Journal of the American Medical Informatics Association, 28(7), 2021, 1374–1382
doi: 10.1093/jamia/ocab008
Advance Access Publication Date: 3 February 2021
Research and Applications

Research and Applications

Pattern discovery, validation, and online experiments: a methodology for discovering television shows for public health announcements

Arash Barfar [1] and Balaji Padmanabhan [2]

[1]Department of Information Systems, College of Business, University of Nevada, Reno, Nevada, USA [2]School of Information Systems and Management, Muma College of Business, University of South Florida, Tampa, Florida, USA

Corresponding Author: Arash Barfar, PhD, Department of Information Systems, College of Business, University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV, 89557, USA (abarfar@unr.edu)

Received 12 October 2020; Revised 29 November 2020; Accepted 14 January 2021

ABSTRACT

Objective: Public Health Announcements (PHAs) on television are a means of raising awareness about risk behaviors and chronic conditions. PHAs' scarce airtime puts stress on their target audience reach. We seek to help health campaigns select television shows for their PHAs about smoking, binge drinking, drug overdose, obesity, diabetes, STDs, and other conditions using available statistics.
Materials and Methods: Using Nielsen's TV viewership database for the entire US panel, we presented a novel show discovery methodology for PHAs that combined (i) pattern discovery from high-dimensional data (ii) nonparametric tests for validation, and (iii) online experiments on Facebook.
Results: The nonparametric tests verified the robustness of the discovered associations between the popularity of certain shows and health conditions. Findings from fifty (independent) online experiments (where our awareness messages were seen by nearly 1.5 million American adults) empirically demonstrated the value of the methodology.
Discussion: For 2016, the methodology identified several shows whose popularities were genuinely associated with certain health conditions, opening up the possibility of health agencies embracing both big data and large-scale experimentation to address an old problem in a new way.
Conclusion: Policy makers can repeatedly apply the methodology as new data streams in, with perhaps different feature sets, pattern discovery techniques, and online experiments running over longer periods. The comparatively lower initial investment in the methodology can pay off by identifying several shows for a potentially national television campaign. As simply a by-product, the initial investment also results in awareness messages that might reach millions of individuals.

Key words: public health announcement (PHA), pattern discovery, validation, online experiments, television advertising

INTRODUCTION

As a leading cause of death and disability, chronic conditions account for nearly 90% of healthcare expenditures in the US.[1] Some of the costs could, however, be avoided through prevention or early intervention. Toward this end, health campaigns use Public Health Announcements (PHAs) to encourage the public to cease negative health behavior (eg, smoking) and enact positive health behavior (eg, healthy diets). A significant number of these PHAs are still shown on television given its popularity in the US. The problem, however, is that PHAs constitute less than 1% of total TV advertisements.[2] This miniscule ratio stands in stark contrast to the TV content (both shows and advertisements) that promotes major causes of several chronic conditions (eg, junk food) and unhealthy behavior (eg, smoking and binge drinking). It is imperative, therefore, for health campaigns to ensure that their limited intervention messages reach their target audience.

Motivated by this problem, this study presents a novel TV show discovery methodology for PHAs. The methodology design is in part governed by a generic hypothesis on the existence of "nonaccidental" associations between the popularity of certain shows and health conditions. The three-step methodology first explores Nielsen's national TV viewership database to discover TV shows whose popularities are associated with a health condition or risk behavior. Step 2 then performs a series of nonparametric tests to examine the robustness of the findings. Step 3 finally applies Facebook's split testing platform to experimentally test the discovered patterns and examine their potential application in public health communication. Of note, the third step of the methodology can be viewed as a health campaign itself, given that the PHAs are delivered to millions of individuals throughout the online experiments. To illustrate, the awareness messages in this study were seen by nearly 1.5 million American adults.

The methodology can be repeatedly applied by policy makers as new data constantly streams in. For 2016 (the year we have complete TV viewership data for), the methodology identified several TV shows whose popularities were genuinely associated with certain risk behaviors and health conditions, thereby opening up the possibility of health agencies embracing both big data and large-scale experimentation to address an old problem in a new way.

The specific risk behaviors and health conditions we examine in this study are smoking, binge drinking, drug overdose, obesity, diabetes, and STDs (ie, chlamydia, gonorrhea, and syphilis) for which health agencies provided reliable state-level statistics in 2016.

OBJECTIVE

PHAs' scarce airtime on television puts stress on their target audience reach; that is, health campaigns must ensure that their limited short messages reach their target TV viewers. Our study attends to this aspect of health campaigns by presenting a methodology that helps identify TV shows whose popularities are meaningfully associated with a health condition; for example, determining which shows are best suited for advertising anti-smoking PHAs within. In that sense, the proposed methodology is minimally intrusive. The only necessary household-specific data is the deidentified TV viewership history, for which the household consent is acquired. The health data is publicly available (by CDC) at the state level.

To examine any association between a certain TV show's popularity and a health condition, we first compute state-level popularity scores of the TV show to map onto the available state-level health statistics. Given a universe with hundreds of TV shows (each with multiple state-level popularity scores) and fifty states with health statistics, our study involves a search for associations in high-dimensional data and further ensures that the discovered associations are meaningful and potentially useful.

Accidental and nonaccidental correlations

The term "spurious correlation" was coined by Pearson to address a concerning issue in organic relationships; specifically, "the correlation which will be found between indices when the absolute values of the organs have been selected purely at random."[3] Yet, the notorious spurious correlation that has dominated in the literature is Simon's reestablishment of the term, which basically underlines the adage "correlation is no proof of causation."[4] This popular idea of spuriousness concerns situations where the correlation between two variables is due solely to the operation of a third "causal" variable (eg, x and y are "spuriously" correlated due to $z \rightarrow x$ and $z \rightarrow y$). In Simon's typology, "genuine" or "true" correlations imply the causal relationship between two variables in a wider system (eg, x and y are "genuinely" correlated in $x \rightarrow z \rightarrow y$, with z as the intervening variable).

There is another type of correlation that is more prevalent in the context of high-dimensional data and yet not addressed in Simon's typology. In high-dimensional data, it becomes more likely to find some dimensions that "happen" to correlate with the target variable due solely to chance. In the present study with hundreds of TV shows and state-level health statistics ($n = 50$), the popularity scores of some shows might "accidentally" match the health statistics in different states. This correlation is acknowledged in the typology suggested by Haig.[5] At the highest level, the typology identifies "accidental" correlations with no proper causal interpretations, and "genuine" (nonaccidental) correlations that are amenable to causal interpretations. Table 1 summarizes the nuances of correlations.

Among the nonaccidental correlations in Table 1, direct correlations are less relevant in the context of this study. The only substantiated direct correlation here pertains to smoking, where there is sufficient evidence to conclude that exposure to on-screen smoking can "cause" the initiation of smoking among young people, which further leads to established smoking.[6,7] Television shows that pertain to this direct correlation can be discovered through content

Table 1. Correlations typology

Correlation		Meaning	Note
Genuine (nonaccidental)	Direct	x and y are correlated through causation; ie, $x \rightarrow z \rightarrow y$.	Corresponds to Simon's genuine correlations
	Indirect	x and y are correlated due to a third (latent) causal variable; ie, $z \rightarrow x$ and $z \rightarrow y$.	Corresponds to Simon's spurious correlations
Accidental	Spurious	x and y are correlated due to errors of sampling, computation, or measurement.	Corresponds to Pearson's spurious organic correlation
	Nonsense	x and y are correlated due solely to chance.	Relevant in high-dimensional data

analysis of the shows themselves. For example, recent content analyses (by Truth Initiative) of the popular shows indicate that the Netflix shows *Stranger Things* and *Orange is the New Black* display much more smoking than broadcast and cable shows. In the wake of these analyses, Netflix announced that it will decrease portrayals of smoking in its original programming intended for general audiences.

In contrast to limited cases of direct correlations, a multitude of latent factors generate several "indirect" correlations between the popularity of certain TV shows and health conditions, which can potentially inform selecting shows to advertise PHAs within. Such nonaccidental correlations may stem from a combination of (not necessarily known) latent variables and/or demographics. Given these observations, the proposed methodology analytically and experimentally tests whether some of the discovered associations from the high-dimensional data are "nonaccidental" and can potentially inform selecting TV shows for PHAs.

MATERIALS AND METHODS

Figure 1 presents a schematic of the methodology along with its application in this study to discover TV shows that were potentially effective to advertise PHAs on smoking, binge drinking, drug overdose, obesity, and diabetes within (while the methodology discovered TV shows associated with STDs, we did not conduct experiments on them given the sensitive nature). The authors will provide any nonprofit health campaign with the code for feature engineering, nonparametric tests, and analyses in the proposed methodology at no cost, and can serve as liaisons to Nielsen Inc.

Data

We use Nielsen's database for the entire US panel live and time-shifted TV viewership throughout 2016. The (deidentified) data is at the household level; that is, the household daily viewership of all televised broadcasts (telecasts) from both broadcast and cable stations were tracked and recorded in minutes. The data contained nearly 1.5 million telecasts spanning nearly 7 billion minutes of total viewership, representing the largest data that have been used for this purpose to date. We focus on a list of 600 popular shows nationwide in the US in 2016, which we compiled using several rankings (IMDB, TV.com, Nielsen, and TVGuide). The shows in the compiled list belong to a wide range of genres such as drama, cooking, talk, sports, news, reality shows, game, and politics. Of note, the methodology can be practiced with any list of shows that the campaign selects (eg, based on the advertising cost).

Methodology: pattern discovery

In contrast to Nielsen's household-level TV viewership data, public health statistics are available at the aggregate level. Thus, we first design a set of aggregate-level features that capture the popularity of each TV show in states. Each state-level popularity score (Supplementary Appendix 1) is computed using the household-level show viewership data and addresses a specific aspect of a show's popularity in a state. The aggregate scores allow us to examine the associations between a show's popularity and health conditions at the state-level. The resulting data contains 2400 popularity scores (600 shows, each with four 2016 state-level popularity scores) per state. The set of health conditions contains the 2016 rates of smoking, binge drinking, drug overdose, diabetes, obesity, chlamydia, gonorrhea, and syphilis. In our methodology, we consider feature engineering as part of Step 1, thereby leaving the methodology open to embracing other novel features.

Correlations between variables are generally measured with Pearson's or Spearman's coefficients.[8] Pearson's coefficient measures the strength of the linear relationship between two variables under the assumption that they are (bivariate) normally distributed. Spearman's coefficient, on the other hand, is a nonparametric rank statistic that estimates the degree of the monotone relationship between two variables. In practice, if the distribution is not normal, the data are first converted to rankings to compute the Spearman's correlation coefficient. The Spearman's coefficient between *Breaking Bad* and drug overdose would be equal to $+1.0$ if the show was most popular in the state with the highest condition rate, and was second-most popular in the state with second-highest rate and so on.

However, in this study we search for a simple yet broader association between a TV show's popularity and a health condition; for (a hypothetical) example, the drug-overdose rate is higher in the states where *Breaking Bad* is more popular (eg, where the *Breaking Bad* popularity is above a certain level). Such broad associations do not demand linear correlation between variables (as in Pearson's coefficient) or alignment between the ranked data (as in Spearman's coefficient). To this end, in the first step of the methodology we use single-split binary regression trees to search for such patterns that might exist between a show's popularity scores and a health condition. For each TV show and health condition, the regression tree essentially selects the show's popularity score (from the four popularity scores in Supplementary Appendix 1) that is most useful in highlighting the association between the show and the health condition (in terms of reducing the Root Mean Square Error [RMSE]). It is important to note that any pattern discovery approach that can highlight the association between a show's popularity and a health condition can be used in this step.

In summary, for each health condition we build 600 stand-alone, single-show models. The data for each single-show model contains a feature vector of four popularity scores and a target variable that pertains to the 2016 rate of a specific health condition. The data contains 50 rows (one per state) and updates 4800 times throughout the model building process (600 TV shows and 8 conditions). Compared to the standard deviation of each health condition, the RMSEs of the related single-show models help discover strong (and potentially useful) associations between certain shows' popularity and the health condition. However, it is important to examine whether the discovered patterns are merely proxies for demographics (eg, race and income) or if they capture other latent factors as well (eg, onscreen smoking or binge drinking). Findings from our detailed analyses (Supplementary Appendix 2) suggest that TV show popularity is a proxy for *both* demographics and other (known and unknown) factors, which in turn indicates the potential value of advertising PHAs within TV shows whose popularities are associated with the health condition.

Methodology: validation

A vector of 2400 features (600 shows, each with four 2016 state-level popularity scores) with only 50 observations makes it likely to find TV shows whose popularity scores "happen" to align with a health statistic due to some "accident." We accordingly need to ensure that at least some of the signals carried by the discovered shows are resulting from nonaccidental associations. Step 2 of the methodology performs a series of nonparametric tests that attend to this problem. The premise of the nonparametric tests is that if all discov-

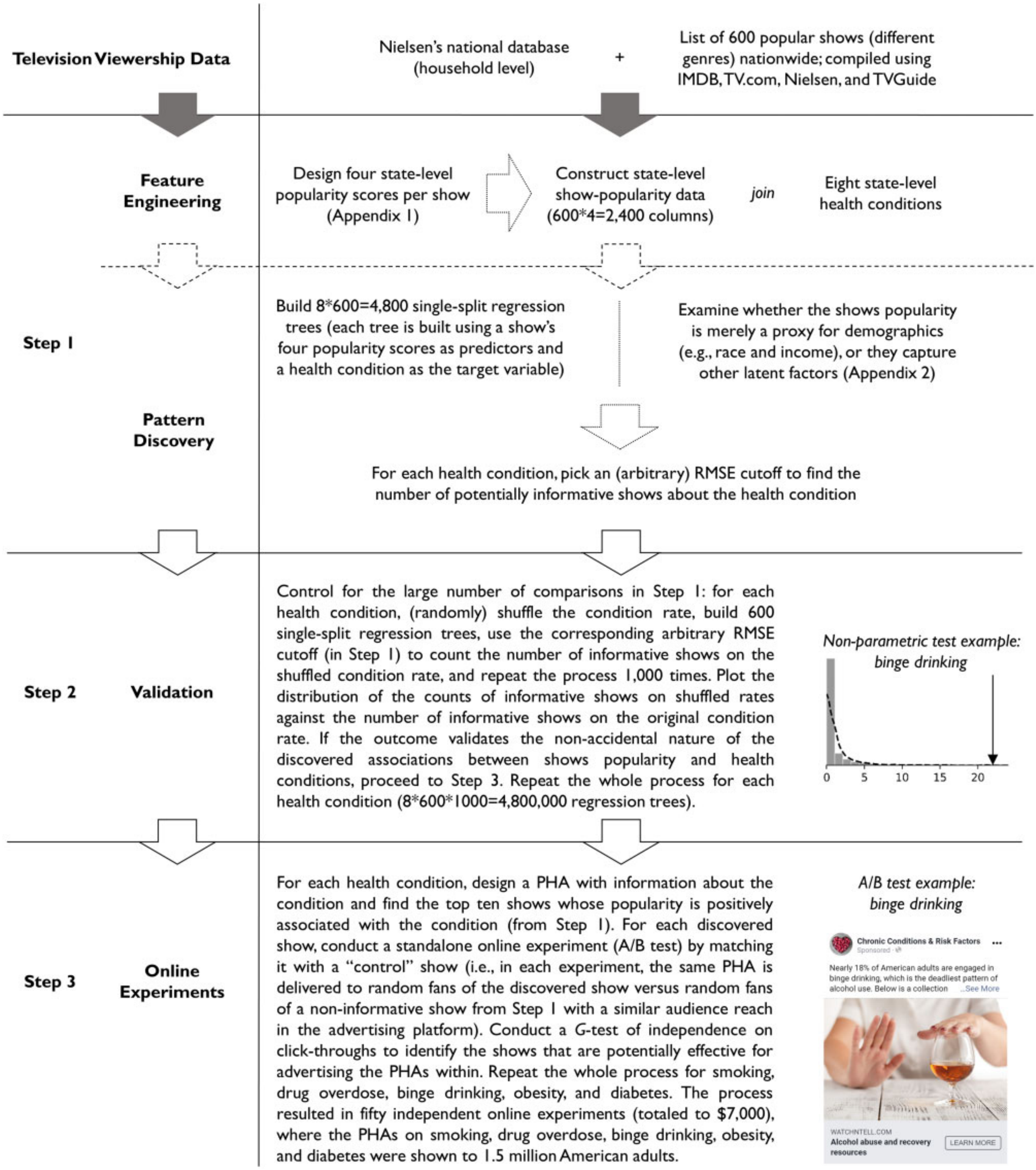

Figure 1. Methodology schematic and its application in this study.

ered patterns are accidental, we should expect a similar number of patterns in the same popularity scores and *shuffled* condition rates. More precisely, if we randomly shuffle condition rates (eg, smoking rates) among the fifty states and run the same model-building process on the shuffled rates, we should find a similar number of potentially informative shows (whose regression tree RMSEs fall below an arbitrary cutoff). If, on the other hand, across many such randomly shuffled runs, the count of potentially informative shows is significantly lower than what it used to be on the original condition rates, we might conclude that at least some of the signals suggested by the discovered shows are not due solely to accidents; in fact, they might stem from genuine associations between the health condition and popularity of certain shows.

Following the above premise, for each condition we preserve the shows' popularity scores but randomly shuffle the health condition (eg, smoking) rates among the 50 states. We then repeat building single-show models on the shuffled rates, register the RMSE of each single-show model, and use the same RMSE cutoff (that we used to count potentially informative shows about the original condition rates) to find the count of shows that seem to carry some signal about the shuffled condition rates. Of note, the robustness tests do not depend on the arbitrary RMSE cutoffs as long as we use the same cutoff to call a TV show potentially informative about both original and shuffled condition (eg, smoking) rates.

After running 1000 iterations of the simulation for each specific health condition, we plot the distribution of the count of potentially informative shows on the shuffled health condition rates. In sum, the simulation process results in 4800000 single-show models (eight health conditions, 600 shows, and 1000 iterations of random shuffling per health condition). This allows us to benchmark the count of potentially informative shows that we observe on an actual condition rates against the distribution of the count of potentially informative shows on its shuffled rates (Figure 2).

In each plot in Figure 2, the vertical solid arrow (plotted typically in the far right tail) marks the number of potentially informative shows on the original condition rate. Panel "a," for example, indicates that, in the original data, there are more than 20 potentially informative shows with respect to binge drinking. Across 1000 shuffled smoking rates and using the same arbitrary RMSE cutoff, however, the distribution indicates that less than 5 shows are deemed informative—suggesting that at least some of the discovered shows in the original data are nonaccidental and, in turn, might be of value. Only the case of drug overdose (panel "c") provides an instance where several shows with similar RMSEs were discovered in the shuffled datasets. Of note, the distributions of the count of informative shows on shuffled data are right-skewed since, in most cases, the resulting regression trees are just roots (ie, no split), and hence their RSMEs are equal to the standard deviation of the corresponding condition rate (which is below any arbitrary cutoff).

Methodology: online experiments

In step 3 we use Facebook's split testing platform to experimentally examine the value of the discovered patterns. If some of the discovered associations are indeed nonaccidental (as suggested by Step 2), we expect them to also manifest in how the fans of the corresponding shows in Facebook respond to information about the health conditions. In an online experiment, specifically, if a PHA (delivered in the form of a Facebook advertisement) is presented to a random set of fans of a discovered show versus a random set of fans of a

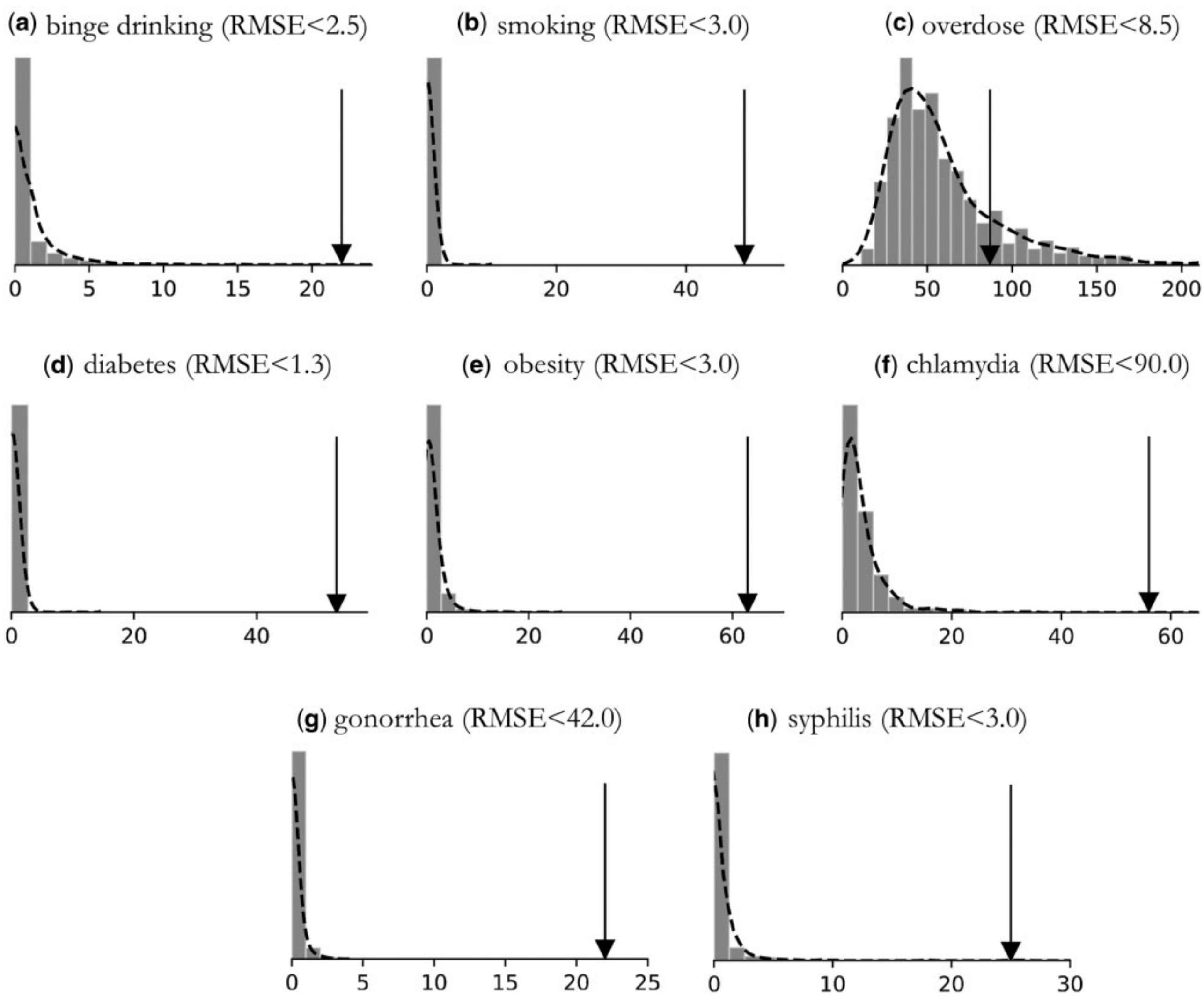

Figure 2. Observed counts of potentially informative shows on original versus shuffled conditions rates.

"control" show, we expect the PHA to attract more responses from the followers of the discovered show if the association with the health condition is indeed nonaccidental. Toward this end, Facebook's split testing platform allows us to test the same PHA on nonoverlapping audience sets. In our online experimental design, accordingly, audience set is the sole test variable and the rest of the advertising parameters are identical.

For every health condition (except for STDs, which are more sensitive than other conditions), we design a PHA that provides users with an awareness message. Similar to the television PHAs that contain follow-up information like toll-free numbers, we include a follow-up link in the PHA, which further allows us to examine how different audience sets respond to the awareness message. To make the experiments worthwhile for users, we launched a portal website with directories to free resources on drug overdose, smoking, binge drinking, diabetes, and obesity. Clicking on the 'Learn More' button in the PHA redirects the user to the related directory in the website. Figure 3 presents the designed PHAs and a sample landing page in the website. All these, and the experiments, were conducted under IRB approvals from two universities in the US. This step of the methodology can be viewed as a health campaign itself as the PHAs were seen by nearly 1.5 million American adults throughout the experiments.

For each health condition, we apply this design separately to every show among the top 10 shows with strong positive association with the condition (identified in the previous steps). The control group in each experiment concerns the random fans of a show (from the same list of 600 popular shows) that (i) was deemed noninformative in Step 1 of the methodology and (ii) has the same size of audience reach in Facebook's advertising platform. We conduct 50 independent online experiments (ie, 5 conditions, 10 discovered TV shows each), each running for a complete week (seven days). All 50 experiments were conducted in Fall 2019 (prior to the Covid-19 pandemic).

RESULTS

The PHAs in the 50 experiments were delivered to nearly 1.5 million American adults on Facebook (30000 users per experiment on average). Table 2 summarizes the experimental results where each row concerns an independent Facebook A/B test that delivered the same PHA on a health condition (Figure 3) to nonoverlapping fans of the "discovered" show versus a "control" show. For example, to experimentally test the discovered association between *Reba*'s popularity and smoking, it is matched with *Once Upon a Time*, and Facebook's split-testing platform randomly chooses individuals within each subpopulation to show the same PHA to. The clickthrough rate (CTR) difference (ΔCTR = *Reba*'s CTR−*Once Upon a Time*'s CTR) is 0.1% and the *P* value of the *G*-test of independence[9] is 0.055, indicating moderate statistical significance at $\alpha = 0.1$. The (seemingly) small ΔCTR should be considered in the context of Facebook, where the average CTR for ads across all industries is 0.9%. The cost-per-click difference is −\$0.74, pointing to a better performance of the smoking PHA among *Reba*'s fans.

In 34 (out of 50) experiments, the PHAs received better CTRs from fans of the discovered TV shows. In 10 experiments, specifically, the CTRs from fans of the discovered shows are significantly greater than the ones from fans of control shows. On the other hand, in 11 online experiments, the control shows have better CTRs than the discovered shows, none of which is statistically significant.

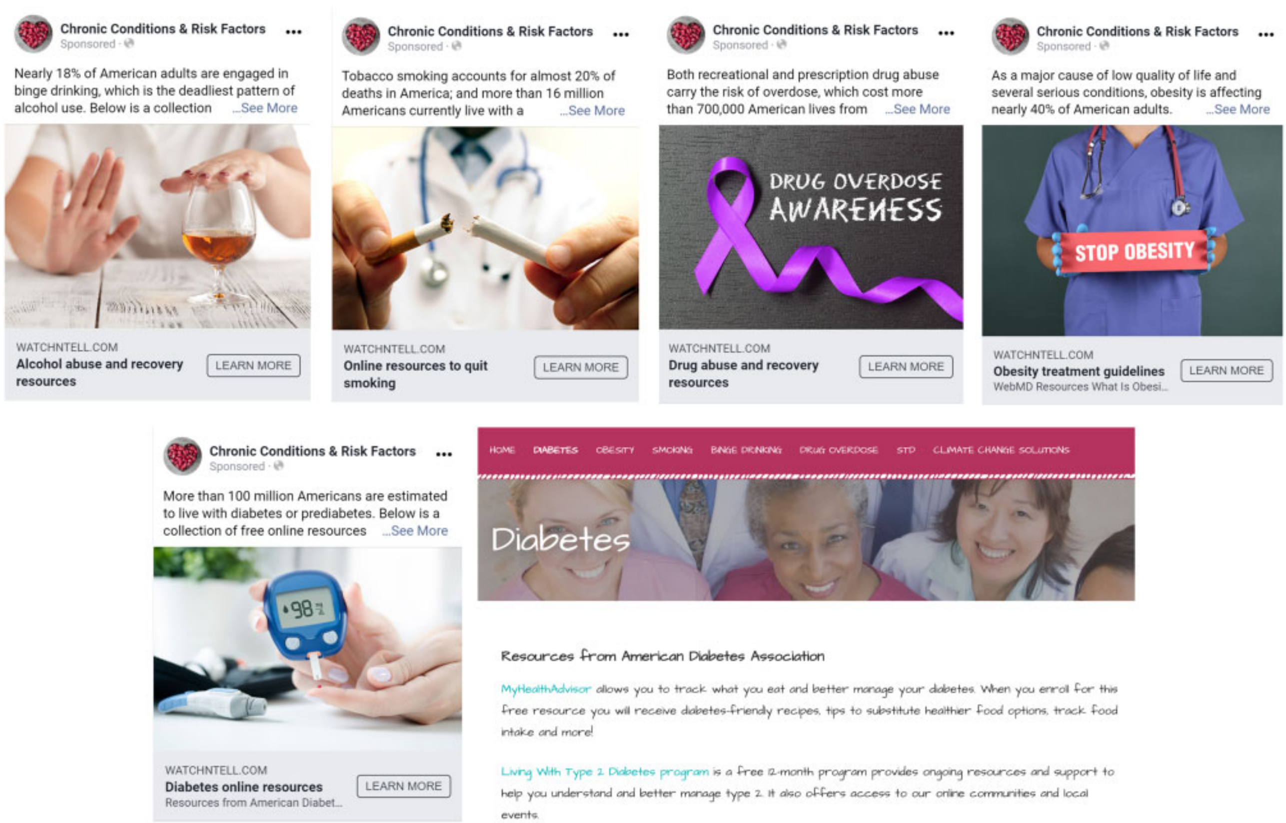

Figure 3. Facebook ads on different conditions and the diabetes landing page in the portal website.

Table 2. Experimental results

	Discovered Show	vs	Control Show	ΔCTR	Δ(Cost per click)
Smoking	Reba	vs	Once Upon a Time	.1%*	−\$.74
	Gunsmoke	vs	Workaholics	.09%	−\$.12
	The Price is Right	vs	The Originals	.14%**	−\$1.58
	Let's Make a Deal	vs	The Venture Bros.	.11%	−\$1.29
	The Young & The Restless	vs	Pretty Little Liars	.00%	\$0
	Duck Dynasty	vs	Shameless	.11%***	−\$1.79
	The Three Stooges	vs	Teen Wolf	.06%	−\$.33
	Dog the Bounty Hunter	vs	New Girl	.01%	−\$.60
	College Gameday	vs	Gotham	−.10%	\$.59
	The Talk	vs	Brooklyn Nine-Nine	.04%	−\$.17
Binge Drinking	Chicago Med	vs	Ash vs Evil Dead	−.03%	\$1.18
	Inside Amy Schumer	vs	20/20	.05%	−\$1.11
	Mom	vs	iZombie	−.03%	\$.81
	Fox NFL	vs	NCIS	.00%	−\$.27
	2 Broke Girls	vs	The Last Ship	.14%*	−\$3.50
	The Catch	vs	Agent Carter	.08%	−\$1.50
	Hollywood Game Night	vs	What Would You Do?	.15%*	−\$1.62
	Jimmy Kimmel	vs	Family Guy	.05%	−\$1.11
	The Americans	vs	Charmed	.13%**	−\$1.20
	American Pickers	vs	Breaking Bad	.08%*	−\$1.89
Drug overdose	Iron Chef America	vs	Hell's Kitchen	.02%	−\$.45
	Jersey Shore	vs	Scandal	.00%	−\$.84
	Bar Rescue	vs	General Hospital	.01%	−\$1.58
	Pit Bulls & Parolees	vs	Leverage	.04%	−\$1.22
	Kroll Show	vs	Nashville	.13%*	−\$1.50
	Full Frontal	vs	Baby Daddy	−.03%	\$2.68
	Conan	vs	Arrow	.03%	−\$.15
	Talking Dead	vs	Chicago Med	−.03%	\$.51
	Grimm	vs	The Blacklist	.06%	−\$2.66
	Modern Family	vs	Fox & Friends	.07%	−\$1.41
Obesity	College GameDay	vs	Chicago Fire	−.01%	\$.11
	Reba	vs	Arrow	.11%**	−\$.83
	The Young & The Restless	vs	The Hills	.08%	−\$1.28
	Duck Dynasty	vs	Elementary	.00%	−\$.29
	Supernatural	vs	Kitchen Nightmares	.05%	−\$.47
	Charmed	vs	1000 Ways to Die	.13%**	−\$1.07
	The 700 Club	vs	Big Brother	.06%	−\$.21
	The Price is Right	vs	Robot Chicken	.05%	−\$.21
	Raising Hope	vs	Squidbillies	.22%**	−\$.50
	Chrisley Knows Best	vs	Suits	−.03%	\$.15
Diabetes	Home Made Simple	vs	The Magicians	.07%	−\$1.58
	Raising Hope	vs	Wheeler Dealers	−.02%	\$.67
	Reba	vs	Survivor	−.03%	\$.66
	Dog the Bounty Hunter	vs	Outlander	−.02%	−\$.35
	Beyond Scared Straight	vs	Silicon Valley	.03%	−\$1.23
	Buffy Vampire Slayer	vs	16 & Pregnant	.06%	−\$.42
	Basketball Wives	vs	Face Off	.01%	−\$.42
	College GameDay	vs	American Ninja Warrior	−.04%	\$.30
	Joel Osteen	vs	Once Upon a Time	.01%	−\$.15
	In Living Color	vs	Leverage	.00%	−\$.40

Notes: *G*-test of independence
$^{*}P < .1$,
$^{**}P < .05$,
$^{***}P < .01$.

DISCUSSION

The results of a Welch's Analysis of Variance[10] confirm that, overall (in all 50 experiments), delivering the PHAs to fans of the discovered shows yields a significantly cheaper cost-per-click (P value = .005). On average, the cost-per-click is \$0.61 less expensive when the PHA is delivered to fans of a discovered TV show. The most significant difference concerns the binge-drinking experiments, which is \$1.02 cheaper on average when the PHA is delivered to fans of the discovered shows. The results in part indicate the potential value of the discovered associations in reaching the audience intended for the PHA, which can potentially help health campaigns complement the TV shows they select (through different methods like content analysis) for their PHAs about the nation's epidemics.

The findings in this study were subject to the following limitations, some of which can be mitigated by employing richer TV viewership data in the methodology.

1. The panel data at our disposal was at the household (and not individual) level, which might introduce noise to the viewership data (and discoveries). Aggregate-level popularity features (in Step 1) can, however, be computed using individual-level viewership data per availability.
2. Our TV viewership data could reliably roll up to the state level. In this study, accordingly, Step 1 (pattern discovery) of the methodology was conducted at the state level. It is however important to note that US states are heterogeneous entities with diverse rates of health conditions. Similarly, television viewership may differ based on population demographics (downtown, suburbs, college towns, and rural areas). Given these, the finer the aggregation level, the less confounding its impact on the methodology's pattern discovery. Health campaigns can practice the methodology at finer granularities (eg, county or census tract) per data availability.
3. Given our limited budget, we examined the practical value of the discovered shows in Facebook, which imposed limitations on our findings for the following reasons:
   i. Recent research[11] provides evidence of discriminatory behavior of "neutral" ad delivery algorithms in Facebook. To mitigate such issues with roots in Facebook's "black box" ad ecosystem, in each experiment we matched a discovered show with a noninformative show with a similar audience size. Still, ad algorithms biases might have affected our experimental outcomes in unknown ways.
   ii. Facebook is not a perfect proxy for television audience given their demographic mismatch. Particularly, older adults in the US spend more time watching live and time-shifted television (which corresponds to our data) and less time online. In contrast, the percentage of younger adults that use Facebook daily is larger.
4. We made specific decisions on aspects of the methodology, like popularity features and models, rather than optimize for those over many feature sets or modeling techniques. However, this is less relevant as a concern, since our contribution is not necessarily the specific findings but the broader methodology.

It is also important to note that television shows and audience interests evolve, making it necessary to periodically practice the methodology with up-to-date data. This is especially important given that the experiments in this study were conducted prior to the Covid-19 pandemic, and public health communication on social media has become much stricter during the pandemic. Television shows usually run on an annual schedule, and, in general, we expect this methodology to be re-applied once a year. There are likely to be a combination of two effects: (i) some "stable" shows that remain associated with a health condition across years, given their stable audiences, and (ii) new shows that periodically draw certain audiences, which might be of value for PHAs. Updating the list of shows to target on an annual basis can potentially capture new shows in addition to confirming the value of the older shows.

Finally, all 50 experiments in this study were conducted with participants throughout the country. Local health agencies, however, can conduct their experiments in specific regions (eg, television market areas) to examine the discovered patterns between certain shows' popularity and the health condition before airing their message in the shows in those areas.

CONCLUSION

Policy makers can apply the proposed methodology to identify TV shows to advertise health messages within. The total Facebook ad cost in our experiments was approximately \$7000 and the authors will work with nonprofit health campaigns to provide the code for the proposed methodology and can serve as liaisons to Nielsen Inc. At a budget of approximately \$100K for experiments, the methodology would identify several shows for a potentially national television campaign that involves millions of dollars in ad spending. The comparatively much lower initial investment in the methodology can pay off in four ways. First, it might identify shows (that were perhaps previously on a shortlist based on current practice) that are likely to be ineffective, saving potentially wasteful ad spending. Second, it might identify new shows, not previously known, that can reach the at-risk population. Third, it can save costs by identifying a broader set of shows, from which campaigns could select cheaper ones. Fourth, as simply a by-product, the initial investment would also result in awareness messages on social media that might reach millions of individuals.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The online experiments were paid for from the authors' own research funds from their respective universities.

AUTHOR CONTRIBUTIONS

Both authors conceptualized the study, 3-step methodology, features, nonparametric tests, and experimental design. Data curation, feature engineering, coding, and analyses were conducted by Arash Barfar. The first draft of the manuscript was written by Arash Barfar and revised critically by both authors.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *the Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENT

The authors have a deep gratitude toward the anonymous reviewing team for their detailed and critical suggestions.

DATA AVAILABILITY STATEMENT

The data used in this research is owned by a third party, Nielsen Inc. and was provided to the authors based on a research partnership between Nielsen and the University of South Florida and therefore cannot be publicly shared.

Nielsen Inc. is a private company that has a 100-year history of being a data provider. The company has a voluntary television viewership panel from which the data is collected. Nielsen Inc. can be contacted directly for access to the original data used in this work. The link below describes the different data sources the company collects and provides to different clients. https://www.nielsen.com/eu/en/solutions/capabilities/nielsenmarketingcloud-daas/ Nielsen also offers charities and nonprofits pro bono marketing campaign data: https://www.nielsen.com/us/en/news-center/2020/nielsen-offers-charities-pro-bono-marketing-campaign-data/

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Goldwater JC, Kwon NJ, Nathanson A, *et al.* Open source electronic health records and chronic disease management. *J Am Med Inform Assoc* 2014; 21 (e1): e50–4.
2. Fuhrel-Forbis A, Nadorff PG, Snyder LB. Analysis of public service announcements on national television, 2001–2006. *Soc Marketing Quart* 2009; 15 (1): 49–69.
3. Pearson K. Mathematical contributions to the theory of evolution—on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc Royal Soc London* 1897; 60 (359-367): 489–98.
4. Simon HA. Spurious correlation: A causal interpretation. *J Am Stat Assoc* 1954; 49 (267): 467–79.
5. Haig BD. What is a spurious correlation? *Understanding Stat Statis Issue Psychol Edu Soc Sci* 2003; 2 (2): 125–32.
6. Sargent JD, Stoolmiller M, Worth KA, *et al.* Exposure to smoking depictions in movies: its association with established adolescent smoking. *Arch Pediatr Adolesc Med* 2007; 161 (9): 849–56.
7. Dalton MA, Beach ML, Adachi-Mejia AM, *et al.* Early exposure to movie smoking predicts established smoking by older teens and young adults. *Pediatrics* 2009; 123 (4): e551– 8.
8. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae* 2011; 30 (2): 87–93.
9. McDonald JH. *Handbook of Biological Statistics*. (3rd ed.). Baltimore, MD: Sparky House Publishing 2014: 68–76.
10. Delacre M, Lakens D, Leys C. Why psychologists should by default use Welch's t-test instead of Student's t-test. *Rips* 2017; 30 (1): 92.
11. Lambrecht A, Tucker C. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Manage Sci* 2019; 65 (7): 2966–81.