## Practice of Epidemiology

# Estimating the Effects of Potential Public Health Interventions on Population Disease Burden: A Step-by-Step Illustration of Causal Inference Methods

**Jennifer Ahern, Alan Hubbard, and Sandro Galea**

Causal inference methods allow estimation of the effects of potential public health interventions on the population burden of disease. Motivated by calls for epidemiologic research to be presented in ways that are more informative for intervention, the authors present a didactic discussion of the steps required to estimate the population effect of a potential intervention using an imputation-based causal inference method and discuss the assumptions of and limitations to its use. An analysis of neighborhood smoking norms and individual smoking behavior is used as an illustration. The implementation steps include the following: 1) modeling the adjusted exposure and outcome association, 2) imputing the outcome probability for each individual while manipulating the exposure by "setting" it to different values, 3) averaging these probabilities across the population, and 4) bootstrapping confidence intervals. Imputed probabilities represent counterfactual estimates of the population smoking prevalence if neighborhood smoking norms could be manipulated through intervention. The degree to which temporal ordering, randomization, stability, and experimental treatment assignment assumptions are met in the illustrative example is discussed, along with ways that future studies could be designed to better meet the assumptions. With this approach, the potential effects of an intervention targeting neighborhoods, individuals, or other units can be estimated.

causality; intervention studies; methods; population; residence characteristics; smoking; social environment

Abbreviations: GEE, generalized estimating equation; OR, odds ratio.

Most analyses of epidemiologic data apply a regression model such as linear or logistic regression. These models have in common that they estimate differences (relative or absolute) between outcomes (in terms of rates, risks, odds, or prevalences) associated with variations in exposure, while holding constant a set of covariates (1–3). These models estimate differences in outcomes that are stratum specific, because they are estimated within strata of the covariates specified in the model. Although such findings constitute the backbone of modern epidemiologic research (3), they represent only 1 approach to capturing the association between an exposure and an outcome. This approach tells us little about population disease burden or about how the disease burden might change if the exposure were modified.

One alternate approach, which could be more informative, would assess how a particular potential intervention on the exposure being studied might reduce disease burden across the population (2, 4). Several methods can estimate population parameters under hypothetical interventions. In simple situations, standardization can estimate a population-level causal effect (5, 6). Certain causal inference methods generalize standardization to situations with covariates that are continuous as well as categorical, covariates that are time dependent, models that include multiplicative interactions, and nonlinear model forms (5–10). Although many causal inference methods were developed to control time-dependent confounding, the machinery allows the estimation of population parameters under hypothetical interventions for cross-sectional studies. Causal inference analyses of epidemiologic data start with the specification of a causal effect that is of interest. The population average causal effect is specified as the difference in the outcome (e.g., the

Correspondence to Dr. Jennifer Ahern, School of Public Health, University of California, Berkeley, 101 Haviland Hall, Berkeley, CA 94720-7358 (e-mail: jahern@berkeley.edu).

proportion of the population with a health outcome) that would have been observed in the population if there had been 1 intervention implemented as opposed to another (or to no intervention), with all else being equal. This is done by estimating different counterfactual distributions of exposures and outcomes (7, 11, 12).

There have been recent calls for epidemiologic research to be conducted or presented in ways that are more informative for those considering intervention (4, 13–18). Although some recent publications have clearly described aspects of using causal inference methods for estimating intervention effects (6, 7, 19–23), we are not aware of any publication that combines discussion of the theoretical utility of these methods, the details of steps required for implementation, and consideration of the assumptions and limitations underlying these methods that is accessible to a broad readership. In this didactic piece, we use an imputation-based causal inference method. Although imputation has long been used to fill in missing data on observed subjects (24), in this setting imputation is used to fill in missing counterfactual observations of subjects observed in only 1 exposure state. This general technique has been called the "g-computation algorithm" in the longitudinal setting and has been discussed in statistical detail elsewhere (8, 25). It is our goal in this paper to present an intuitive and practical discussion of this approach that illuminates the utility of this method for estimating changes in the distribution outcomes that might be of interest when considering specific potential interventions, the steps required for implementation, and some of the assumptions of and limitations to its use.

We use an analysis of neighborhood smoking norms and individual smoking behavior to illustrate the application of these methods. The estimation methods applied here allow us to examine how the distribution of smoking would be different in the population if we were able to change neighborhood smoking norms.

## MATERIALS AND METHODS

### Data source

We illustrate the application of these methods using data from the New York Social Environment Study, a multilevel study designed to examine neighborhood-level exposures that include economic, social, and structural characteristics and substance use in New York, New York (referred to hereafter as "New York City"). The New York Social Environment Study was conducted between June and December 2005. Random digit dialing methods were used to contact and interview 4,000 New York City residents. One randomly selected adult 18 years or older was interviewed by telephone in each household.

Smoking behavior was assessed from each respondent by using the tobacco module in the World Mental Health Comprehensive International Diagnostic Interview (26, 27). Neighborhood smoking norms were measured with a question modified from the National Survey on Drug Use and Health (28). The neighborhood measure is the proportion of residents who believe it is "unacceptable" for adults to smoke cigarettes regularly in each neighborhood.

Respondents provided their residential address or nearest cross-streets so that their locations could be geocoded and linked to their neighborhoods of residence. The neighborhood units for this analysis were the 59 community districts in New York City. Further details about the New York Social Environment Study are available elsewhere (29, 30).

### Analysis step 1

The first step in this analysis was to estimate the association between the exposure, neighborhood smoking norms, and the outcome, current smoking. A full elaboration of this and other related analyses is the subject of another paper (30). For illustrative purposes here, the details of 1 model are presented in Table 1. Briefly, a generalized estimating equation (GEE) logistic regression model with an exchangeable working correlation was used to account for potential clustering by neighborhood and to estimate the association between neighborhood smoking norms and individual smoking (31–33). (If the model is linear, GEE, random effects, or simple linear regression models could all be used to estimate the association of interest in step 1. If the model is logistic, GEE or simple logistic models can be used in a straightforward way because they both produce marginal estimates. With a random effects logistic model, the predicted values produced in step 2 need to incorporate the predicted value of the random effect for each neighborhood, because random effects models produce neighborhood-specific estimates). In this model, we adjusted for confounders and included interaction terms based on an hypothesis posed in the original analysis that the effects of neighborhood smoking norms on smoking would vary depending on individuals' smoking history (30). We found an inverse association between more prohibitive neighborhood smoking norms and current smoking, and there was an interaction between smoking norms and history of smoking. For those with no history of smoking before living in their current neighborhood, the reference group in this model, there was the strongest protective effect of antismoking norms on smoking (odds ratio (OR) $= e^{(\text{norms beta} \times (\text{standard deviation} \times 2))} = e^{(-0.688 \times (0.096 \times 2))} = 0.27$ for a 2-standard deviation increase in the strength of antismoking norms). However, among those who tried smoking before living in the current neighborhood but never smoked regularly, there was almost no association (OR $= e^{(\text{norms beta} \times (\text{standard deviation} \times 2) + \text{tried smoking} \times \text{norms interaction beta} \times (\text{standard deviation} \times 2))} = e^{(-0.688 \times (0.096 \times 2) + 6.48 \times (0.096 \times 2))} = 0.93$). Similarly, there was essentially no association with neighborhood smoking norms among those who smoked weekly (OR $= 1.14$) or daily (OR $= 0.90$) before living in the current neighborhood. As with any multivariable logistic model, this model presents the separate contributions of each covariate to the odds of smoking; for example, men had 2.09 times the odds of smoking compared with women, and those with less than a high school education had 2.70 times the odds of smoking compared with those who had done graduate work.

These results highlight the 2 central reasons why this type of analysis falls short when the interest is in the effect of a specific potential intervention on the population levels of the outcome. First, neighborhood smoking norms are

**Table 1.** Generalized Estimating Equation Logistic Regression Model of the Association Between Neighborhood Smoking Norms and Individual Current Smoking, New York, New York, 2005

| | Beta[a] | Standard Error | Odds Ratio |
|---|---|---|---|
| Intercept | −4.00*** | 0.30 | |
| Neighborhood smoking norms | −6.88*** | 1.04 | 0.27 |
| Smoking before moved to neighborhood | | | |
| Never smoked | | | 1.00 |
| Ever smoked/tried smoking | −0.02 | 0.16 | 0.98 |
| Weekly smoker | 2.72*** | 0.24 | 15.21 |
| Daily smoker | 2.86*** | 0.15 | 17.49 |
| Age, years | | | |
| 18–24 | 1.28*** | 0.25 | 3.58 |
| 25–34 | 0.81*** | 0.21 | 2.25 |
| 35–44 | 0.34 | 0.18 | 1.40 |
| 45–54 | | | 1.00 |
| 55–64 | −0.55* | 0.23 | 0.57 |
| ≥65 | −1.72*** | 0.25 | 0.18 |
| Missing | 0.76 | 0.76 | 2.14 |
| Race | | | |
| White | | | 1.00 |
| African American | −0.04 | 0.14 | 0.96 |
| Asian | −0.32 | 0.28 | 0.73 |
| Hispanic | 0.11 | 0.19 | 1.12 |
| Other | 0.53 | 0.34 | 1.69 |
| Missing | 0.37 | 0.40 | 1.45 |
| Sex | | | |
| Female | | | |
| Male | 0.48*** | 0.10 | 1.61 |
| Marital status | | | |
| Married | | | 1.00 |
| Divorced | 0.34 | 0.21 | 1.41 |
| Separated | 0.41 | 0.24 | 1.50 |
| Widowed | 0.60* | 0.28 | 1.82 |
| Never married | 0.35* | 0.16 | 1.42 |
| Missing | 0.42 | 0.86 | 1.53 |

**Table continues**

**Table 1.** Continued

| | Beta[a] | Standard Error | Odds Ratio |
|---|---|---|---|
| Birthplace | | | |
| New York, New York | | | 1.00 |
| Other US location | −0.02 | 0.15 | 0.98 |
| Different country | −0.68*** | 0.16 | 0.51 |
| Missing | −0.85 | 0.71 | 0.43 |
| Survey language | | | |
| English | | | 1.00 |
| Spanish | −0.69** | 0.26 | 0.50 |
| Years lived in neighborhood | | | |
| 0–7 | | | 1.00 |
| 8–21 | 0.31* | 0.14 | 1.36 |
| >21 | 0.85*** | 0.17 | 2.35 |
| Missing | 0.19 | 0.84 | 1.21 |
| Income | | | |
| >$80,000 | | | 1.00 |
| $40,001–$80,000 | 0.10 | 0.17 | 1.10 |
| ≤$40,000 | 0.22 | 0.20 | 1.24 |
| Missing | −0.40 | 0.23 | 0.67 |
| Education | | | |
| Graduate work | | | 1.00 |
| College graduate | 0.41* | 0.18 | 1.50 |
| Some college | 0.73*** | 0.20 | 2.07 |
| High school/GED | 0.35 | 0.21 | 1.42 |
| Less than high school | 0.99*** | 0.28 | 2.70 |
| Missing | 1.45** | 0.53 | 4.26 |
| Unemployed | | | |
| No | | | 1.00 |
| Yes | 0.74*** | 0.16 | 2.09 |
| Smoking history × smoking norms | | | |
| Never smoked × smoking norm | | | N/A |
| Tried smoking × smoking norm | 6.48** | 2.16 | N/A |
| Weekly smoker × smoking norm | 7.55** | 2.72 | N/A |
| Daily smoker × smoking norm | 6.35*** | 1.24 | N/A |

Abbreviations: GED, general equivalency diploma; N/A, not applicable.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.
[a] Although parameter values are rounded in this table and in the equations in the text, unrounded values should be used in any analysis.

associated with individual smoking behavior only among those with no history of smoking; we cannot tell what effect changing norms would have on the whole population because only some persons in the population would be affected. This analysis does not incorporate both the differences in impacts of the exposure on subgroups and how common the subgroups are in the population. Second, each covariate in the multivariable model makes its contribution to the odds of smoking by an individual. However, each covariate's separate contribution gives the reader no sense of how these ''bits'' of risk accumulate for any given

individual in predicting his/her overall probability of being a smoker or for the population overall in predicting smoking levels.

**Analysis step 2**

Moving beyond ''bits'' of risk, the second step was to use the model from step 1 to impute the probabilities of smoking for each individual in the data set incorporating all of the individual's particular characteristics. These probabilities were estimated while ''setting'' or fixing the neighborhood

norms to different levels that correspond to the range of the observed data (34). In this analysis, we set the norm values across the range from 40% to 75% in 5% intervals (percent who believe it is unacceptable for adults to smoke cigarettes), covering the range observed in the data. Each individual's probability was based on his/her individual covariates and the risks that they contribute, as well as on the risk from the "set" norm value in the neighborhood. We can think about the unobserved counterfactual probability of smoking given a neighborhood norm level that a particular individual did not experience as missing data. With this method, we are imputing each individual's probability of smoking if he/she had experienced a norm value that he/she did not experience to estimate the missing counterfactuals. As an example, the calculations for when 70% of persons in each neighborhood believe it is unacceptable to smoke are presented below for someone with no history of smoking and for someone who smokes daily to illustrate how the interaction combines with the "set" norm value.

The predicted log odds of smoking for each individual ($i$) when 70% believe it is unacceptable to smoke (plo70) (the smoking norm variable in the model) (Table 1) are centered around the mean of 58%, so that, for 70% unacceptable, the variable is "set" to the value of 0.12 or 12%:

$$\text{plo70}_{i, \text{no history of smoking}} = \text{intercept beta}$$
$$+ (0.12 \times \text{neighborhood smoking norms beta})$$
$$+ \text{other covariate betas} = -4.00 + (0.12 \times -6.88) + \ldots$$

$$\text{plo70}_{i, \text{daily smoker}} = \text{intercept beta}$$
$$+ (0.12 \times \text{neighborhood smoking norms beta})$$
$$+ \text{daily smoker beta} + (0.12$$
$$\times \text{daily smoker smoking norm interaction beta})$$
$$+ \text{other covariate betas} = -4.00 + (0.12 \times -6.88)$$
$$+ 2.86 + (0.12 \times 6.35) + \ldots.$$

The predicted probability of smoking for each individual ($i$) when 70% believe it is unacceptable to smoke (Pprob70) is shown as follows:

$$\text{Pprob70}_i = 1/(1 + \exp(-1 \times \text{plo70}_i)).$$

In the original analysis (Table 1), the effects of norms on smoking were presented for different subgroups of the population (because of the interaction between smoking history and norms), and the contributing effects of the individual covariates were presented separately. In contrast, we have now considered together the attributes that shaped each individual's overall probability of smoking. Moreover, we have estimated the counterfactual probabilities of smoking for each individual if the norms in his/her neighborhood had taken different values.

## Analysis step 3

The third step involved a simple averaging of the imputed probabilities of smoking for each individual across the whole population, for each "set" level of smoking
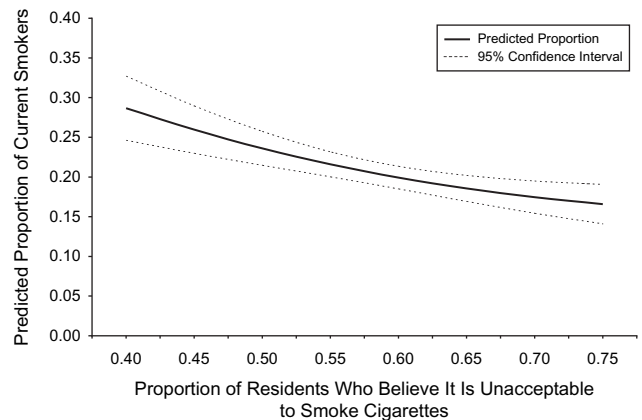
**Figure 1.** Predicted smoking prevalence corresponding to counterfactually "set" levels of neighborhood smoking norms, New York, New York, 2005.

norms. These averages of the individual probabilities estimated the prevalence of smoking for each level of neighborhood smoking norms if, contrary to fact, that "set" level of norms had been present in the whole population. These predicted probabilities are presented in Figure 1. If norms were at their most permissive level (across the range observed) in all neighborhoods, the imputed prevalence of smoking in the whole population would be 29%. If norms were at their most prohibitive level in all neighborhoods, the imputed prevalence of smoking in the whole population would be 17%.

We now have a population-wide predicted effect of changing norm levels on the prevalence of smoking in the whole population. Underlying this population-wide prediction are all of the individual contributions to risk, including the stronger effects of norms on some individuals (those with no prior history of smoking) than on others. We have estimated the predicted population-wide net effect of changing norms, while allowing any heterogeneity in associations due to the interaction to exist at the individual level.

## Analysis step 4

The fourth step was the calculation of confidence intervals around the estimate of the population-level effect. There is typically no straightforward analytical estimate of the standard error available for this population-level effect. However, the standard error can be estimated with a bootstrapping technique (35). The technique is based on resampling from the study population with replacement (following the original sampling design), estimation of the imputed population probabilities of smoking for each "set" level of norms in the new sample, and then repetition of this process 1,000 times, so that the imputed probabilities across all of the repeated samples capture the sampling distribution from which we can calculate a standard error. The bootstrapped 95% confidence interval for the predicted prevalence of smoking in this example is presented in Figure 1.

The Appendix provides further details on the principles and mechanics of bootstrapping.

## RESULTS AND DISCUSSION

In this analysis, we presented 1 approach to estimate the effects of a specific potential intervention with neighborhood smoking norms on the burden of smoking in the whole population. We examined what smoking levels would be if we could manipulate smoking norms in neighborhoods and set them across a range of values. This is in contrast to the effect we were able to estimate with a traditional regression analysis, which produced stratum-specific odds ratios that varied by subgroups depending on smoking history.

In the illustrative example, we found that, if smoking norms were changed to the most permissive level in all neighborhoods, the prevalence of smoking in the whole population would be about 29%. In contrast, if neighborhood smoking norms could be changed to their most prohibitive level in all neighborhoods, the prevalence of smoking would be about 17%. With the tools in hand for conducting the analysis presented here, the potential effects of any specific intervention on neighborhoods, individuals, or other units of interest can be estimated. Considering this illustrative example, another interesting analysis could have estimated the change in smoking prevalence if we had reduced the exposure to a certain level among neighborhoods with particularly permissive smoking norms. More generally, this approach can estimate what would happen to the outcome if researchers were able to change the exposure for any subgroups of interest.

There are several assumptions that need to be met to allow us to interpret the predicted values causally (8). First, we assume that the confounders came before the exposure and that the exposure came before the outcome. This is known as the "temporal ordering assumption"; temporality is a commonly cited requirement for causality. Second, we assume that there are no unmeasured confounders for the exposure–outcome relation being studied. This is known as the "randomization assumption" because, if all confounders have been measured, within strata of the confounders the exposure is effectively randomized. Third, we assume that the outcome of any individual is independent of the exposures and outcomes (or counterfactuals) of other individuals, an assumption known as the "stability assumption" (or "stable unit treatment value assumption") (36, 37). This means that any individual conceptually has a set of counterfactual exposure–outcome combinations that could have been observed which are not affected by the exposures and outcomes of others. Finally, we assume that all exposures are possible for all members of the study population, an assumption known as the "experimental treatment assignment assumption." Practically, this means that, within the subgroups defined by combinations of covariates, some individuals have to be observed as exposed and others as unexposed to meet the experimental treatment assignment assumption. If one is comfortable making parametric model assumptions (i.e., extrapolating beyond the data), this assumption is not required when using this analysis method;

however, caution is advisable when interpreting results beyond the range of the observed data. Whether and to what degree these assumptions are met for any analysis should inform the strength with which the findings are interpreted.

It is clear that the data used in the illustrative example here do not meet all of the above assumptions. However, considering the assumptions explicitly helps to clarify how future studies could be designed to strengthen the possibility of estimating a causal effect. Many neighborhood studies, including the one presented here, have a cross-sectional design (38, 39). The temporal ordering assumption required for a causal interpretation of the results is not met by the structure of the data, so the likely temporal ordering of the variables should be considered in assessing this assumption. Most covariates in this analysis are fixed or tend to be static over long time periods (e.g., age, race, sex, marital status, education), and it is reasonable to assume that they came before the exposure. If this assumption were untrue, the estimated parameter could differ from the true causal parameter in either direction. For the exposure of smoking norms and the outcome of smoking, we must assume that it is norms that change in advance of smoking behavior to have appropriate temporal ordering; this is a reasonable assumption, but the reverse may also be true to some extent. Were this assumption untrue, we would infer the wrong causal direction for the parameter estimated. Longitudinal studies, including time-varying individual- and group-level data, will be an important future step in neighborhood analyses to establish temporality.

Unmeasured confounders are always a concern in epidemiology, and several authors have raised particular concern about this issue for social exposures, such as the neighborhood smoking norms of interest in this analysis (40, 41). Although we measured all the confounders we identified based on our knowledge of the literature, there still may be unmeasured and mismeasured confounders. In addition, structural relations among variables (measured and unmeasured) that are not accounted for in the analysis can render the effects of the exposure on the outcome unidentifiable or biased in unpredictable directions (41). For example, in this analysis, we treated all of our covariates as confounders but, if any are on the causal pathway between neighborhood smoking norms and smoking, this creates a different causal structure where a different analysis approach would be appropriate. Although this does not seem a likely problem for the exposure and outcome examined here, it might be more of a problem for other exposures. Sensitivity analyses considering different structural relations and unmeasured variables may be important tools for quantifying these uncertainties (7, 41, 42).

In the context of the stability assumption, this model and the other models used in epidemiology assume that the outcome of any individual is independent of the exposures and outcomes of other individuals. However, this assumption of independence is not met when such phenomena as contagion and positive or negative feedback are at play (43). These issues have been dealt with in infectious disease epidemiology by necessity for a long time; any analysis that ignores the contagion of infectious diseases is clearly likely to be misleading (44, 45). Notably, recent discussions have begun to consider how contagion and feedback may play

important roles in the context of chronic disease and health behaviors (46–48). Smoking is a behavior that likely has an element of contagion because of the social nature of the behavior; the smoking of any individual likely affects the smoking behaviors of others. Recently, it has been suggested that, for neighborhood-level exposures, the stability assumption applies to the neighborhood units rather than to the individuals within the neighborhoods (19). The estimates presented in this paper are based on the prevalences of smoking observed in neighborhoods that currently have norm levels across the range presented in Figure 1. Thus, these estimates assume that the neighborhoods were at equilibrium when they were observed, whatever dynamic processes take place at the individual or neighborhood level as norms and smoking prevalence change. This assumption may or may not be reasonable. Applications of models that explicitly model these dynamic processes may be an interesting complementary approach to consider for anticipating the effects of interventions.

To consider the distribution of exposure across covariate subgroups (assessing the experimental treatment assignment assumption), we examined the distribution of participants between prohibitive and permissive smoking norm neighborhoods using propensities based on their individual covariates, and minimal social stratification was observed. People of all "types" based on their individual characteristics lived in both prohibitive and permissive norm neighborhoods. Thus, there were no extrapolations made in estimating the associations between neighborhood smoking norms and smoking in step 1. Moreover, in this analysis, we limited our consideration of counterfactual levels of neighborhood smoking norms to those actually observed among the neighborhoods in our study (range, 40%–75%). This analysis approach allows extrapolation beyond the observed values; for example, we could impute the counterfactual prevalence of smoking if 90% of the population believed it was unacceptable to smoke. However, it is our opinion, consistent with the work of others in the area (49), that caution should be exercised when considering making such extrapolations.

Naturally, the concept of neighborhood-level smoking norms is functionally and conceptually the aggregate of the norms of neighborhood residents. Two issues about an aggregate exposure merit discussion. First, defining an exposure by a proportion of the population raises a question about how the counterfactual is defined. Typically, counterfactuals are unique (e.g., an individual is either exposed or unexposed); however, if 40% believe it is unacceptable to smoke, it could be any 40% of the overall population. The method that we used, where we apply imputation at the individual level, assumes that the causal effect of the neighborhood exposure is the same no matter which 40% believe smoking is unacceptable; effectively, we assume that the different possible 40%s are exchangeable. Second, the aggregate nature of the exposure means that, when neighborhood smoking norms change, by definition the norms of individuals change. For that reason, we did not adjust the underlying model (step 1) (Table 1) for individual smoking norms. If the underlying individual model had been adjusted for individual smoking norms, the counterfactual question

would have been quite different; it would have been a question about a change in the norms of the neighborhood around the individuals in the study, but with the requirement that their individual norms remained the same. This is an interesting question (and is the subject of another analysis (30)). It is not, however, a question that makes much sense from the perspective of intervention, because one would likely not intend to modify social norms while leaving the norms of a subgroup unchanged. In contrast, by leaving individual norms out of the model, we ask a question about a change in the norms of the neighborhood, allowing the corresponding individual norms to change as they would by definition when neighborhood norms change. We do assume that it is neighborhood norms that influence individual norms; teasing apart the causal relation between individual and neighborhood norms would require detailed time-dependent data.

In this analysis, we have taken a model that was built to test a specific hypothesis and started from this basis to estimate the population-wide effect of changing neighborhood smoking norms. However, research on the optimal approach to selecting the model that provides the basis for the population-wide effect is in its infancy. We could create a model at the individual level that simply provides the best fit to the data regardless of whether it makes any sense from the perspective of the subject matter, or in relation to any hypotheses of interest. One strength of this approach is that it allows any underlying individual-level model that will best control for confounding to be applied, no matter how complicated that model may be (e.g., including interactions and nonlinear terms), and all of that complexity in confounding is summarized in a nuisance parameter. So-called "black-box" model selection techniques are being explored for this purpose, and these techniques provide estimates that are somewhat robust to model misspecification (50).

Ultimately, complex analytical approaches do not get around fundamental issues of good study design and conduct, careful measurement, and consideration of the required assumptions (40). However, the analytical approaches that are traditionally used in any discipline tend to shape and constrain the types of questions that are asked in a field (43). This analytical approach expands the range of questions that we might ask to include more that may be directly pertinent to the effects of interventions on a population, and it presents an interesting complement to the traditional estimates from regression models. Certainly this approach should be applied to high-quality studies with data that meet the assumptions to the greatest extent possible if a causal interpretation of the effect is of interest. Whether such estimates can be informative about the potential results of intervention remains to be determined and perhaps can be assessed with comparison to actual interventions.

School of Public Health, University of California, Berkeley, Berkeley, California (Alan Hubbard); Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan (Sandro Galea); Survey Research Center, Institute for Social Research, Ann Arbor, Michigan (Sandro Galea); and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York (Sandro Galea).

## REFERENCES

1. Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott-Raven Publishers; 1998.
2. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 2001;30(3):427–432.
3. Szklo M, Nieto FJ. *Epidemiology. Beyond the Basics*. Boston, MA: Jones and Bartlett Publishers; 2007.
4. Rockhill B. Theorizing about causes at the individual level while estimating effects at the population level. *Epidemiology*. 2005;16(1):124–129.
5. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*. 2003;14(6):680–686.
6. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006; 60(7):578–586.
7. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31(2):422–429.
8. van der Laan MJ, Robins JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York, NY: Springer Verlag; 2002.
9. Tsiatis A. *Semiparametric Theory and Missing Data*. New York, NY: Springer; 2006.
10. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran EM, Berry D, eds. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. New York, NY: Springer; 2000: 95–113.
11. Rubin DB. Estimating the causal effects of treatment in randomized and nonrandomized studies. *J Educ Psychol*. 1974; 66(5):688–701.
12 Rubin DB. Practical implications of the modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*. 1991;47(4):1213–1234.
13. Tickner JA. Commentary: barriers and opportunities to changing the research agenda to support precaution and primary prevention. *Int J Occup Med Environ Health*. 2004; 17(1):163–173.
14. Shenassa E, Daskalakis C. Smoking prevention: implications of study design, research setting, and goals. *Nicotine Tob Res*. 2004;6(2):321–325.
15. Leung MW, Yen IH, Minkler M. Community based participatory research: a promising approach for increasing epidemiology's relevance in the 21st century. *Int J Epidemiol*. 2004;33(3):499–506.
16. Loewenson R. Epidemiology in the era of globalization: skills transfer or new skills? *Int J Epidemiol*. 2004;33(5):1144–1150.
17. Kaufman JS, Kaufman S, Poole C. Causal inference from randomized trials in social epidemiology. *Soc Sci Med*. 2003; 57(12):2397–2409.
18. Merzel C, D'Afflitti J. Reconsidering community-based health promotion: promise, performance, potential. *Am J Public Health*. 2003;93(4):557–574.
19. Vanderweele TJ. Ignorability and stability assumptions in neighborhood effects research. *Stat Med*. 2008;27(11): 1934–1943.
20. Hernan MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health*. 2004;58(4): 265–271.
21. Höfler M. Causal inference based on counterfactuals [electronic article]. *BMC Med Res Methodol*. 2005;5:28.
22. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000;21:121–145.
23. Hubbard AE, van der Laan MJ. Population intervention models in causal inference. *Biometrika*. 2008;95(1):35–47.
24. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley and Sons, Inc; 1987.
25. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7:1393–1512.
26. Kessler RC, Ustun TB. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res*. 2004;13(2):93–121.
27. Kessler RC, Abelson J, Demler O, et al. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *Int J Methods Psychiatr Res*. 2004;13(2):122–139.
28. Office of Applied Studies, Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services. *National Survey on Drug Use and Health, 2007* [computer file]. ICPSR study no. 23782. Ann Arbor, MI: Inter-university Consortium for Political and Social Research; 2008.
29. Ahern J, Galea S, Hubbard A, et al. "Culture of drinking" and individual problems with alcohol use. *Am J Epidemiol*. 2008; 167(9):1041–1049.
30. Ahern J, Galea S, Hubbard A, et al. Neighborhood smoking norms modify the relation between collective efficacy and smoking behavior. *Drug Alcohol Depend*. 2009;100(1-2): 138–145.
31. Hanley JA, Negassa A, Edwardes MD, et al. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol*. 2003;157(4):364–375.
32. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988; 44(4):1049–1060.
33. Fitzmaurice GM. Laird NM. Ware JH. *Applied Longitudinal Analysis*. Hoboken, NJ: John Wiley & Sons, Inc; 2004.
34. Pearl J. *Causality: Models, Reasoning and Inference*. Cambridge, United Kingdom: Cambridge University Press; 2000.
35. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci*. 1986;1:54–77.
36. Rubin DB. Comment on "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *J Am Stat Assoc*. 1980;75:591–593.
37. Rubin DB. Comment on Neyman (1923) and causal inference in experiments and observational studies. *Stat Sci*. 1990;5: 472–480.
38. Diez Roux AV. Neighborhoods and health: where are we and where do we go from here? *Rev Epidemiol Sante Publique*. 2007;55(1):13–21.

39. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med*. 2002;55(1):125–139.
40. Oakes JM, Kaufman JS. Introduction: advancing methods in social epidemiology. In: Oakes JM, Kaufman JS, eds. *Methods in Social Epidemiology*. San Francisco, CA: Josey-Bass; 2006.
41. Kaufman JS, Kaufman S. Assessment of structured socioeconomic effects on health. *Epidemiology*. 2001;12(2):157–167.
42. MacLehose RF, Kaufman S, Kaufman JS, et al. Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology*. 2005;16(4):548–555.
43. Miller JH, Page SE. *Complex Adaptive Systems*. Princeton, NJ: Princeton University Press; 2007.
44. Koopman J, Longini I. The ecological effect of individual exposures and nonlinear disease dynamics in populations. *Am J Public Health*. 1994;84(5):836–842.
45. Halloran EM, Struchiner CJ. Causal inference in infectious diseases. *Epidemiology*. 1995;6(2):142–151.
46. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357(4):370–379.
47. Ness RB, Koopman JS, Roberts MS. Causal system modeling in chronic disease epidemiology: a proposal. *Ann Epidemiol*. 2007;17(7):564–568.
48. Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. *Am J Public Health*. 2006;96(3):452–458.
49. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med*. 2004;58(10):1929–1952.
50. van der Laan MJ, Rubin D. Targeted maximum likelihood learning. *Int J Biostat*. 2006;2(1):Art. 11. (http://www.bepress.com/ijb/vol2/iss1/11).
51. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. London, United Kingdom: Chapman and Hall; 1993.
52. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med*. 2000;19(9):1141–1164.

## APPENDIX

The principle of bootstrapping is the following. Persons who participated in the study are a representative sample of residents in the target population under study (in this instance, the New York City adult population). We assume, as one always does when using epidemiologic data, that the persons we happened to interview represent everyone else we might have interviewed. We then treat the sample we have as if it were itself a source population. We sample from the *sample* we have *with replacement*. This means that, in the New York Social Environment Study where there are 4,000 respondents, we randomly sample 4,000 people from the sample with replacement. Some persons will get included more than once, and others will not get included at all. This is similar to how, if we repeated the original study sampling from the source population, we might have drawn a slightly different sample by chance. If any form of complex sampling was used in the initial study (e.g., clustered or stratified sampling), the selection of individuals for bootstrapping has to mimic the original sampling design of the study. Because the study in the illustrative example used random sampling, random sampling from the sample works here for bootstrapping.

This sample of our sample can be called a bootstrap sample. We take this bootstrap sample and go through all of the steps described above, estimating the adjusted association between neighborhood smoking norms and individual smoking in the bootstrap sample, predicting the probability of smoking for each individual "setting" smoking norms to different (counterfactual) levels, and averaging those probabilities to estimate the effect of different levels of smoking norms in the whole population. When we have estimated the average probabilities from the bootstrap sample, we save them in a data set. We then draw a new bootstrap sample from the study sample and repeat this process again. We have to repeat this process about 1,000 times to get a reasonably precise standard error. In the end, we have estimates of each probability for the corresponding counterfactual norm level from 1,000 different samples of the data. The variability in these estimates captures the sampling variability that is estimated by a standard error. We then use the distribution of these probabilities to get standard errors and confidence intervals for the parameter estimates, either by calculating the standard error of the 1,000 average probabilities (for each "set" norm value) or by directly taking the 2.5 percentile and 97.5 percentile values of those 1,000 average probabilities as the confidence limits. When there is a concern about bias in the confidence intervals, as would be suggested by differences between the percentile method and the standard error method, particularly where the percentile method produces asymmetric confidence intervals, there are variations on these approaches devised to remove the bias. For example, the bias corrected and accelerated (BCa) confidence interval allows for asymmetry and skewness, although the calculations can be quite involved (51, 52). While this process may sound cumbersome, some statistical packages provide bootstrap commands (e.g., STATA; StataCorp LP, College Station, Texas), and bootstrapping can be implemented in other programs with relatively simple coding (e.g., SAS; SAS Institute, Inc., Cary, North Carolina).