

Health misinformation seems to have become an accepted part of the digital world. But it doesn't have to stay that way. Section 230 offers internet companies the immunity they need to take action on harmful content on their platforms, and rapid advances in digital technology have made it possible to identify and address misinformation at scale. However, these factors have not been sufficient to minimize the harm caused by online health misinformation. Novel regulatory mechanisms are needed.

Performance-based regulation is one regulatory framework for addressing health misinformation that, if implemented properly, could work for companies, regulators, and most importantly, the public. This proposal has identified the following recommendations which would help ensure federal regulators implement successful PBR of online health misinformation:

1) Regulators must take a proactive, ongoing role in defining what constitutes online health misinformation by partnering with medical experts and professionals.
2) Regulators should focus on moderating the dissemination of health misinformation among internet users.
3) Regulators should target social media platforms specifically to mitigate the spread of health misinformation.
4) Regulators should incentivize internet platforms to leverage their Section 230 immunity to mitigate the spread of health misinformation.
5) Regulators should implement performance-based regulation to address the harms of online health misinformation.
6) Regulators should recognize differences among internet platforms in size, design, and company policy, and implement performance metrics which keep those differences in mind. PBR is not a one-size-fits-all solution.
7) Regulators should minimize the "dangers" of performance-based regulation by incorporating relevant stakeholders such as companies, technical experts, and the public into the PBR design process.
8) Regulators should expect companies to act in good faith, and companies should work with regulators to develop effective standards for audit and verification. Trust and transparency are key.

9) Regulators should establish standards defining harmful content so that internet platforms have a structure within which to develop appropriate tracking and reporting mechanisms of health misinformation.
10) The FDA should implement performance-based regulation as part of a deliberate effort across federal agencies to reduce harmful online health misinformation.

The form of performance-based regulation that this proposal advocates for is only possible if companies maintain their ability to self-moderate content under Section 230. Ongoing discussions by both Republicans and Democrats suggest that this ability may be in jeopardy[115]. Politicians on both sides of the aisle would be well advised to apply the existing power of Section 230 in a new way, and to advance performance-based regulation of online health misinformation as a potentially effective and powerful tool for keeping American's informed, safe, and healthy.

[115] Pietsch 2020

8. References

Bao, Yanping, Yankun Sun, Shiqiu Meng, Jie Shi, and Lin Lu. “2019-NCoV Epidemic: Address Mental Health Care to Empower Society.” *The Lancet* 395, no. 10224 (February 22, 2020): e37–38. https://doi.org/10.1016/S0140-6736(20)30309-3.

Basch, C. H., P. Zybert, R. Reeves, and C. E. Basch. “What Do Popular YouTubeTM Videos Say about Vaccines?” *Child: Care, Health and Development* 43, no. 4 (2017): 499–503. https://doi.org/10.1111/cch.12442.

BBC Monitoring. “How Anti-Vaccine Movements Threaten Global Health.” *BBC News*, June 20, 2019, sec. Health. https://www.bbc.com/news/health-48585036.

BBC News. “Norwich Filipino Nurse ‘dumbfounded’ by Coronavirus Abuse.” *BBC News*, February 11, 2020, sec. Norfolk. https://www.bbc.com/news/uk-england-norfolk-51465604.

Belluz, Julia. “Minnesota Is Fighting Its Largest Measles Outbreak in Nearly 30 Years. Blame Vaccine Deniers.” Vox, May 8, 2017. https://www.vox.com/2017/5/8/15577316/minnesota-measles-outbreak-explained.

Belluz, Julia. “The Real Reason Measles Cases Spiked in 2014: The Ohio Amish.” Vox, January 29, 2015. https://www.vox.com/2015/1/29/7929791/measles-outbreak-2014.

Belluz, Julia. “Why Minnesota Lost a Battle against Anti-Vaccine Campaigners.” Vox, May 10, 2017. https://www.vox.com/science-and-health/2017/5/10/15591410/minnesota-measles-anti-vaccine-public-health-lessons.

Bickert, Monica. “Charting A Way Forward: Online Content Regulation.” Facebook, February 1, 2020. https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf.

Bowman, C. Douglas, and Margaret S. McKay. “Performance-Based Regulation (PBR) Delivers Maximum Benefits - POWERGrid International.” Power Grid International. Accessed December 8, 2020. https://www.power-grid.com/smart-grid/performance-based-regulation-pbr-delivers-maximum-benefits/#gref.

Bradford, Ben, Florian Grisel, Tracey L Meares, Emily Owens, Baron L Pineda, Jacob N Shapiro, Tom R Tyler, and Danieli Evans Peterman. “Report Of The Facebook Data Transparency Advisory Group.” The Justice Collaboratory at Yale Law School, April 2019. https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf.

Brennen, J. Scott, Felix M. Simon, Philip N. Howard, and Rasmus Kleis Nielsen. “Types, Sources, and Claims of COVID-19 Misinformation.” Reuters Institute for the Study of Journalism, April 7, 2020. http://www.primaonline.it/wp-content/uploads/2020/04/COVID-19_reuters.pdf.

Bundorf, M. Kate, Todd H. Wagner, Sara J. Singer, and Laurence C. Baker. “Who Searches the Internet for Health Information?” *Health Services Research* 41, no. 3p1 (2006): 819–36. https://doi.org/10.1111/j.1475-6773.2006.00510.x.

Burki, Talha. “The Online Anti-Vaccine Movement in the Age of COVID-19.” *The Lancet Digital Health* 2, no. 10 (October 1, 2020): e504–5. https://doi.org/10.1016/S2589-7500(20)30227-2.

CDC. “Measles Cases and Outbreaks.” Centers for Disease Control and Prevention, December 2, 2020. https://www.cdc.gov/measles/cases-outbreaks.html.

CDC. “Measles Complications.” Centers for Disease Control and Prevention, November 5, 2020. https://www.cdc.gov/measles/symptoms/complications.html.

Center for Devices and Radiological Health. “About the Digital Health Center of Excellence.” *FDA*, September 22, 2020. https://www.fda.gov/medical-devices/digital-health-center-excellence/about-digital-health-center-excellence.

Center for Drug Evaluation and Research. “Internet/Social Media Platforms: Correcting Independent Third-Party Misinformation About Prescription Drugs and Medical Devices.” U.S. Food and Drug Administration. FDA, February 8, 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/internetsocial-media-platforms-correcting-independent-third-party-misinformation-about-prescription.

Chou, Wen-Ying Sylvia, April Oh, and William M. P. Klein. “Addressing Health-Related Misinformation on Social Media.” *JAMA* 320, no. 23 (December 18, 2018): 2417–18. https://doi.org/10.1001/jama.2018.16865.

Coglianese, Cary, Jennifer Nash, and Todd Olmstead. “PERFORMANCE-BASED REGULATION: PROSPECTS AND LIMITATIONS IN HEALTH, SAFETY, AND ENVIRONMENTAL PROTECTION.” *Administrative Law Review* 55, no. 4 (2003): 705–29.

Coglianese, Cary. “The Limits of Performance-Based Regulation.” *University of Michigan Journal of Law Reform* 50, no. 3 (March 1, 2017): 525–63.

Collins, Damian. “Facebook Must Not Be Allowed to Dictate How It Gets Regulated.” *Wired UK*, February 20, 2020. https://www.wired.co.uk/article/facebook-eu-regulation.

Commissioner, Office of the. “What Does FDA Inspect?” *FDA*, February 19, 2020. https://www.fda.gov/about-fda/fda-basics/what-does-fda-inspect.

Communications Decency Act of 1996, (CDA), Pub. L. No. 104-104 (Tit. V), 110 Stat. 133 (Feb. 8, 1996), codified at 47 U.S.C. §§ 230.

Dahl, Stephan, and Lynne Eagle. “Empowering or Misleading? Online Health Information Provision Challenges.” *Marketing Intelligence & Planning* 34, no. 7 (January 1, 2016): 1000–1020. https://doi.org/10.1108/MIP-07-2015-0127.

Diamond, John, and James Primm. “Rediscovering Traditional Tort Typologies to Determine Media Liability for Physical Injuries: From the Mickey Mouse Club To

Hustler Magazine." *Hastings Communications and Entertainment Law Journal* 10, no. 4 (January 1, 1988): 969. Available at: https://repository.uchastings.edu/hastings_comm_ent_law_journal/vol10/iss4/

Donzelli, Gabriele, Giacomo Palomba, Ileana Federigi, Francesco Aquino, Lorenzo Cioni, Marco Verani, Annalaura Carducci, and Pierluigi Lopalco. "Misinformation on Vaccination: A Quantitative Analysis of YouTube Videos." *Human Vaccines & Immunotherapeutics* 14, no. 7 (July 3, 2018): 1654–59. https://doi.org/10.1080/21645515.2018.1454572.

Facebook. "Community Standards." Facebook, 2020. https://www.facebook.com/communitystandards/.

Facebook. "Community Standards Enforcement Report." Facebook, 2020. https://transparency.facebook.com/community-standards-enforcement.

Facebook. "Measuring Prevalence of Violating Content on Facebook," May 23, 2019. https://about.fb.com/news/2019/05/measuring-prevalence/.

Facebook. "Understanding the Community Standards Enforcement Report." Facebook, 2020. https://transparency.facebook.com/community-standards-enforcement/guide.

Federal Communications Commission. "The FCC's Authority to Interpret Section 230 of the Communications Act." Federal Communications Commission, October 21, 2020. https://www.fcc.gov/news-events/blog/2020/10/21/fccs-authority-interpret-section-230-communications-act.

Federal Communications Commission. "What We Do." Federal Communications Commission, November 22, 2010. https://www.fcc.gov/about-fcc/what-we-do.

Federal Trade Commission. "FTC Seeks to Halt 10 Operators of Fake News Sites from Making Deceptive Claims About Acai Berry Weight Loss Products." Federal Trade Commission, April 19, 2011. https://www.ftc.gov/news-events/press-releases/2011/04/ftc-seeks-halt-10-operators-fake-news-sites-making-deceptive.

Federal Trade Commission. "Truth In Advertising." Federal Trade Commission, August 5, 2013. https://www.ftc.gov/news-events/media-resources/truth-advertising.

Frenkel, Sheera, Davey Alba, and Raymond Zhong. "Surge of Virus Misinformation Stumps Facebook and Twitter." *The New York Times*, March 8, 2020, sec. Technology. https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html.

Fritch, John W., and Robert L. Cromwell. "Evaluating Internet Resources: Identity, Affiliation, and Cognitive Authority in a Networked World." *Journal of the American Society for Information Science and Technology* 52, no. 6 (2001): 499–507. https://doi.org/10.1002/asi.1081.

Gandhi, Lakshmi. "A History Of 'Snake Oil Salesmen.'" NPR.org, August 26, 2013. https://www.npr.org/sections/codeswitch/2013/08/26/215761377/a-history-of-snake-oil-salesmen.

Garg, Neetika, Anand Venkatraman, Ambarish Pandey, and Nilay Kumar. “YouTube as a Source of Information on Dialysis: A Content Analysis.” *Nephrology* 20, no. 5 (2015): 315–20. https://doi.org/10.1111/nep.12397.

Goodman, Jack, and Flora Carmichael. “Coronavirus: Bill Gates ‘Microchip’ Conspiracy Theory and Other Vaccine Claims Fact-Checked.” *BBC News*, May 29, 2020, sec. Reality Check. https://www.bbc.com/news/52847648.

Keefe, Patrick. “The Cost of People Who Don’t Get Vaccinations.” Healthline, October 24, 2016. https://www.healthline.com/health-news/vaccinations-cost-us-economy-7-billion-a-year.

Kim, Alfred H.J., Jeffrey A. Sparks, Jean W. Liew, Michael S. Putman, Francis Berenbaum, Alí Duarte-García, Elizabeth R. Graef, et al. “A Rush to Judgment? Rapid Reporting and Dissemination of Results and Its Consequences Regarding the Use of Hydroxychloroquine for COVID-19.” *Annals of Internal Medicine* 172, no. 12 (March 30, 2020): 819–21. https://doi.org/10.7326/M20-1223.

Kinsora, A., K. Barron, Q. Mei, and V. G. V. Vydiswaran. “Creating a Labeled Dataset for Medical Misinformation in Health Forums.” In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 456–61, 2017. https://doi.org/10.1109/ICHI.2017.93.

Koebler, Jason, and Joseph Cox. “The Impossible Job: Inside Facebook’s Struggle to Moderate Two Billion People.” Motherboard, August 23, 2018. https://www.vice.com/en/article/xwk9zd/how-facebook-content-moderation-works.

Li, Heidi Oi-Yee, Adrian Bailey, David Huynh, and James Chan. “YouTube as a Source of Information on COVID-19: A Pandemic of Misinformation?” *BMJ Global Health* 5, no. 5 (May 1, 2020): e002604. https://doi.org/10.1136/bmjgh-2020-002604.

Lo, Nathan C., and Peter J. Hotez. “Public Health and Economic Consequences of Vaccine Hesitancy for Measles in the United States.” *JAMA Pediatrics* 171, no. 9 (September 1, 2017): 887–92. https://doi.org/10.1001/jamapediatrics.2017.1695.

Logie, Carmen H., and Janet M. Turan. “How Do We Balance Tensions Between COVID-19 Public Health Responses and Stigma Mitigation? Learning from HIV Research.” *AIDS and Behavior* 24, no. 7 (July 1, 2020): 2003–6. https://doi.org/10.1007/s10461-020-02856-8.

Loiaconi, Stephen. “Social Media Sites under Scrutiny as They Struggle with Election Misinformation.” KUTV, November 18, 2020. https://kutv.com/news/nation-world/social-media-sites-under-scrutiny-as-they-struggle-with-election-misinformation.

Love, Jennifer S., Adam Blumenberg, and Zane Horowitz. “The Parallel Pandemic: Medical Misinformation and COVID-19.” *Journal of General Internal Medicine* 35, no. 8 (August 1, 2020): 2435–36. https://doi.org/10.1007/s11606-020-05897-w.

Lyons, Tessa. “Hard Questions: What’s Facebook’s Strategy for Stopping False News?” *About Facebook* (blog), May 23, 2018. https://about.fb.com/news/2018/05/hard-questions-false-news/.

Madden, Mary, and Susannah Fox. “Finding Answers Online in Sickness and in Health.” *Pew Research Center: Internet, Science & Tech* (blog), May 2, 2006. https://www.pewresearch.org/internet/2006/05/02/finding-answers-online-in-sickness-and-in-health/.

Madden, Mary, and Susannah Fox. “Health Online 2013.” *Pew Research Center: Internet, Science & Tech* (blog), January 15, 2013. https://www.pewresearch.org/internet/2013/01/15/health-online-2013/.

May, Peter J. “Performance-Based Regulation and Regulatory Regimes: The Saga of Leaky Buildings.” *Law & Policy* 25, no. 4 (2003): 381–401. https://doi.org/10.1111/j.0265-8240.2003.00155.x.

McKee, Martin, and Pascal Diethelm. “How the Growth of Denialism Undermines Public Health.” *BMJ* 341 (December 14, 2010): c6950. https://doi.org/10.1136/bmj.c6950.

Miller, Michele, Tanvi Banerjee, Roopteja Muppalla, William Romine, and Amit Sheth. “What Are People Tweeting About Zika? An Exploratory Study Concerning Its Symptoms, Treatment, Transmission, and Prevention.” *JMIR Public Health and Surveillance* 3, no. 2 (June 19, 2017). https://doi.org/10.2196/publichealth.7157.

Minnesota Department of Health. “News Release: Health Officials Declare End of Measles Outbreak.” Minnesota Department of Health, August 25, 2017. https://www.health.state.mn.us/news/pressrel/2017/measles082517.html.

MITRE. “Secure Code Review,” *Systems Engineering Guide*, August 28, 2013. https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/systems-engineering-for-mission-assurance/secure-code-review.

Myers, Joe. “Anti-Vaccine Activists Spark a US State’s Worst Measles Outbreak in Decades.” World Economic Forum, May 19, 2017. https://www.weforum.org/agenda/2017/05/a-us-measles-outbreak-is-being-blamed-on-anti-vac-activists/.

Office of the FDA Commissioner. “MOU 225-14-017.” *FDA*, June 17, 2019. https://www.fda.gov/about-fda/domestic-mous/mou-225-14-017.

Office of the FDA Commissioner. “What Does FDA Regulate?” *FDA*, August 3, 2020. https://www.fda.gov/about-fda/fda-basics/what-does-fda-regulate.

Office of the FDA Commissioner. “What We Do.” FDA. FDA, November 3, 2018. https://www.fda.gov/about-fda/what-we-do.

Organisation for Economic Co-operation and Development. “Regulatory Policies in OECD Countries: From Interventionism to Regulatory Governance.” Text, October 14, 2002. https://www.oecd-ilibrary.org/governance/regulatory-policies-in-oecd-countries_9789264177437-en.

Ortega-Sanchez, Ismael R., Maya Vijayaraghavan, Albert E. Barskey, and Gregory S. Wallace. “The Economic Burden of Sixteen Measles Outbreaks on United States Public Health Departments in 2011.” *Vaccine* 32, no. 11 (March 5, 2014): 1311–17. https://doi.org/10.1016/j.vaccine.2013.10.012.

Ortutay, Barbara. “Facebook Bans Anti-Vaccination Ads but Not Antivax Posts.” ABC News, October 13, 2020. https://abcnews.go.com/Health/wireStory/facebook-bans-anti-vaccination-ads-antivax-posts-73594039.

Oyeyemi, Sunday Oluwafemi, Elia Gabarron, and Rolf Wynn. “Ebola, Twitter, and Misinformation: A Dangerous Combination?” *BMJ* 349 (October 14, 2014): g6178. https://doi.org/10.1136/bmj.g6178.

O’Sullivan, Donie, Kaya Yurieff, and Kelly Bourdet. “Misinformation Watch on the 2020 Election.” CNN, December 8, 2020. https://www.cnn.com/business/live-news/election-2020-misinformation/index.html.

Pietsch, Bryan. “Trump and Biden Both Want to Revoke Section 230, but for Different Reasons.” Business Insider. Accessed December 8, 2020. https://www.businessinsider.com/trump-biden-want-to-revoke-section-230-for-different-reasons-2020-5.

Rajkumar, Ravi Philip. “COVID-19 and Mental Health: A Review of the Existing Literature.” *Asian Journal of Psychiatry* 52 (August 1, 2020): 102066. https://doi.org/10.1016/j.ajp.2020.102066.

Reddit. “Content Policy - Reddit.” Reddit, 2020. https://www.redditinc.com/policies/content-policy.

Reiss, Dorit, and John Diamond. “Measles and Misrepresentation in Minnesota: Can There Be Liability for Anti-Vaccine Misinformation That Causes Bodily Harm?” *San Diego Law Review* 56, no. 3 (October 15, 2019): 531.

Richtel, Matt. “W.H.O. Fights a Pandemic Besides Coronavirus: An ‘Infodemic.’” *The New York Times*, February 6, 2020, sec. Health. https://www.nytimes.com/2020/02/06/health/coronavirus-misinformation-social-media.html.

Roberts, Michelle. “Vaccines: Low Trust in Vaccination ‘a Global Crisis.’” *BBC News*, June 19, 2019, sec. Health. https://www.bbc.com/news/health-48512923.

Rocha, Veronica, Brian Ries, and Amanda Wills. “Live: Mark Zuckerberg Testifies before Congress.” CNN, April 10, 2018. https://www.cnn.com/politics/live-news/mark-zuckerberg-testifies-congress/index.html.

Roose, Kevin, Mike Isaac, and Sheera Frenkel. “Facebook Struggles to Balance Civility and Growth - The New York Times.” The New York Times, November 24, 2020. https://www.nytimes.com/2020/11/24/technology/facebook-election-misinformation.html.

Silber, Denise. “[Medicine 2.0: the stakes of participatory medicine].” *Presse Medicale (Paris, France: 1983)* 38, no. 10 (October 2009): 1456–62. https://doi.org/10.1016/j.lpm.2009.06.011.

Spring, Marianna. “Coronavirus: The Human Cost of Virus Misinformation.” *BBC News*, May 27, 2020, sec. Stories. https://www.bbc.com/news/stories-52731624.

Spring, Marianna. “Social Media Firms Fail to Act on Covid-19 Fake News.” *BBC News*, June 3, 2020, sec. Technology. https://www.bbc.com/news/technology-52903680.

Steinberg, Brian. “Seattle Judge Tosses Suit That Tried to Gag Fox News Commentary - Variety.” Variety, May 27, 2020. https://variety.com/2020/tv/news/fox-news-lawsuit-washlite-dismissed-1234617658.

Stewart Loane, Susan, and Steven D’Alessandro. “Empowered and Knowledgeable Health Consumers: The Impact of Online Support Groups on the Doctor–Patient Relationship.” *Australasian Marketing Journal (AMJ)*, Health Care in Service Science, 22, no. 3 (August 1, 2014): 238–45. https://doi.org/10.1016/j.ausmj.2014.08.007.

Strychowsky, Julie E., Smriti Nayan, Forough Farrokhyar, and Jonathan MacLean. “YouTube: A Good Source of Information on Pediatric Tonsillectomy?” *International Journal of Pediatric Otorhinolaryngology* 77, no. 6 (June 1, 2013): 972–75. https://doi.org/10.1016/j.ijporl.2013.03.023.

Swoyer, Alex. “Judge Tosses Lawsuit against Fox News over COVID-19 Coverage - Washington Times.” The Washington Times, May 28, 2020. https://www.washingtontimes.com/news/2020/may/28/judge-tosses-lawsuit-against-fox-news-over-covid-1/.

Tustin, Jordan Lee, Natasha Sarah Crowcroft, Dionne Gesink, Ian Johnson, Jennifer Keelan, and Barbara Lachapelle. “User-Driven Comments on a Facebook Advertisement Recruiting Canadian Parents in a Study on Immunization: Content Analysis.” *JMIR Public Health and Surveillance* 4, no. 3 (2018): e10090. https://doi.org/10.2196/10090.

Twitter. “The Twitter Rules.” Twitter, 2020. https://help.twitter.com/en/rules-and-policies/twitter-rules.

U.S. Supreme Court. *Jacobson v. Massachusetts*, 197 U.S. 11 (1905). Accessed December 8, 2020. https://supreme.justia.com/cases/federal/us/197/11/.

U.S. Supreme Court. *United States v. Alvarez*, 567 U.S. 709 (2012). Accessed December 8, 2020. https://supreme.justia.com/cases/federal/us/567/709/.

U.S. Supreme Court. *United States v. O’Brien*, 391 U.S. 367 (1968). Accessed December 8, 2020. https://supreme.justia.com/cases/federal/us/391/367/.

Wakefield, AJ, SH Murch, A Anthony, J Linnell, DM Casson, M Malik, M Berelowitz, et al. “RETRACTED: Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children.” *The Lancet* 351, no. 9103 (February 28, 1998): 637–41. https://doi.org/10.1016/S0140-6736(97)11096-0.

Waldrop, Theresa, Dave Aslup, and Eliot C McLaughlin. “Fearing Coronavirus, Arizona Man Dies after Taking a Form of Chloroquine Used in Aquariums.” CNN, March 25, 2020. https://www.cnn.com/2020/03/23/health/arizona-coronavirus-chloroquine-death/index.html.

Wang, Yuxi, Martin McKee, Aleksandra Torbica, and David Stuckler. “Systematic Literature Review on the Spread of Health-Related Misinformation on Social Media.” *Social*

Science & Medicine 240 (November 1, 2019): 112552. https://doi.org/10.1016/j.socscimed.2019.112552.

Wardle, Clarire, and Hossein Derakhshan. “INFORMATION DISORDER: Toward an Interdisciplinary Framework for Research and Policy Making.” Council of Europe, September 27, 2017. https://firstdraftnews.org/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-de%CC%81sinformation-1.pdf?x25689.

Wetsman, Nicole. “Masks May Help Prevent the Spread of Coronavirus, but the Messaging around Them Has Been Very Bad - The Verge.” The Verge, April 3, 2020. https://www.theverge.com/2020/4/3/21206728/cloth-face-masks-white-house-coronavirus-covid-cdc-messaging.

Wilson, Rebecca F, Sharma AJ, Schluechtermann S, et al. “Factors Influencing Risk for COVID-19 Exposure Among Young Adults Aged 18–23 Years — Winnebago County, Wisconsin, March–July 2020.” *MMWR. Morbidity and Mortality Weekly Report* 69 (2020). https://doi.org/10.15585/mmwr.mm6941e2.

Witteman, Holly O., and Brian J. Zikmund-Fisher. “The Defining Characteristics of Web 2.0 and Their Potential Influence in the Online Vaccination Debate.” *Vaccine*, Special Issue: The Role of Internet Use in Vaccination Decisions, 30, no. 25 (May 28, 2012): 3734–40. https://doi.org/10.1016/j.vaccine.2011.12.039.

Wolfe, Robert M, and Lisa K Sharp. “Anti-Vaccinationists Past and Present.” *BMJ : British Medical Journal* 325, no. 7361 (August 24, 2002): 430–32.

Wong, Tessa. “Sinophobia: How a Virus Reveals the Many Ways China Is Feared.” *BBC News*, February 20, 2020, sec. Asia. https://www.bbc.com/news/world-asia-51456056.

World Health Organization. “Advice for the Public on COVID-19 – World Health Organization.” World Health Organization, December 8, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public.

World Health Organization. “Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation,” September 23, 2020. https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.

Xiang, Yu-Tao, Yuan Yang, Wen Li, Ling Zhang, Qinge Zhang, Teris Cheung, and Chee H. Ng. “Timely Mental Health Care for the 2019 Novel Coronavirus Outbreak Is Urgently Needed.” *The Lancet Psychiatry* 7, no. 3 (March 1, 2020): 228–29. https://doi.org/10.1016/S2215-0366(20)30046-8.

Zadrozny, Brandy. “Facebook Banned Big Anti-Vaccination Pages. Researchers Warn It’s Too Little, Too Late.” NBC News, November 30, 2020. https://www.nbcnews.com/tech/tech-news/covid-19-vaccines-face-varied-powerful-misinformation-movement-online-n1249378.

Zhang, Jun, Weili Wu, Xin Zhao, and Wei Zhang. “Recommended Psychological Crisis Intervention Response to the 2019 Novel Coronavirus Pneumonia Outbreak in China: A Model of West China Hospital.” *Precision Clinical Medicine* 3, no. 1 (March 25, 2020): 3–8. https://doi.org/10.1093/pcmedi/pbaa006.

Ziv, Shahar. “Coronavirus Pandemic Will Cost U.S. Economy $8 Trillion.” Forbes. Forbes Magazine, June 2, 2020. https://www.forbes.com/sites/shaharziv/2020/06/02/coronavirus-pandemic-will-cost-us-economy-8-trillion/.

Zuckerberg, Mark. “I Wanted to Share an Update on the Steps We’re Taking to Respond to the Coronavirus.” Facebook, March 3, 2020. https://www.facebook.com/zuck/posts/10111615249124441.

Zuckerberg, Mark. “Opinion | Mark Zuckerberg: The Internet Needs New Rules. Let’s Start in These Four Areas.” *Washington Post*. Accessed December 9, 2020. https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html.