

TECHNICAL ADVANCE

Open Access



Semiparametric estimation of the attributable fraction when there are interactions under monotonicity constraints

Wei Wang^{1*} , Dylan S. Small² and Michael O. Harhay^{1,3}

Abstract

Background: The population attributable fraction (PAF) is the fraction of disease cases in a sample that can be attributed to an exposure. Estimating the PAF often involves the estimation of the probability of having the disease given the exposure while adjusting for confounders. In many settings, the exposure can interact with confounders. Additionally, the exposure may have a monotone effect on the probability of having the disease, and this effect is not necessarily linear.

Methods: We develop a semiparametric approach for estimating the probability of having the disease and, consequently, for estimating the PAF, controlling for the interaction between the exposure and a confounder. We use a tensor product of univariate B-splines to model the interaction under the monotonicity constraint. The model fitting procedure is formulated as a quadratic programming problem, and, thus, can be easily solved using standard optimization packages. We conduct simulations to compare the performance of the developed approach with the conventional B-splines approach without the monotonicity constraint, and with the logistic regression approach. To illustrate our method, we estimate the PAF of hopelessness and depression for suicidal ideation among elderly depressed patients.

Results: The proposed estimator exhibited better performance than the other two approaches in the simulation settings we tried. The estimated PAF attributable to hopelessness is 67.99% with 95% confidence interval: 42.10% to 97.42%, and is 22.36% with 95% confidence interval: 12.77% to 56.49% due to depression.

Conclusions: The developed approach is easy to implement and supports flexible modeling of possible non-linear relationships between a disease and an exposure of interest.

Keywords: Attributable fraction, B-splines, Interaction, Monotonicity constraint, Quadratic programming

Background

The population attributable fraction (PAF) is the fraction of disease cases in a sample that can be attributed to the exposure. The PAF is an important measure of the public health impact of an exposure on disease burden, and thus it is useful to prioritize public health interventions

[1, 2]. The maximum likelihood method is commonly used to estimate the PAF. However, this approach to estimation requires a correct model for the probability of disease given the exposure and other covariates subject to the ignorable treatment assignment assumption [3] (to be reviewed in the next section). Logistic regression is typically used to model the probability of disease given the exposure and the other covariates [1, 4, 5].

The PAF of an exposure must fall between 0 and 1 by definition. A zero PAF indicates that the disease risk is irrelevant of the exposure levels. In contrast, a higher PAF

*Correspondence: wei.wang@pennterms.upenn.edu

¹Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

value indicates a stronger association between disease risk and the exposure level. In the extreme, a PAF equal to 1 implies no disease risk when there is no exposure. If the disease risk is higher in the absence of the exposure than in the presence of the exposure, the PAF then has no proper meaning. Thus, the definition of the PAF itself also suggests a monotone relationship between the disease risk and the exposure level (a justification is presented at the end of the next section). Incorporating the monotonicity assumption into the estimation of the PAF provides performance gains when there are no other covariates [6].

In many research fields, the exposure is thought to have a monotone effect on the probability of having the disease, i.e., the probability is a monotone function of the exposure. For instance, a dose-response curve is often thought to be non-decreasing [7, 8]. One such example is the probability of suicidal ideation, which is thought to be an increasing function of both hopelessness and depression [9–12]. Hence, it is desirable to model the probability of suicidal ideation under the monotonicity constraint of both hopelessness and depression.

This situation also presents an analytic challenge. When there are other covariates, they can interact with the exposure, for instance, the interaction of two drugs [13]. In our example, hopelessness is a system of negative expectations concerning one's future life and can cause depression. On the other hand, current depression can influence one's hopelessness towards the future [14, 15]. Thus, an interaction between hopelessness and depression can present to have a joint effect on suicidal ideation.

These examples highlight, how in many studies, the effect of the exposure can be complicated and is not necessarily linear, including the common analysis of drug interactions [16]. We bring together several past innovations to propose a novel analytic solution. [17] study the PAF when there are joint effects or interactions. [18] develop an approach of estimating logistic regression models with interactions and monotonicity constraints. The authors [18] apply the approach to estimating the PAF and achieve substantially more accurate estimates in some settings than the usual approach which uses logistic regression without monotonicity constraints.

Semiparametric approaches have been applied to estimate relative risk functions [19], to calculate odds ratios [20], and to estimate effect measures in the presence of interactions [21]. Herein, we use B-splines [22] (see also [23, 24] and references therein), to develop a semiparametric approach to estimate the PAF in the presence of interactions with confounding. The model fitting procedure is formulated as a well studied quadratic programming problem, and, thus, can be easily solved using standard optimization packages. We implement the approach using the *R* function *solve.QP* in the package *quadprog* [25, 26]. After a simulation study, we illustrate our new

method by examining hopelessness and depression for suicidal ideation among elderly depressed patients from the PROSPECT (Prevention of Suicide in Primary Care Elderly: Collaborative Trial) study [27]. Specifically, we model the interaction between hopelessness and depression under the monotonicity constraint.

Methods

We develop the semiparametric approach to estimate PAF accounting for interactions under the monotonicity constraint using B-splines in the last two subsections. We compare the performance of the following three approaches: the approach we developed (monB), the conventional B-splines (conB) approach without the monotonicity constraint but with the same knots, and the logistic regression approach (logit). Comparisons are made through simulation studies and a case study of estimating the PAF for suicidal ideation attributable to hopelessness or depression. To save computation time, we use a small number of basis functions, quadratic B-splines with knots placed at quantiles of the distribution of unique predictor values [28, P. 24]. We use quartiles {0, 1/4, 1/2, 3/4 and 1} as the knot locations.

Simulations

Let Y denote presence (1) or absence (0) of the outcome, Z the exposure, X a confounder, and V an additional covariate. The two interacting covariates Z and X are simulated independently from the Uniform $[0, 1]$ distribution where a 0 value is always part of the simulated z . The additional covariate V is simulated from a Bernoulli distribution with $p = 1/2$. The outcome is simulated from a Bernoulli distribution with the following probability functions, A: $0.1 + 0.4z + 0.3x - 0.2xz + 0.2v$; B: $2(0.1 + 2 \log(z/2 + 1))x/3 + 0.3v$; and C: $0.8\sqrt{z} + 0.1x^2 + 0.1v$. Shapes of the models at a fixed v are provided in Figure S1 of the supplementary material.

We examined 1000 simulations for each sample size 100 and 200. We compared the absolute value of the bias, the variance, and the mean squared error (MSE) of estimating the PAF attributable to the exposure Z . The true PAF values from these models are respectively 0.3, 0.4404, and 0.4616. They are calculated from (4) where the integrals are computed using *R* functions *integral* and *integral2* [29].

Illustrative case study

In the PROSPECT study, we focus on suicidal ideation four months after the beginning of the study. We consider the 592 patients in the study with no missing data. An event was observed if the score for suicidal ideation is greater than zero [27]. Following [30], we use the Beck Hopelessness Scale [31, BHS] to measure hopelessness, and the Beck Depression Inventory score [12, BDI] to

measure depression. The BHS and the BDI range from 0 to 19 and from 0 to 17, respectively, where a higher value means more hopelessness or depression.

Figure 1 shows the sample average of suicidal ideation by BDI and BHS scores, while Fig. 2 shows the corresponding patient frequency. Many patients with low BDI and BHS scores experienced no suicidal ideation. Overall, the risk of suicidal ideation increases with BDI and BHS scores, though the patient frequency decreases. The figures also show that high BHS scores are associated with high BDI scores. The PAF for hopelessness is the proportion of suicide ideation that would be prevented if all patients' hopelessness was reduced to 0 on the BHS scale, while keeping BDI fixed. Similarly, the PAF for depression is the proportion of suicide ideation that would be prevented if all patients' depression was reduced to 0 on the BDI scale, while BHS fixed.

Semiparametric estimation of the probability

Suppose Y is distributed in the exponential family with mean μ [32]. The logit link function connects μ with the exposure and the confounder through a smooth function f

as $\log[\mu/(1 - \mu)] = f(z, x)$. [18] models $f(z, x)$ parametrically assuming linearity as $f(z, x) = \beta_0 + \beta_1 z + \beta_2 x + \beta_3 x z$ and estimates the coefficients under the constraint that $\beta_1 z + \beta_3 x z \leq \beta_1 z^* + \beta_3 x z^*$ when $z \leq z^*$ at every x . We use a flexible semiparametric approach to model a possible non-linear $f(z, x)$ by linear combinations of B-splines basis functions [22] under the constraint $f(z, x) \leq f(z^*, x)$ when $z \leq z^*$ at every x . Let $\psi(z) = (\psi_1(z), \dots, \psi_p(z))'$ be a set of B-spline basis functions. Let $\phi(x) = (\phi_1(x), \dots, \phi_Q(x))'$ be another set of B-spline basis functions. We model $f(z, x)$ as

$$f(z, x) = \sum_{p=1}^P \sum_{q=1}^Q \psi_p(z) b_{p,q} \phi_q(x) \equiv \psi(z)' \mathbf{B} \phi(x), \mathbf{B}[p, q] = b_{p,q}, \tag{1}$$

where \mathbf{B} is the unknown coefficient matrix. Using Kronecker product \otimes and the vectorization operator vec described in Section S1 of the supplementary material, we further obtain

$$f(z, x) = (\phi(x)' \otimes \psi(z)') \text{vec} \mathbf{B} \equiv \varphi(z, x)' \beta,$$

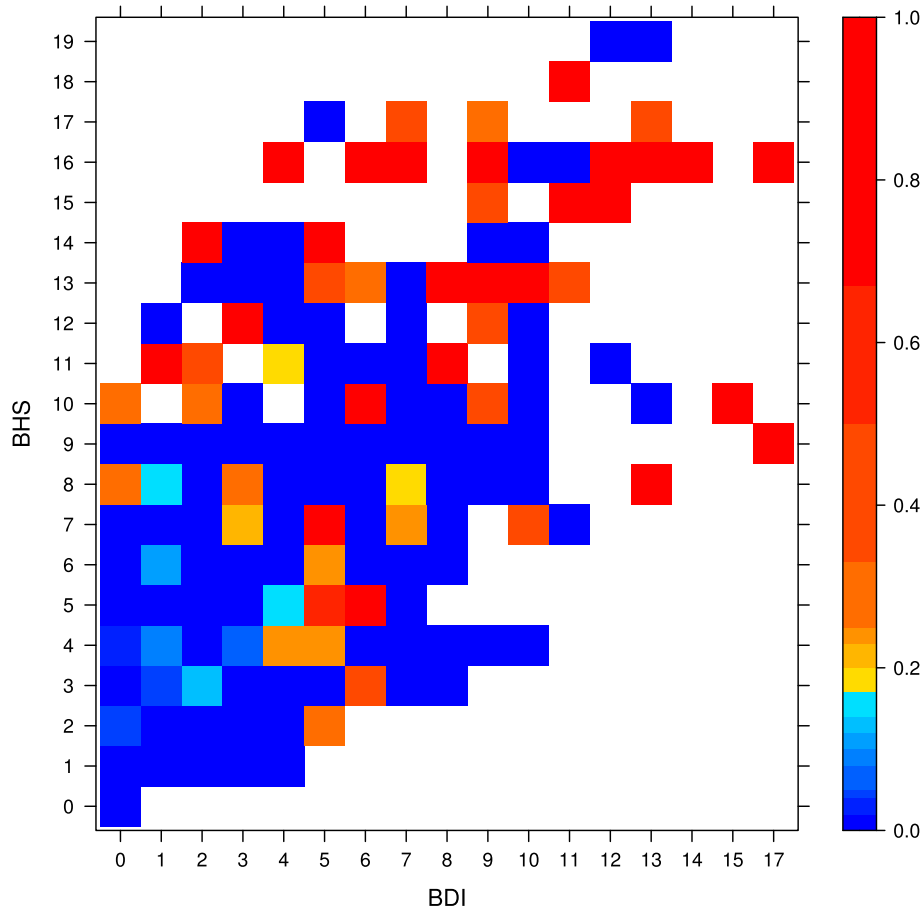


Fig. 1 Sample average of suicidal ideation by BDI and BHS scores

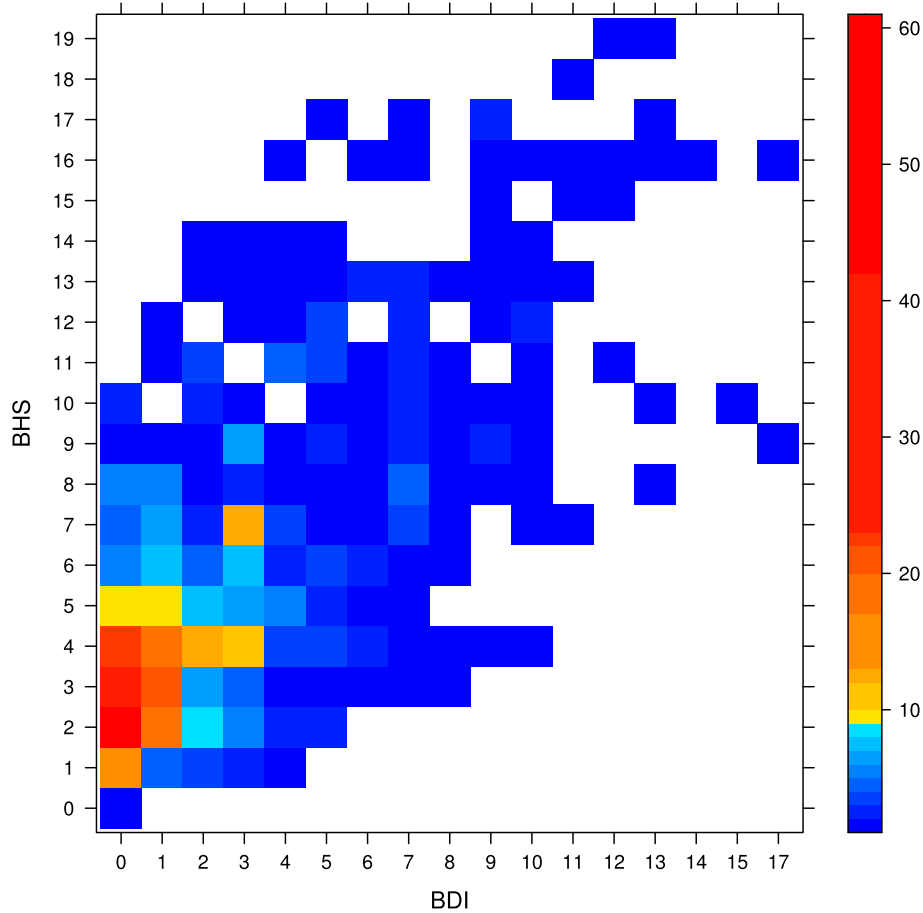


Fig. 2 Patient frequency by BDI and BHS scores

where $\varphi(z, x)' = \phi(x)' \otimes \psi(z)'$ and $\beta = \text{vec}B$. (2)

The maximum likelihood estimate of β without the constraint can be viewed as a modified iteratively weighted least squares problem [33, 34]. Derivation in Section S2 of the supplementary material shows that the constraint $f(z, x) \leq f(z^*, x)$ when $z \leq z^*$ at every x can be expressed as $A\beta \geq 0$, where the matrix A has a special pattern. The derivation in Section S3 of the supplementary material shows that the estimation procedure can be expressed as a quadratic programming problem. The approach is also capable of finding the estimate under the additional constraint that $f(z, x)$ is monotone in x .

Estimation of the PAF

To be general, let X be a vector of measured covariates. Let Y_0 denote what the presence or absence would be if the exposure were to be eliminated. Let P be the probability measure. In particular, let $P(Y_0 = 1)$ be the “hypothetical probability of disease in the same population but with all exposure eliminated” [35, 36]. The PAF for the exposure

Z is the proportion of disease that would be eliminated if $Z = 0$,

$$PAF = 1 - \frac{P(Y_0 = 1)}{P(Y = 1)} = 1 - \frac{\int P(Y_0 = 1|X = x)dP}{P(Y = 1)}. \tag{3}$$

To identify the PAF based on observed data, we make the assumption that all confounders of the disease-exposure relationship are measured and contained in X . We assume consequently that the ignorable treatment assignment [3] holds: $P(Y_0 = 1|X = x) = P(Y_0 = 1|X = x, Z = 0) = P(Y = 1|X = x, Z = 0)$. Under this assumption, the PAF can be written as

$$PAF = 1 - \frac{\int P(Y = 1|Z = 0, X = x)dP}{P(Y = 1)}, \tag{4}$$

which is equal to the expression (2.2) for the PAF in [36]. The PAF is also written as

$$PAF = \int \left[1 - \frac{P(Y = 1|Z = 0, \mathbf{X} = \mathbf{x})}{P(Y = 1|Z = z, \mathbf{X} = \mathbf{x})} \right] dP(\cdot|Y = 1), \tag{5}$$

where $P(\cdot|Y = 1)$ is the conditional probability given $Y = 1$ in the subpopulation of people with the disease [1, 5, 37]. We provide a proof of the equivalence between (4) and (5) in Section S4 of the supplementary material.

From a random sample of the population of size n , $i = 1, \dots, n$, an estimate of the PAF then follows as

$$p\hat{A}F = \frac{1}{\sum_{i=1}^n I_{\{Y_i=1\}}} \sum_{\{i: Y_i=1, 1 \leq i \leq n\}} \left[1 - \frac{\hat{P}(Y_i = 1|Z_i = 0, \mathbf{X}_i = \mathbf{x}_i)}{\hat{P}(Y_i = 1|Z_i = z_i, \mathbf{X}_i = \mathbf{x}_i)} \right], \tag{6}$$

where $I_{\{Y_i=1\}}$ is an indicator function [18]. Justification of the convergence of the estimate (6) to the PAF in (5) is provided by [6] in the scenario of no other covariates. A similar proof adjusting for the \mathbf{X} can also be derived. Thus, an accurate estimate of the probability of the disease would provide a reasonable estimate of the PAF. The defined PAF in (3) as a proportion has no practical meaning if $P(Y_0 = 1) > P(Y = 1)$. Similarly in (6), occurrences of $\hat{P}(Y_i = 1|Z_i = 0, \mathbf{X}_i = \mathbf{x}_i) > \hat{P}(Y_i = 1|Z_i = z_i, \mathbf{X}_i = \mathbf{x}_i)$ can result in an estimated PAF out of the $[0, 1]$ range. Hence, an estimate of the probability with the monotonicity constraint ensures a reasonable estimate of the PAF.

Results

Simulation

Table 1 summarizes the proportion of the times that the estimated PAF was within the $[0, 1]$ range for the three approaches. As discussed, the monotonicity constraint ensures the estimated PAF is within $[0, 1]$. The lack of a constraint results in PAF estimates that are sometimes out of the range using the logistic regression and the conventional B-splines approaches.

Table 2 shows the performance of the three approaches of estimating the PAF. In the comparison, results from the conventional B-splines approach are not shown due

Table 1 Proportion of the times out of the 1000 simulations that the estimated PAF is between 0 and 1

Sample size	Approach	Model		
		A	B	C
n=100	logit	0.867	0.521	0.402
	conB	0.326	0.221	0.250
	monB	1	1	1
n=200	logit	0.939	0.540	0.390
	conB	0.778	0.697	0.306
	monB	1	1	1

to its large scale. Instead, we show the results from this approach by censoring the original estimate at 0 if it is negative, or at 1 if it is bigger than 1. Similarly, we obtain the censored estimate from the logistic regression approach. In general, the censored estimate improves the original estimate. Overall, the developed approach outperformed the other approaches in the settings that we examined.

Case study

The estimated probability of suicidal ideation is shown in Fig. 3 separately by the three approaches. Both the logistic regression and the developed approach demonstrate a monotone relationship. Fitted probabilities numerically 0 or 1 occurred using the conventional B-splines approach. Table 3 summarizes the estimated PAF by the three approaches. The logistic regression produces lower PAF attributable to BHS and higher PAF attributable to BDI. Due to the numerical instability, the conventional B-splines estimated PAF is out of the $[0, 1]$ range. We examine the numerical instability in the “Discussion” section.

We use the bootstrap method to obtain the 95% confidence intervals of the estimates. The lower bound is the 2.5% quantile of the 1000 bootstrap estimates and the upper bound is the 97.5% quantile. The lower bound of the logistic regression estimated PAF attributable to BDI is negative. In fact, 28 of the estimated PAF’s are negative with the minimum being -32.45% . Among the 1000 estimates, two of the estimated PAF attributable to BHS are negative with the minimum being -4.27% .

Discussion

We used B-splines to develop a semiparametric estimate of the probability of an event under the monotonicity constraint and accounting for interactions. The approach is solved as a quadratic programming problem and can be easily implemented using the statistical software *R*. Using a boosting technique [38, 39] implement a similar approach to estimate β under the same constraint $A\beta \geq 0$ to solve the problem defined in equation (3) of the supplementary material. In the settings that have been tried, a comparison of fitting a univariate generalized linear model under the monotonicity constraint showed that the boosting algorithm is computationally intensive [23]. A summary of other approaches in the literature for estimation when there are monotonicity constraints and interactions is provided in Section S8 of the supplementary material.

Throughout our study, we placed the knots at the quartiles of the distribution of the unique predictors. We observed similar performance of the conventional B-splines approach and the developed approach when the knots are placed at the tertiles $\{0, 1/3, 2/3, 1\}$. In the

Table 2 Comparison of the absolute value of the bias ($|Bias|$), the variance and the MSE of estimating the PAF among the logistic regression approach (logit), the conventional B-splines approach (conB), and the developed approach (monB)

Sample size	Approach	$ Bias (\times 10^{-1})$			Variance ($\times 10^{-2}$)			MSE ($\times 10^{-2}$)		
		A	B	C	A	B	C	A	B	C
n=100	logit	1.63	3.62	8.75	10.80	37.19	94.47	13.45	50.24	170.94
	logit*	1.86	1.28	2.40	7.98	12.54	9.53	11.44	14.17	15.29
	conB*	1.74	2.84	2.64	5.18	10.09	12.86	8.20	18.17	19.80
	monB	1.48	1.53	1.65	3.69	4.94	5.25	5.88	7.28	7.96
	logit	2.07	3.67	8.53	6.27	26.80	68.97	10.53	40.25	141.63
n=200	logit*	2.14	1.74	2.85	5.35	9.79	6.96	9.93	12.81	15.09
	conB*	0.33	1.43	2.78	5.43	7.95	11.02	5.54	9.99	18.74
	monB	1.16	1.00	1.17	2.75	3.20	3.88	4.08	4.20	5.26

The logit* and conB* estimates are obtained by censoring the original estimates at 0 or at 1

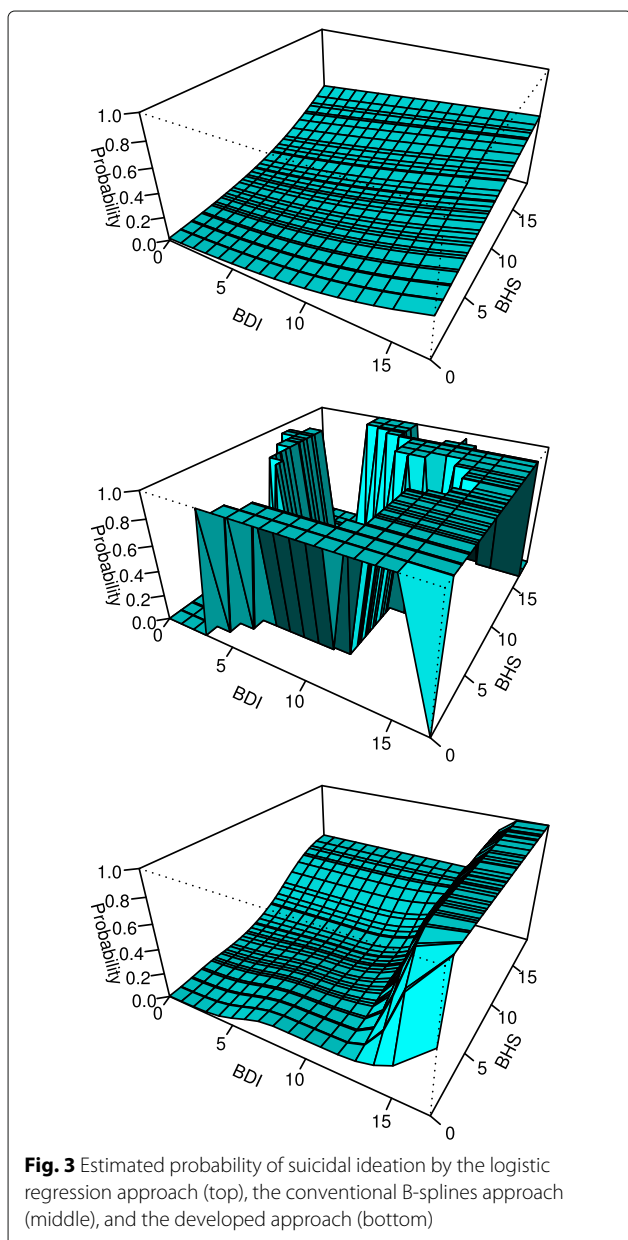


Fig. 3 Estimated probability of suicidal ideation by the logistic regression approach (top), the conventional B-splines approach (middle), and the developed approach (bottom)

simulation study, Tables S1 and S2 of the supplementary material show that overall the developed approach has better performance. In the case study, using the developed approach, the estimated PAF attributable to BHS was 67.66% (43.41%, 95.13%), and the estimated PAF attributable to BDI was 23.51% (10.92%, 55.62%) (Table S3). Similar numerical instability was observed using the conventional B-splines approach (Figure S2 and Table S3). The estimated PAF is also out of the defined range between 0 and 1 using the conventional approach when knots are placed at the quantiles {0, 1/2, 1} (Table S3). Performance of the developed approach is robust to the knots placement.

Conclusions

We developed a semiparametric estimate of the probability of an event using B-splines. Our approach can model a monotone relationship between the response and covariates, and can account for interactions. We applied the approach to estimate the PAF, and compared the performance of the estimator with the logistic regression approach and the conventional approach without the monotonicity constraint. Simulation studies showed that the developed estimator outperforms the other two approaches.

Table 3 Estimated PAF attributable to BHS and to BDI by the logistic regression approach (logit), the conventional B-splines approach (conB), and the developed approach (monB). The 95% confidence intervals are obtained from 2.5% and 97.5% quantiles of 1000 bootstrap estimates

Estimator	Estimated PAF	
	Attributable to BHS (%)	Attributable to BDI (%)
logit	63.57 (27.35, 82.20)	32.50 (-2.56, 58.13)
conB	< -100 (<-100, 100)	100 (<-100, 100)
monB	67.99 (42.10, 97.42)	22.36 (12.77, 56.49)

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12874-020-01118-4>.

Additional file: The supplementary material contains the following sections: S1, a review of the Kronecker product and the vectorization operator; S2 and S3, technical details of the developed approach; S4, proof of the equivalence between Eqs. (4) and (5); S5, a figure of the models used in the simulation studies; S6, additional simulation results where the knots are placed at the tertiles; S7, additional data analysis results where the knots are placed at the tertiles and at the quantiles {0, 1/2, 1}; S8 literature review of relevant statistical methods; and S9, R code to implement the developed approach.

Abbreviations

BDI: Beck depression inventory; BHS: Beck hopelessness scale; conB: Conventional B-splines approach without the monotonicity constraint; logit: Logistic regression; monB: Developed monotone B-splines approach; PAF: Population attributable fraction.

Acknowledgements

The authors thank the two anonymous reviewers whose comment helped improving the overall quality of the manuscript.

Authors' contributions

All the authors made substantial contributions to the manuscript. WW conducted the data analysis and simulation studies, and drafted the manuscript. DSS and MOH reviewed it and revised it. All three authors approved the manuscript to be submitted.

Funding

Michael O. Harhay was supported by grant R00 HL141678 from the National Heart, Lung, and Blood Institute (NHLBI) of the US National Institutes of Health (NIH). The funding body had no role in study design, data collection and analysis, or decision to publish.

Availability of data and materials

The R code to implement the developed approach is available in the supplementary material. The data that support the findings of this study are available from Martha Bruce, but are not publicly available. Data are however available from the authors upon reasonable request and with permission of Martha Bruce.

Ethics approval and consent to participate

Data from the PROSPECT (Prevention of Suicide in Primary Care Elderly: Collaborative Trial) study [27] was used for this study. The research protocol received full review and approval from the institutional review board of each of the 3 universities participating in the PROSPECT trial and written informed consent was obtained for all participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ²Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Biostatistics, Epidemiology, and Informatics, Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

Received: 5 May 2020 Accepted: 9 September 2020

Published online: 21 September 2020

References

1. Deubner DC, Wilkinson WE, Helms MJ, Tyroler HA, Hames CG. Logistic model estimation of death attributable to risk factors for cardiovascular disease in Evans County, Georgia. *Am J Epidemiol.* 1980;112(1):135–43.
2. Rothman KJ, Greenland S. Causation and causal inference. In: Rothman KJ, Greenland S, editors. *Modern Epidemiology*. Philadelphia, Pennsylvania: Lippincott, Williams & Wilkins; 1998.
3. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.
4. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol.* 1985;122(5):904–14.
5. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. *Biometrics.* 1993;49(3):865–72.
6. Wang W, Small D. A comparative study of parametric and nonparametric estimates of the attributable fraction for a semi-continuous exposure. *Int J Biostat.* 2012;8(1):1–22.
7. Kelly C, Rice J. Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics.* 1990;46(4):1071–85.
8. Kong M, Eubank RL. Monotone smoothing with application to dose-response curve. *Commun Stat Simul Comput.* 2006;35:991–1004.
9. Beck AT. *Depression: Clinical, Experimental, and Theoretical Aspects*. New York: Hoeber Medical Division, Harper & Row; 1967, p. 370.
10. Wetzel RD. Hopelessness, depression, and suicide intent. *Arch Gen Psychiatr.* 1976;33(9):1069–73.
11. Beck AT, Steer RA, Kovacs M, Garrison B. Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. *Am J Psychiatr.* 1985;142(5):559–63.
12. Beck AT, Brown G, Berchick RJ, Stewart BL, Steer RA. Relationship between hopelessness and ultimate suicide: A replication with psychiatric outpatients. *Am J Psychiatr.* 1990;147(2):190–5.
13. Kong M, Lee JJ. A generalized response surface model with varying relative potency for assessing drug interaction. *Biometrics.* 2006;62:986–95.
14. Stotland E. *The Psychology of Hope*. San Francisco, California: Jossey-Bass; 1969, p. 284.
15. Beck AT, Kovacs M, Weissman A. Hopelessness and suicidal behavior: an overview. *J Am Med Assoc.* 1975;234(11):1146–9.
16. Kong M, Lee JJ. A semiparametric response surface model for assessing drug interaction. *Biometrics.* 2008;64:396–405.
17. Hössjer O, Kockum I, Alfredsson L, Hedström AK, Olsson T, Lekman M. A general framework for and new normalization of attributable proportion. *Epidemiol Methods.* 2017;6(1):1–25.
18. Traskin M, Wang W, Ten Have TR, Small DS. Efficient estimation of the attributable fraction when there are monotonicity constraints and interactions. *Biostatistics.* 2013;14(1):173–88.
19. Zhao LP, Kristal AR, White E. Estimating relative risk functions in case-control studies using a nonparametric logistic regression. *Am J Epidemiol.* 1996;144(6):598–609.
20. Figueiras A, Cadarso-Suárez C. Application of nonparametric models for calculating odds ratios and their confidence intervals for continuous exposures. *Am J Epidemiol.* 2001;154(3):264–75.
21. Cadarso-Suárez C, Roca-Pardiñas J, Figueiras A. Effect measures in non-parametric regression with interactions between continuous exposures. *Stat Med.* 2006;25(4):603–21.
22. de Boor C. *A Practical Guide to Splines*. New York: Springer; 1978.
23. Wang W, Small D. Monotone B-spline smoothing for a generalized linear model response. *Am Stat.* 2015;69(1):28–33.
24. Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M. A review of spline function procedures in R. *BMC Med Res Methodol.* 2019;19(1):46.
25. Goldfarb D, Idrani A. Dual and primal-dual methods for solving strictly convex quadratic programs. In: Hennart JP, editor. *Numerical Analysis*. Berlin: Springer; 1982. p. 226–39.
26. Goldfarb D, Idrani A. A numerically stable dual method for solving strictly convex quadratic programs. *Math Program.* 1983;27:1–33.
27. Bruce ML, Ten Have TR, Reynolds CF, Katz IL, Schulberg HC, Mulsant BH, Brown GK, McAvay GJ, Pearson JL, Alexopoulos GS. Reducing suicidal ideation and depressive symptoms in depressed older primary care patients: a randomized controlled trial. *J Am Med Assoc.* 2004;291(9):1081–91.

28. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Boca Raton, Florida, USA: Chapman & Hall/CRC; 1990.
29. Borchers HW. *Pracma: Practical Numerical Math Functions*. 2019. R package version 2.2.9. <https://CRAN.R-project.org/package=pracma>. Accessed 15 Dec 2019.
30. Cheung YB, Law CK, Chan B, Liu KY, Yip PS. Suicidal ideation and suicidal attempts in a population-based study of Chinese people: risk attributable to hopelessness, depression, and social factors. *J Affect Disord*. 2006;90(2-3):193–9.
31. Beck AT, Steer RA. *Manual for the Beck hopelessness scale*. San Antonio, TX: Psychological Corporation; 1988.
32. McCullagh P, Nelder JA. *Generalized Linear Models*, 2nd edn. Boca Raton, Florida: Chapman & Hall; 1989.
33. O'Sullivan F, Yandell BS, Raynor WJJ. Automatic smoothing of regression functions in generalized linear models. *J Am Stat Assoc*. 1986;81(393):96–103.
34. Eilers PHC, Marx BD. Flexible smoothing with b-splines and penalties. *Stat Sci*. 1996;11:89–121.
35. Benichou J. A review of adjusted estimators of attributable risk. *Stat Methods Med Res*. 2001;10:195–216.
36. Sjölander A, Vansteelandt S. Doubly robust estimation of attributable fractions. *Biostatistics*. 2011;12:112–21.
37. Benichou J, Gail MH. Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. *Biometrics*. 1990;46(4):991–1003.
38. Hofner B, Mayr A, Robinzonov N, Schmid M. Model-based boosting in R: a hands-on tutorial using the R package mboost. *Comput Stat*. 2014;29:3–35.
39. Hofner B, Kneib T, Hothorn T. A unified framework of constrained regression. *Stat Comput*. 2016;26:1–14.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

