Title

Joint Indirect Standardization When Only Marginal Distributions are Observed in the Index Population

Journal

Journal of the American Statistical Association, 114(526)

ISSN

0162-1459

Authors

Wang, Yifei
Tancredi, Daniel J
Miglioretti, Diana L

Publication Date

2019-04-03

DOI

10.1080/01621459.2018.1506340

Peer reviewed

HHS Public Access

Author manuscript

J Am Stat Assoc. Author manuscript; available in PMC 2020 January 01.

Published in final edited form as:

J Am Stat Assoc. 2019 ; 114(526): 622–630. doi:10.1080/01621459.2018.1506340.

Joint Indirect Standardization when Only Marginal Distributions are Observed in the Index Population

Yifei Wang,
Department of Radiology, University of California, San Francisco

Daniel J. Tancredi,
Department of Pediatrics, University of California, Davis

Diana L. Miglioretti
Department of Public Health Sciences, University of California, Davis

Abstract

It is a common interest in medicine to determine whether a hospital meets a benchmark created from an aggregate reference population, after accounting for differences in distributions of multiple covariates. Due to the difficulties of collecting individual-level data, however, it is often the case that only marginal distributions of the covariates are available, making covariate-adjusted comparison challenging.

We propose and evaluate a novel approach for conducting indirect standardization when only marginal covariate distributions of the studied hospital are known, but complete information is available for the reference hospitals. We do this with the aid of two existing methods: iterative proportional fit, which estimates the cells of a contingency table when only marginal sums are known, and synthetic control methods, which create a counterfactual control group using a weighted combination of potential control groups. The proper application of these existing methods for indirect standardization would require accounting for the statistical uncertainties induced by a situation where no individual-level data is collected from the studied population. We address this need with a novel method which uses a random Dirichlet parametrization of the synthetic control weights to estimate uncertainty intervals for the standard incidence ratio.

We demonstrate our novel methods by estimating hospital-level standardized incidence ratios for comparing the adjusted probability of computed tomography examinations with high radiations doses, relative to a reference standard and we evalauate out methods in a simulation study.

Keywords

Hospital Profiling; Iterative Proportional Fit; Synthetic Control; Survey Sampling; Causal Inference

1 Introduction

For assessing hospital performance (g.e., hospital profiling), it is often of interest to compare the incidence of a binary outcome in an index hospital to the expected incidence of this outcome based on patterns of a wider population (Ash et al. (2011)), represented by a

collection of multiple well-studied reference hospitals. The outcome incidences may be affected by covariates, such as patient characteristics, whose distributions may vary greatly between hospitals (Christiansen and Morris (1997), Woodard et al. (2007)), motivating the need to adjust for them before fair comparisons can be made.

One commonly-used method for making this adjusted comparison is indirect standardization, using the "Standardized Mortality Ratio" or "Standardized Incidence Ratio" (SIR), given the first name because it was developed to compare death rates (Farr (1859)). Traditional indirect standardization methods generally adjust for a single confounding variable, but often multiple covariates must be considered. For example, for radiation safety quality assurance, a hospital may wish to evaluate the probability of computed tomography exams with high radiation dose relative to other hospitals while adjusting for several patient factors such as body size, age, and sex (Smith-Bindman et al. (2015)).

If the full joint distribution of all covariates were known, such a problem could still be addressed using existing indirect standardization methods by applying a risk prediction model that is a function of this full joint distribution. However, the index hospital may not have the capacity to collect, or may not have the willingness or permission to share, individual-level patient data, making it challenging to observe even a sample from this joint distribution. On the other hand, index marginal covariate distributions may still be observed or approximated, and may be used along with supplementary data from reference hospitals to estimate the index joint covariate distribution.

A related challenge exists in survey sampling, where individual-level data of multiple covariates are sampled from a population of interest, and the population covariate marginal distributions are known from auxiliary sources. The goal is to estimate the joint covariate distribution of the population by adjusting the joint covariate distribution observed in the sample. This adjusted estimate should be consistent with known population marginal distributions but minimize divergence from the sampled joint distribution. This problem was first popularized and addressed through a computationally simple algorithm called iterative proportional fit, or raking (Deming and Stephan (1940)). Alternative approaches (Stephan (1942)) and expansions (Deville et al. (1993)) have been proposed, but traditional raking continues to be commonly used (Singh and Rao (1995)).

One important issue, however, limits our ability to directly use raking and related methods for hospital profiling. In survey sampling, the population of interest and the population from which individual-level data are drawn are the same or assumed to be very similar. This assumption cannot be made about our index and reference populations. While some reference hospitals may behave similarly to the index hospital, assuming the same for all reference hospitals or for "the average hospital" may be inappropriate when the reference hospitals are very different from each other.

One solution to this is viewing the index hospital as similar to a weighted combination of the reference hospitals. In causal inference, synthetic control methods (Abadie et al. (2010)) weight clusters of a data set combined from multiple sources to create a control group comparable to the population of interest. The motivating goal of this method is to evaluate

an intervention whose effect cannot be observed through experimentation, such as estimating the effect of enacted gun control law in Connecticut (Rudolph et al. (2015)). This is done by developing a counterfactual state, formed by a weighted combination of states which did not enact gun control to synthesize a "control group Connecticut." Our hospital profiling problem does not share this motivation, and direct application of synthetic control methods would require information of our index hospital which is more difficult to obtain (sometimes impossible obtain). Nonetheless, we would like to weight reference hospitals to synthesize a variety of possible control groups, or counterfactual hospitals, for the index hospital.

In this paper, we propose an approach for estimating the covariate-adjusted expected probability of a binary outcome in an index hospital for which only marginal covariate distributions are observed. Our approach exploits the known joint distributions in each of a collection of reference hospitals, using concepts drawn from two traditionally unrelated statistical methodologies, iterative proportional fit and synthetic control methods. This paper begins with a mathematical description of the hospital profiling problem, then reviews iterative proportional fit and the inadequacies it presents in this context, addresses these inadequacies by exploring ideas drawn from synthetic control methods, and proposes a novel random parametrization of synthetic control weights to overcome remaining limitations. We demonstrate our novel methodology by applying to an example dataset of radiation doses from computed tomography exams performed at 151 hospitals and evaluate our methods with a simulation study. We conclude with an exploration of possible future extensions and the challenges of making the methods presented in this paper accessible to its predominantly non-statistical audience.

2 Mathematical Description of the Problem

We observe a binary outcome, Y, in both an index hospital and a reference population, consisting of a large number of reference hospitals. This outcome is influenced by C covariates, called $\boldsymbol{X}$, whose distributions vary between hospitals. Any fair comparison of $\Pr(Y)$ between hospitals must account for differences in $\boldsymbol{X}$.

For simplicity, and consistent with existing work in this area, we assume all covariates are categorical, with covariate $c \in 1,\ldots, C$ having L_c categories. Any continuous covariates may be binned in categorical covariates. Let $\boldsymbol{\pi}_j = \{\pi_{j|l_1,l_2,\ldots,l_C}\}_{l_c \in 1,\ldots,L_c}$ denote the joint probability vector of $\boldsymbol{X}$ in the j^{th} reference hospital, $j \in 1,\ldots, J$. We order the elements of $\boldsymbol{\pi}_j$ lexicographically such that element $\pi_{j|l_1,l_2,\ldots,l_C} = \Pr(X_1 = l_1, X_2 = l_2, \ldots, X_C = l_C)$ in reference hospital j. That is, it denotes the probability of belonging to category $l_1 \in 1,\ldots, L_1$ of the first covariate, category $l_2 \in 1,\ldots, L_2$ of the second covariate, and so on.

Let $\tilde{p}_{l_1,l_2,\ldots,l_C} = \Pr(Y|X_1 = l_1, X_2 = l_2, \ldots, X_C = l_C$ in the overall reference population, and let vector $\tilde{\boldsymbol{p}} = \{\tilde{p}_{l_1,l_2,\ldots,l_C}\}_{l_c \in 1,\ldots,L_c}$ be indexed identically to $\boldsymbol{\pi}_j$. These probabilities can be computed in several ways. The most direct way is using the observed outcome proportions in the reference data for each covariate combination, assuming a large enough sample to allow precise estimation. In the case of rare outcomes or small sample sizes, $\tilde{\boldsymbol{p}}$ may be estimated using multi-level logistic regression (Ash et al. (2011)).

In the index hospital, let $p = \Pr(Y)$ and $\tilde{\mathrm{E}}[p]$ be the expectation of $\Pr(Y)$ if $\Pr(Y|\boldsymbol{X}) = \tilde{\boldsymbol{p}}$. The value $p/\tilde{\mathrm{E}}[p]$ is called the SIR. Traditional indirect standardization views $\tilde{\mathrm{E}}[p]$ as fixed, while p is estimated using the observed proportion of Y in the index; call this observed value $\hat{p}$. There are many existing methodologies for quantifying the uncertainty in $\hat{p}$, but in our hospital profiling problem, we shall see that the $\hat{p}$ is not the only source of uncertainty in the SIR, as $\tilde{\mathrm{E}}[p]$ is also not observed and thus must be estimated with some uncertainty. Herein lies a unique challenge which has not been addressed in traditional indirect standardization.

The source of this uncertainty is $\boldsymbol{\pi}_0 = \{\pi_{0|l_1,l_2,\ldots,l_C}\}_{l_c \in 1,\ldots,L_c}$, which we use to denote the index joint covariate probability vector, indexed identically to $\boldsymbol{\pi}_j$. When $\boldsymbol{\pi}_0$ is observed, one may simply take $\tilde{\boldsymbol{p}}' \times \boldsymbol{\pi}_0$ as the value of $\tilde{\mathrm{E}}[p]$. If this induces an SIR significantly greater than 1, we may say the outcome occurs more often than expected given the patient population. In our motivating problem, however, no individual-level data is collected from the index hospital, leaving $\boldsymbol{\pi}_0$ unobserved. We instead observe only the index covariate marginal probabilities. Denote these marginals as vector $\boldsymbol{\pi}_{\mathbf{0}}|+$, which lists relative frequencies of each category for the first covariate, then relative frequencies of each category for the second covariate, and so on. Let $\mathbf{A}$ be the matrix of 1s and 0s which transforms $\boldsymbol{\pi}_0$ to its marginal summaries $\boldsymbol{\pi}_{\mathbf{0}}|+$, i.e. $\mathbf{A} \times \boldsymbol{\pi}_0 = \boldsymbol{\pi}_{\mathbf{0}}|+$. For example, in the case where there are two covariates, each with two categories, we have

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \boldsymbol{\pi}_0 = \begin{bmatrix} \pi_{0|1,1} \\ \pi_{0|1,2} \\ \pi_{0|2,1} \\ \pi_{0|2,2} \end{bmatrix} \boldsymbol{\pi}_{\mathbf{0}}|+ = \begin{bmatrix} \pi_{0|1,+} \\ \pi_{0|2,+} \\ \pi_{0|+,1} \\ \pi_{0|+,2} \end{bmatrix}$$

where $p_{l_1,+} = \sum_{l_2=1}^{L_2} p_{l_1,l_2}$ and $p_{+,l_2} = \sum_{l_1=1}^{L_1} p_{l_1,l_2}$ for any vector $\boldsymbol{p}$ indexed as $\{p_{l_1,l_2}\}_{l_1 \in 1,\ldots,L_1, l_2 \in 1,\ldots,L_2}$.

The goal of this paper is to synthesize a hospital counterfactual to the index - call its joint covariate probability vector $\hat{\boldsymbol{\pi}}$ - such that it is reasonable to estimate $\tilde{\mathrm{E}}[p]$ using $\tilde{\boldsymbol{p}}' \times \hat{\boldsymbol{\pi}}$, and the SIR $(p/\tilde{\mathrm{E}}[p])$ using $\hat{p} \,/\, [\tilde{\boldsymbol{p}}' \times \hat{\boldsymbol{\pi}}]$. That is to say, $\hat{\boldsymbol{\pi}}$ should have the following properties

1. For consistency with existing knowledge of the index, we must have $\mathbf{A} \times \hat{\boldsymbol{\pi}} = \boldsymbol{\pi}_{0|+}$.
2. An uncertainty interval created for $p \,/\, [\tilde{\boldsymbol{p}}' \times \boldsymbol{\pi}_0]$ using $\hat{p} \,/\, [\tilde{\boldsymbol{p}}' \times \hat{\boldsymbol{\pi}}]$ must have the same probability of including $p \,/\, [\tilde{\boldsymbol{p}}' \times \boldsymbol{\pi}_0]$ as a confidence interval (of the same level) created using $\hat{p} \,/\, [\tilde{\boldsymbol{p}}' \times \boldsymbol{\pi}_0]$.

3 Using Observed Information from the Index

The problem of finding $\hat{\boldsymbol{\pi}}$ such that $\mathbf{A} \times \hat{\boldsymbol{\pi}} = \boldsymbol{\pi}_{0|+}$ can be addressed using an iterative process called the iterative proportional fit algorithm, or raking. The original work (Deming and Stephan (1940)) surrounding raking first considered the case with only two covariates.

In this context, raking operates by placing all elements of $\hat{\boldsymbol{\pi}}^{(0)}$ - the initial estimate - into a contingency table, with L_1 rows denoting the categories of the first covariate and L_2 columns denoting the categories of the second. Each row is then multiplied by whatever number is required to meet the known marginal totals for the first covariate induced by $\boldsymbol{\pi}_{\mathbf{0}|+}$. The same is then done for the columns, meeting known marginal totals for the second covariate. The process then alternates between adjusting the rows, then columns, until marginal totals for both covariates are met simultaneously.

To put this in mathematical terms, let $\hat{\boldsymbol{\pi}}^{(t)} = \{\hat{\pi}^{(t)}_{l_1, l_2}\}_{l_1 \in 1, \ldots, L_1, l_2 \in 1, \ldots, L_2}$ denote the candidate for $\hat{\boldsymbol{\pi}}$ on the t^{th} iteration. The initial candidate, $\hat{\boldsymbol{\pi}}^{(0)}$, is formed from a sample of the population of interest. The first two iterative steps are:

$$\begin{aligned} \hat{\pi}^{(1)}_{l_1 l_2} &= \hat{\pi}^{(0)}_{l_1 l_2} \cdot \frac{\pi_{0 \mid l_1, +}}{\left[\mathbf{A} \times \hat{\boldsymbol{\pi}}^{(0)}\right]_{l_1, +}} \\ \hat{\pi}^{(2)}_{l_1 l_2} &= \hat{\pi}^{(1)}_{l_1 l_2} \cdot \frac{\pi_{0 \mid +, l_2}}{\left[\mathbf{A} \times \hat{\boldsymbol{\pi}}^{(1)}\right]_{+, l_2}} \end{aligned} \quad (1)$$

This forms a cycle which is repeated, with $\hat{\pi}^{(0)}$, $\hat{\pi}^{(1)}$, $\hat{\pi}^{(2)}$ replaced by $\hat{\pi}^{(s)}$, $\hat{\pi}^{(s+1)}$, $\hat{\pi}^{(s+2)}$ for even-valued s. The algorithm stops when $\mathbf{A} \times \hat{\boldsymbol{\pi}}^{(T)}$ is sufficiently close to $\boldsymbol{\pi}_{\mathbf{0}|+}$ for some T. The convergence of this algorithm has been proven at varying degrees of generalization by multiple authors (Bishop (1967), Ireland and Kullback (1968), Fienberg (1970), Ruschendorf (1995)), though it is necessary that $\hat{\boldsymbol{\pi}}^{(0)}$ has no elements equal to zero; this is a restriction we shall place as well. In any cases where $\hat{\boldsymbol{\pi}}^{(0)}$ would be observed to have elements equal to zero, we set the zero-values to 10^{-10} instead, then normalize so that the elements of $\hat{\boldsymbol{\pi}}^{(0)}$ add up to one.

The primary utility of raking is to find an estimate $\hat{\boldsymbol{\pi}}^{(T)}$ that not only satisfies $\mathbf{A} \times \hat{\boldsymbol{\pi}}^{(T)} = \boldsymbol{\pi}_{0 \mid +}$, but also minimizes the Kullback-Leibler Divergence from $\hat{\boldsymbol{\pi}}^{(0)}$ (Ireland and Kullback (1968)). In the process of minimizing this divergence, raked estimates obtain a number of other useful properties. Most importantly for this paper, as odd steps deal only with restrictions set by the first covariate and even steps only the second, any iteration, including the solution, may be written as

$$\hat{\pi}^{(t)}_{l_1 l_2} = \delta^{(t)}_{1, l_1} \cdot \delta^{(t)}_{2, l_2} \cdot \hat{\pi}^{(0)}_{l_1 l_2}$$

Where

$$\delta_{1,l_1}^{(t)} = \prod_{u \le t,\, u \bmod 2 = 0} \frac{\pi_{0 \mid l_1,\, +}}{\left[\mathbf{A} \times \hat{\boldsymbol{\pi}}^{(u)}\right]_{l_1,\, +}}$$
$$\delta_{2,l_2}^{(t)} = \prod_{u \le t,\, u \bmod 2 = 1} \frac{\pi_{0 \mid\, +,\, l_2}}{\left[\mathbf{A} \times \hat{\boldsymbol{\pi}}^{(u)}\right]_{+,\, l_2}} \quad (2)$$

Later authors (Deville and Särndal (1992), Deville et al. (1993)) extended this algorithm to cases with C covariates using different notation, with each cycle having C operations instead of the two operations in equation 1, giving the relationship

$$\hat{\pi}_{l_1,\ldots,l_C}^{(t)} = \delta_{1,l_1}^{(t)} \cdot \delta_{2,l_2}^{(t)} \cdot \ldots \cdot \delta_{C,l_C}^{(t)} \cdot \hat{\pi}_{l_1,\ldots,l_C}^{(0)} \quad (3)$$

The $\delta^{(t)}$ values are defined similarly to equation 2, except the product conditions on u become $u \le t$, $u \bmod C = 0,1, \ldots, C-1$.

From this, we see that raking preserves the cross-product ratios of $\hat{\boldsymbol{\pi}}^{(0)}$ (Mosteller (1968)), a measure of association between covariates. By equation 3, without loss of generality, observe that for any pairs of categories from the first two covariates $a_1, a_2 \in 1, \ldots, L_1$, $b_1, b_2 \in 1, \ldots, L_2$, and $l_c \in 1, \ldots, L_c \; \forall \; c \in 3, 4, \ldots, C$. We have, for any t,

$$\begin{aligned}\frac{\hat{\pi}_{a_1,b_1,l_3,\ldots,l_C}^{(t)} \cdot \hat{\pi}_{a_1,b_2,l_3,\ldots,l_C}^{(t)}}{\hat{\pi}_{a_2,b_1,l_3,\ldots,l_C}^{(t)} \cdot \hat{\pi}_{a_2,b_2,l_3,\ldots,l_C}^{(t)}} &= \frac{\hat{\pi}_{a_1,b_1,l_3,\ldots,l_C}^{(0)} \cdot \hat{\pi}_{a_1,b_2,l_3,\ldots,l_C}^{(0)}}{\hat{\pi}_{a_2,b_1,l_3,\ldots,l_C}^{(0)} \cdot \hat{\pi}_{a_2,b_2,l_3,\ldots,l_C}^{(0)}} \\ &\quad \cdot \frac{\delta_{1,a_1}^{(t)} \cdot \delta_{2,b_1}^{(t)} \cdot \delta_{1,a_1}^{(t)} \cdot \delta_{2,b_2}^{(t)}}{\delta_{1,a_2}^{(t)} \cdot \delta_{2,b_1}^{(t)} \cdot \delta_{1,a_2}^{(t)} \cdot \delta_{2,b_2}^{(t)}} \cdot \frac{\left(\prod_{c=3}^{C} \delta_{c,l_c}^{(t)}\right)^2}{\left(\prod_{c=3}^{C} \delta_{c,l_c}^{(t)}\right)^2} \quad (4) \\ &= \frac{\hat{\pi}_{a_1,b_1,l_3,\ldots,l_C}^{(0)} \cdot \hat{\pi}_{a_1,b_2,l_3,\ldots,l_C}^{(0)}}{\hat{\pi}_{a_2,b_1,l_3,\ldots,l_C}^{(0)} \cdot \hat{\pi}_{a_2,b_2,l_3,\ldots,l_C}^{(0)}}\end{aligned}$$

In the case of $C = L_1 = L_2 = 2$, the cross-product ratio is also known as the odds ratio.

The question then arises as to whether it is desirable to preserve the cross-product ratios of $\hat{\boldsymbol{\pi}}^{(0)}$. If it can be assumed that $\hat{\boldsymbol{\pi}}_0$ shares the cross-product ratios of $\hat{\boldsymbol{\pi}}^{(0)}$, then setting $\hat{\boldsymbol{\pi}} = \hat{\boldsymbol{\pi}}^{(T)}$ would make $\hat{\boldsymbol{\pi}}$ an exact estimate of $\hat{\boldsymbol{\pi}}_0$, satisfying both enumerated requirements set out in section 2. In traditional raking, $\hat{\boldsymbol{\pi}}^{(0)}$ is estimated by a sample of individual-level data from the population in interest. In our motivating problem, the population of interest is the index hospital, from which we have no sample of individual-level data.

We instead have individual-level data from a collection of reference hospitals which have a variety of cross-product ratios represented, many of which are quite different from the cross-product ratio of the overall reference population. Thus, assuming the index comes from the same population of hospitals as the reference hospitals, it would be inappropriate to preserve the cross-product ratios of a $\widehat{\boldsymbol{\pi}}^{(0)}$ computed using the overall reference population dataset. In fact, to assume that the index hospital shares its cross-product ratios with any fixed $\widehat{\boldsymbol{\pi}}^{(0)}$ seems unwarranted and would be untestable, as the index covariate cross-product ratios are unobserved. Therefore, an inferential model on the SIR in the context of our motivating problem must allow for some uncertainty in the choice of $\widehat{\boldsymbol{\pi}}^{(0)}$, or in other words, uncertainty in the denominator of the SIR.

4 Using Supplementary Data from the Reference

To model uncertainty in the selection of $\widehat{\boldsymbol{\pi}}^{(0)}$, we look to the individual reference hospitals, rather than the full reference dataset, for candidates. Let Θ_j denote the set of all cross-product ratios of covariates in the j^{th} reference hospital, and let Θ_0 denote the unknown set of all cross-product ratios of covariates in the index. The collection $\Theta_1, \ldots, \Theta_J$ provides a collection of candidate values for Θ_0.

Noting that the combination of known marginal summaries and known cross-product ratios induces a unique joint distribution, we see from section 3 that given $\boldsymbol{\pi}_{0|+}$ and the assumption of $\Theta_0 = \Theta_j$ for fixed $j \neq 0$, we may directly compute $\boldsymbol{\pi}_0$ by using $\boldsymbol{\pi}_j$ as the initial estimate in a raking procedure. Refer to the joint probability vector produced this way as $\widetilde{\boldsymbol{\pi}}_j$; let $\widetilde{\boldsymbol{\Pi}}$ be a matrix where the j^{th} column is $\widetilde{\boldsymbol{\pi}}_j$.

The statement $\Theta_0 = \Theta_j$ is unlikely to be true for any single j ≠ 0, making it also unlikely that any $\widetilde{\boldsymbol{\pi}}_j = \boldsymbol{\pi}_0$. However, it is also true that Θ_0 would be more similar to some $\{\Theta_j\}_{j\neq 0}$ than to others, so it becomes natural, in the process of synthesizing $\widehat{\boldsymbol{\pi}}$ similar to $\boldsymbol{\pi}_0$, to place greater importance on some $j \neq 0$ than others. That is, have

$$\widehat{\boldsymbol{\pi}} = \widetilde{\boldsymbol{\Pi}} \times \mathbf{w}, \qquad w_j \geq 0 \; \forall \; j \in 1, \ldots, J \text{ and } \sum_{j=1}^{J} w_j = 1 \quad (5)$$

leading to the estimate

$$\hat{p} \,/\, \widehat{\mathrm{E}}[p] = \hat{p} \,/\, [\widetilde{\boldsymbol{p}}' \times \widetilde{\boldsymbol{\Pi}} \times \mathbf{w}] \quad (6)$$

for the SIR ($p/\tilde{\mathrm{E}}[p]$). Note that since $\mathbf{A} \times \widetilde{\boldsymbol{\pi}}_j = \boldsymbol{\pi}_{0\,|\,+} \quad \forall\, j \neq 0$ we have $\mathbf{A} \times \widetilde{\boldsymbol{\Pi}} \times \mathbf{w} = \mathbf{A} \times \widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}_{0\,|\,+}$ for all possible values of $\mathbf{w}$.

If the index hospital differs sufficiently from the reference hospitals such that no weighted combination of reference hospitals could reflect the index, the formulation in equation 5

would preclude $\hat{\boldsymbol{\pi}}$ from precisely estimating $\boldsymbol{\pi}_0$. Thus, the set of reference hospitals must be sufficiently representative of any potential hospitals of interest, so that the convex hull of $\{\tilde{\boldsymbol{\pi}}\}_{j \neq 0}$ encompasses most reasonable values the index covariate distributions can take.

Synthetic control methods (Abadie et al. (2010)) address a similar issue of weighting a set of auxiliary populations to form a counterfactual for the population of interest. In its standard application, set of predictive values required of the counterfactual (that is, the analog of $\boldsymbol{\pi}_0$) can take many forms, including marginals summaries (such as age distribution), population characteristics (such as the GDP of a state), and time-sensitive data (such as records of state-level cigarette sale volume over the course of a decade). In all cases, however, $\boldsymbol{\pi}_0$ is known, and $\mathbf{w}$ could, for example, simply be chosen to minimize a target function such as $\left\|\boldsymbol{\pi}_0 - \sum_{j=1}^{J} w_j \cdot \boldsymbol{\pi}_j\right\|_2^2$. Under our motivating context, where $\boldsymbol{\pi}_0$, the joint covariate distribution of the index hospital, is unknown. Only $\boldsymbol{\pi}_{0|+}$ is known, and similar marginal distributions may not lead to similar cross-product ratios or similar joint distributions.

Given all values of $\mathbf{w}$ are possible, we assume $\mathbf{w}$ follows a Dirichlet distribution, which is a multivariate distribution characterized by the density function

$$f(\mathbf{w}; \alpha_1, \ldots, \alpha_J) = \frac{\Gamma(\prod_{j=1}^{J} \alpha_j)}{\prod_{j=1}^{J} \Gamma(\alpha_j)} \prod_{j=1}^{J} w_j^{\alpha_j - 1} \quad (7)$$

where $\Gamma(\cdot)$ is the Gamma function, and α_j are positive and finite values called concentration parameters. To gain full utility of this distribution in our motivating problem, we reparametrize the Dirichlet distribution with parameters $\beta_0, \beta_1, \ldots, \beta_J$, such that $\beta_j = \frac{\alpha_j}{\sum_{j=1}^{J} \alpha_j} \forall\, j \neq 0$ and $\beta_0 = \sum_{j=1}^{J} \alpha_j$. The Dirichlet distribution generates vectors of positive values which add up to one. On average, the generated value takes the form $\bar{\mathbf{W}} = \{\beta_j\}_{j \neq 0}$. As $\beta_0 \to \infty$, a generated $\mathbf{w}$ will always take the form $\bar{\mathbf{W}}$. As $\beta_0 \to 0$, generated $\mathbf{w}$ will tend to be vectors which contain a single value close to 1, with remaining values close to 0.

Since a higher $\beta_{j \neq 0}$ induces, on average, a higher realization of W_j, it could be said that $\beta_{j \neq 0}$ describes our confidence in the statement $\Theta_0 = \Theta_j$. Due to lack of any prior information about which hospitals are similar, we assume for this paper that all $\{\beta_j\}_{j \neq 0}$ are equal, which centers the distribution on the unweighted average of all reference hospitals. For example, we see that when all $\{\beta_j\}_{j \neq 0}$ are equal but β_0 is very small, the prevailing candidates for $\mathbf{w}$ will represent cases where the index hospital has some nearly exact mirror among the reference hospitals, though we do not know which is the mirror. Despite averaging out to the case where the index is the unweighted average of all references, this selection of Dirichlet parameters nearly precludes such a possibility. That is, the value of equation 7 decreases for $\mathbf{w} = \{\beta_j\}_{j \neq 0}$ when $\beta \to 0$, $\beta_1 = \ldots = \beta_J$.

5 Point Estimation and Inference of the Standardized Incidence Ratio

Traditional methods of computing a confidence interval for the SIR ($p/\tilde{\mathrm{E}}[p]$) presume $\tilde{\mathrm{E}}[p]$ to be a known constant, and calculate confidence intervals for p by assuming $n \cdot \hat{p} \sim \text{Poisson}(n \cdot p)$, where n is the sample size of the index hospital (Yule (1934)). A confidence interval for the SIR is then estimated by dividing the known $\tilde{\mathrm{E}}[p]$ from a confidence interval for p. Note that if $\tilde{\mathrm{E}}[p]$ is incorrectly observed, this practice would result in an interval that fails to include the true SIR at the selected confidence level.

In our motivating problem, $\tilde{\mathrm{E}}[p]$ is not observed at all, inducing uncertainty in both the numerator and the denominator of the SIR. We've set $\hat{\mathrm{E}}[p] = \tilde{\boldsymbol{p}}' \times \widetilde{\boldsymbol{\Pi}} \times \mathbf{w}$, and must take into account the uncertainty due to $\mathbf{w} \sim \text{Dirichlet}(\beta_0, \beta_1, \ldots, \beta_J)$. We propose an innovative simulation approach to take into account variability due to the estimation of $\mathbf{w}$.

Given known values for $\beta_0, \beta_1, \ldots, \beta_J$, $\mathbf{w}$ can be easily simulated. A simulated value of $\mathbf{w}$, along with a simulated value of p using $n \cdot \hat{p} \sim \text{Poisson}(n \cdot p)$, form a simulated value for the SIR. Multiple candidates for the SIR may be simulated this way. The mean of these values is taken as a point estimate for the SIR, and quantiles of these candidates reflecting the desired confidence level may be chosen to form a uncertainty interval for the SIR; call the desired confidence level γ' and let the chosen percentiles be $(1 - \gamma')/2$ and $1 - (1 - \gamma')/2$. However, though $\beta_1, \ldots, \beta_J$ are known (section 4), the value of β_0 still remains to be chosen.

Let p_j and $\hat{p}_j$ be the true and observed probability of the outcome Y in the j^{th} reference hospital. Let $\tilde{\mathrm{E}}[p_j]$ be the expected value of $\hat{p}_j$ given $\tilde{\boldsymbol{p}}$. Given a fixed β_0, we estimate the coverage rate of a uncertainty interval by having each reference hospital take the role of the index one by one. That is, for each $j \neq 0$, we rake $\{\boldsymbol{\pi}_k\}_{k \neq j}$ similarly to equation 5 to predict $\tilde{\mathrm{E}}[p_j]$ via the random variable

$$\hat{\mathrm{E}}[p_j] = \tilde{\boldsymbol{p}}' \times \widetilde{\boldsymbol{\Pi}}_{-j} \times \mathbf{w}_{-j} \quad (8)$$

where $\widetilde{\boldsymbol{\Pi}}_{-j}$ is a matrix whose columns are $(\boldsymbol{\pi}_k\}_{k \neq j}$, raked to match the marginals of $\boldsymbol{\pi}_j$, and $\mathbf{w}_{-j}$ follows a Dirichlet distribution with mean $\bar{\mathbf{W}}_{-j} = (\beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_J)$. An uncertainty interval can be thus computed, through simulation as described earlier, for the SIR of the j^{th} reference hospital ($p_j/\tilde{\mathrm{E}}[p_j]$). Call this interval (a_j, b_j).

Because each $(\boldsymbol{\pi}_j\}_{j \neq 0}$ is known, the true $\tilde{\mathrm{E}}[p_j]$, and thus the distribution of the SIR under circumstances of traditional indirect standardization, are also known. The expected coverage rate of (a_j, b_j) can be computed by

$$\gamma_j = \int_{\tilde{\mathrm{E}}[p_j]a_j}^{\tilde{\mathrm{E}}[p_j]b_j} f(x)dx \quad (9)$$

where $f(x)$ is some candidate posterior distribution for p given $n \cdot \hat{p} \sim$ Poisson($n \cdot p$). There are many possible candidates for $f(x)$, the one we chose was motivated by the distribution assumption that

$$np \sim \text{Gamma}(n\hat{p} + 1/3, 1) \quad (10)$$

The justification for this new distribution is the Bayesian prior assumption that $np \sim$ Gamma(1/3, 0), which has been shown to be non-informative in estimating the rate parameter of a Poisson random variable (Kerman (2011)). This, when motivated by data observed from $n\hat{p} \sim$ Poisson(np), produces the posterior equation 10. An "overall" coverage rate for the candidate β_0 may then be computed as $\gamma = \frac{1}{J}\sum_{j=1}^{J}\gamma_j$.

Here we run into the problem raised in section 4 where $\boldsymbol{\pi}_j$ may fall outside the convex hull of $\widetilde{\boldsymbol{\Pi}}_{-j}$. Observe that, by equation 8, $\hat{\mathrm{E}}[p_j]$ is bounded by the maximum and minimum of $\{\tilde{E}[p_k]\}_{k \neq j}$. When $\tilde{\mathrm{E}}[p_j]$ falls outside these bounds, equation 9 may not be able to produce a coverage rate of γ'. The values of $\{\gamma_k\}_{k \neq j}$ are usually able to compensate for this and still give $\gamma = \gamma'$, and in fact this was nearly always the case in the data example and simulations of sections 6 and 7. However, it is still theoretically possible $\gamma = \gamma'$ cannot be achieved for any β_0. In such cases, we follow the precedence of synthetic control methods (Abadie et al. (2010)) and attempt to minimize the target function $|\gamma - \gamma'|$.

To do this, we note that the covariance of $\mathbf{w} \sim$ Dirichlet($\alpha_1, \ldots, \alpha_J$) takes the following form (Devroye (1986)):

$$\text{Var}(w_j) = \frac{\alpha_j(\sum_{s=1}^{J}\alpha_s - \alpha_j)}{(\sum_{s=1}^{J}\alpha_s)^2(\sum_{s=1}^{J}\alpha_s + 1)}$$

$$\text{Cov}(w_j, w_k) = \frac{-\alpha_j\alpha_k}{(\sum_{s=1}^{J}\alpha_s)^2(\sum_{s=1}^{J}\alpha_s + 1)}$$

Under the reparametrization to $\beta_0, \beta_1, \ldots, \beta_J$, these equations change to

$$\text{Var}(w_j) = \frac{\beta_j(1-\beta_j)}{\beta_0 + 1} \qquad \text{Cov}(w_j, w_k) = \frac{-\beta_j\beta_k}{\beta_0 + 1}$$

meaning

$$\text{Var}(\hat{\mathrm{E}}[p]) = \left[\sum_{j=1}^{J}(\tilde{\boldsymbol{p}}' \times \tilde{\boldsymbol{\pi}}_j)^2\frac{\beta_j(1-\beta_j)}{\beta_0+1}\right] + \left[\sum_{j=1}^{J}\sum_{k \neq j}(\tilde{\boldsymbol{p}}' \times \tilde{\boldsymbol{\pi}}_j)(\tilde{\boldsymbol{p}}' \times \tilde{\boldsymbol{\pi}}_k)\frac{-\beta_j\beta_k}{\beta_0+1}\right] \quad (11)$$

Since all elements of $\tilde{\mathbf{w}}$, $\tilde{\boldsymbol{p}}$, and $\tilde{\boldsymbol{\Pi}}$ are by definition bounded by (0,1) and are independent of β_0, we see that equation 11 tends to 0 strictly as $\beta_0 \to \infty$. This observation can be easily expanded to $\hat{E}[p_j]$.

As $\mathrm{Var}(\hat{E}[p_j])$ has a one-to-one relationship with β_0, so must γ_j. Thus, we note that when all $\beta_0 > 0$ fail to induce $\gamma = \gamma'$, then $|\gamma - \gamma'|$ will be minimized as $\beta_0 \to 0$ or $\beta_0 \to \infty$.

6 Application to Example Data

We evaluated our proposed algorithm by comparing the 95% uncertainty intervals it produces with 95% confidence intervals produced in traditional indirect standardization, where the full joint covariate distribution of the index hospital is known. The goal is to show that intervals produced by our method cover roughly the same values as intervals produced with full information on the index hospital, and thus would have similar utility.

This application was performed using a dataset of nearly all (377,928) consecutive adult abdomen computed tomography scans from 151 hospitals performed between April 2015 and August 2016. Sample sizes of scans within hospitals varied between 33 and 17,930, with a mean of 2,503 and a standard deviation of 3,452. The hospitals sampled include public, private, academic, and non-academic institutions, from a variety of localities in Europe, Japan, and throughout the United States, representing very diverse demographics and radiological practices.

The outcome of interest is whether a scan had radiation dosage, measured by the dose length product in milligray-centimeters (mGy-cm), above a value of 1140 mGy-cm, which was predetermined to be high. The expected probability of high-dose scans in a hospital is 25%, with wide variability in values across hospitals, ranging from 2.5% to 82%. To evaluate this probability in an index hospital relative to others, we control for the patient case-mix including gender, age, and abdominal diameter, as well as whether the scan was single or multi-phase. Both age and abdominal diameter were categorized into quartiles.

These covariates significantly influence the outcome, producing highly variable elements in $\tilde{\boldsymbol{p}}$, which ranged from 0.4% for single-phase scans performed on small, very young, female patients, to 82% for multi-phase scans performed on very large, young, female patients. The probability vector of high dose exams given all covariates was computed directly (see section 2) due to availability of large sample size in the reference hospitals. The distributions of the covariates themselves also vary greatly across hospitals, which range from hospitals with exceptional obesity rates, to those who mostly treat children (and thus only have young people among the adult scans they do perform), to those which have good representation across all demographics.

We observed the joint covariate distribution for all hospitals, providing an opportunity to evaluate the performance of our algorithm had only marginal distributions been observed. To do this, we randomly selected 25% of hospitals (37) to serve the hypothetical role of index hospitals unable to obtain or unwilling to provide individual-level scan data. The remaining 75% (114) were treated as a referential “training set”. The training set was used to estimate β_0, resulting in a value of $\hat{\beta}_0 = 1.84$. Using this estimated β_0, 95% uncertainty intervals were

then formed for the SIRs of the hypothetical index hospitals using equation 6 and $\mathbf{w} \sim$ Dirichlet (1.84/114, …, 1.84/114). This interval was compared to the 95% SIR confidence intervals produced using traditional indirect standardization, which requires all $\boldsymbol{\pi}_j$ be known.

Figure 1 shows the SIR uncertainty intervals generated by our algorithm using only covariate marginal distributions in unfilled, black-bordered boxes and the SIR confidence intervals produced using traditional indirect standardization using full joint covariate distributions in grey-filled, unbordered boxes. On average, our algorithm's uncertainty intervals were only 6.2% wider than traditional confidence intervals, despite using significantly less information. However, at the extremes, our algorithm produced intervals up to 45% wider than their traditional counterparts. Comparatively wide intervals from our algorithm tend to occur in cases where the index hospital has an exceptionally large sample size or relatively high observed incidence of high dosage; in such cases, the widths of traditional intervals decrease substantially, while the widths using our methods do not.

The intervals estimated with our method also tend to cover the same values as traditional intervals. On average, our intervals shared 89.6% of the values they covered with their traditional counterparts, while the traditional intervals on average shared 94.3% of the values they covered with our counterpart.

7 Simulation Study

We conducted a simulation study to evaluate how well 95% uncertainty intervals estimated with our approach predict the true SIR value. We would like to construct uncerntainty intervals which contain the true SIR at least 95% of the time. Our method was compared to the case where the covariate cross-product ratios of the index hospital are non-random. Specifically, the joint covariate distribution of the index was computed by directly raking the joint covariate distribution of the overall reference dataset. This methodology, which we will call the "fixed denominator method," is computationally simple compared to the one we've proposed, but produces uncertainty intervals which fails to include the true SIR 95% of the time.

To perform these simulations, we assumed the covariate distributions and high dosage incidences (and therefore also the SIR) observed in our example hospitals are true values. From this, we simulated a new set of reference hospitals with the same covariate distributions and sample sizes as the example hospitals, but sampled their observed incidences of high dose exams from $\frac{1}{n_j}\text{Poisson}(n_j p_j)$.

The simulated hospitals were split into 37 index hospitals and 114 reference hospitals, in the same groups as they were in section 6. The algorithm was performed as before, except this time we simulated 1,000 unique datasets. For each dataset, we observed whether each simulated hospital had its true SIR contained within the uncertainty interval estimated by our method and by the fixed denominator method, producing for us a prediction rate for the SIR both overall and for each hospital. Our goal was to produce a prediction rate of 95%. The

phenomenon mentioned in section 4 of this not being theoretically possible due to $\boldsymbol{\pi}_j$ falling outside the convex hull of $\widetilde{\boldsymbol{\Pi}}_{-j}$ was not observed in our simulations.

Figure 2 shows the prediction rates of the true SIR in each of the 37 index hospitals. For our proposed method, these prediction rates averaged 95.1%, and ranged from 85.8% to 98.4%, with an interquartile range of 1.8% (94.5% to 96.3%). For the fixed denominator method, these prediction rates averaged 93.4%, and ranged from 80.3% to 96.0%. Prediction rates produced by the fixed denominator method were significantly and consistently inferior to those produced by our methods among the 37 index hospitals.

Figure 3 shows the biases of the SIR point estimates from each simulated hospital, using our proposed method. On average, the point estimates, with an overall average percent bias of 0.02% and average percent biases within hospitals ranging from −0.08% to 1.8%. The percent biases produced from individual simulated hospital datasets vary greatly at the extremes, with the most extreme as large as −267%, but in general they vary modestly, with an interquartile range of 9.3% (−4.5% to 4.8%). In terms of point estimates, the fixed denominator method had similar performance.

As figures 2 and 3 show, the algorithm performed poorest in terms of lower prediction rates and higher percent biases when the true SIR was low. Performance gradually improved as the true SIR approached 1, and performed very well for SIRs as high as 5.97 (the largest value observed in our example dataset).

8 Conclusion and Expansion

We demonstrated that a combination of raking and Dirichlet parametrization of synthetic control weights can be used to conduct indirect standardization for a hospital with only marginal covariate distributions available, when supplementary data on joint covariate distributions are provided from reference hospitals. An exploration with example data and a simulation study showed that our approach is unbiased, on average, and has uncertainty interval coverage rates consistent with expectations. On average, our algorithm’s uncertainty intervals were only slightly wider than traditional confidence intervals, despite using significantly less information.

Future work will include developing methods for selecting measures of similarity between the reference hospitals and the index hospital, with respect to associative structure of covariates. Such a measure can be quantified in a variety of ways depending on the outcome variable. For our motivating problem, hospital characteristics that may correlate to similar covariate cross-product ratios between hospitals may include geographic location, whether the hospital is public or private, whether the hospital is academic or non-academic, or certain numerical summaries of covariate marginal distributions. Sufficiently-informed estimates of these measures should improve SIR point estimates, as well as decrease the width of the uncertainty intervals.

Our methodology can also be extended to utilize continuous covariates, and there are extensions of raking which allow for this development (Deville et al. (1993)). The greatest

challenge of this extension would come from context surrounding the motivating medical problem, which involves asking a user who's likely not versed in statistics to provide marginal summaries for covariate distributions. The most informative marginal summary for a continuous covariate would be a density function, a concept not well-understood among our target audience. On the other hand, statistical concepts which are well-understood may be insufficiently informative to obtain a good estimate. Balancing these two objectives will be a pressing issue going forward, one whose solution may not lie exclusively in the world of statistics.

Our next step is to develop a web application where hospitals can enter their marginal covariate distributions and percentage of high dose exams to get an estimated SIR and uncertainty interval, given their case-mix. The hospitals included in our example will serve as the reference population. The web site will provide important feedback to hospitals on the quality of their performance in computed tomography radiation safety relative to a large number of reference hospitals.

The advantages of the index hospital only needing to provide marginal covariate distributions are numerous. Full joint covariate distributions are often infeasible to obtain due to difficulties in extracting data from electronic health records. Even when available, full joint distributions may be tedious to enter into a web application. As the number of (categorical) covariates being considered increases, the number of cells in a contingency table detailing the covariate distribution increases exponentially. Even among index hospitals who are able to obtain full joint covariate distributions, there may still exist those who wish to be profiled, but are apprehensive about providing individual-level data due to human subject protections. By asking only for marginal covariate distributions, we offer protection to patient confidentiality.

Our approach will be used to develop a web application which will be efficient to use while protecting patient confidentiality. The application will be entirely operated by the user, without the need to communicate with either a statistician or the institutions providing referential data. Our algorithm runs very quickly, taking on average three seconds to provide an uncertainty interval for the SIR, when 114 reference hospitals and four covariates (taking on 64 possible combinations) are considered, and when any information intrinsic only to the reference hospitals (such as the reference joint covariate distributions) are computed ahead of time. We hope the availability of such easily-accessible software will encourage more hospitals to compare their computed tomography scans to those of other hospitals, and in doing so lead to improved quality of service.

Code used to produce the results of this paper has been provided in the accompanying "RSCfiles.zip" file; the "README.txt" file within contains descriptions for this code.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by multi-site studies evaluating radiation exposure from computed tomography imaging and methods to reduce radiation exposure (PI: Rebecca Smith-Bindman, MD: NCI 1R01CA181191, PCORI CA-0054498). We thank the study investigators for their willingness to share their data for this work.

References

Abadie A, Diamond A, and Hainmueller J (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. Journal of the American Statistical Association 105, 493–505.

Ash A, Fienberg S, Louis T, Normand S, Stukel T, and Utts J (2011). Statistical issues in assessing hospital performance. Technical report, Committee of Presidents of Statistical Societies.

Bishop Y (1967). Multidimensional contingency tables: cell estimates (Ph.D. dissertation). Harvard University.

Christiansen C and Morris C (1997). Improving the statistical approach to health care provider profiling. Annals of Internal Medicine 127, 764–768. [PubMed: 9382395]

Deming W and Stephan F (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. The Annals of Mathematical Statistics 11, 427–444.

Deville J and Särndal C (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association 87, 376–382.

Deville J, Särndal C, and Sautory O (1993). Generalized raking procedures in survey sampling. Journal of the American Statistical Association 88, 1013–1020.

Devroye L (1986). Non-Uniform Random Variate Generation. Springer-Verlag.

Farr W (1859). Letter to the registrar-general on the causes of death in england in 1857. Twentieth Annual Report of the Registrar-General of Births, Deaths, and Marriages in England 5, 163–178.

Fienberg S (1970). An iterative procedure for estimation in contingency tables. The Annals of Mathematical Statistics 41, 907–917.

Ireland C and Kullback S (1968). Contingency tables with given marginals. Biometrika 55, 179–188. [PubMed: 5661043]

Kerman J (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions. Electronic Journal of Statistics 5, 1450–1470.

Mosteller F (1968). Association and estimation in contingency tables. Journal of the American Statistical Association 63, 1–28.

Rudolph K, Stuart E, Vernick J, and Webster D (2015). Association between connecticut's permit-to-purchase handgun law and homocides. American Journal of Public Health 105, 49–54. [PubMed: 25393182]

Ruschendorf L (1995). Convergence of the iterative proportional fitting procedure. The Annals of Statistics 23, 1160–1174.

Singh A and Rao J (1995). On the adjustment of gross flow estimates for classification error with application to data from the canadian labour force survey. Journal of the American Statistical Association 90, 478–488.

Smith-Bindman R, Moghadassi M, Wilson N, Nelson T, Boone J, Cagnon C, Gould R, Hall D, Krishnam M, Lamba R, McNitt-Gray M, Seibert A, and Miglioretti D (2015). Radiation doses in consecutive ct examinations from five university of california medical centers. Radiology 277, 134–141. [PubMed: 25988262]

Stephan F (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. The Annals of Mathematical Statistics 13, 166–178.

Woodard D, Gelfrand A, Barlow W, and Elmore J (2007). Performance assessment for radiologists interpreting screening mammography. Statistics in Medicine 26, 1532–1551. [PubMed: 16847870]

Yule G (1934). On some points relating to vital statistics, more especially statistics of occupational mortality. Journal of the Royal Statistical Society 97, 1–84.

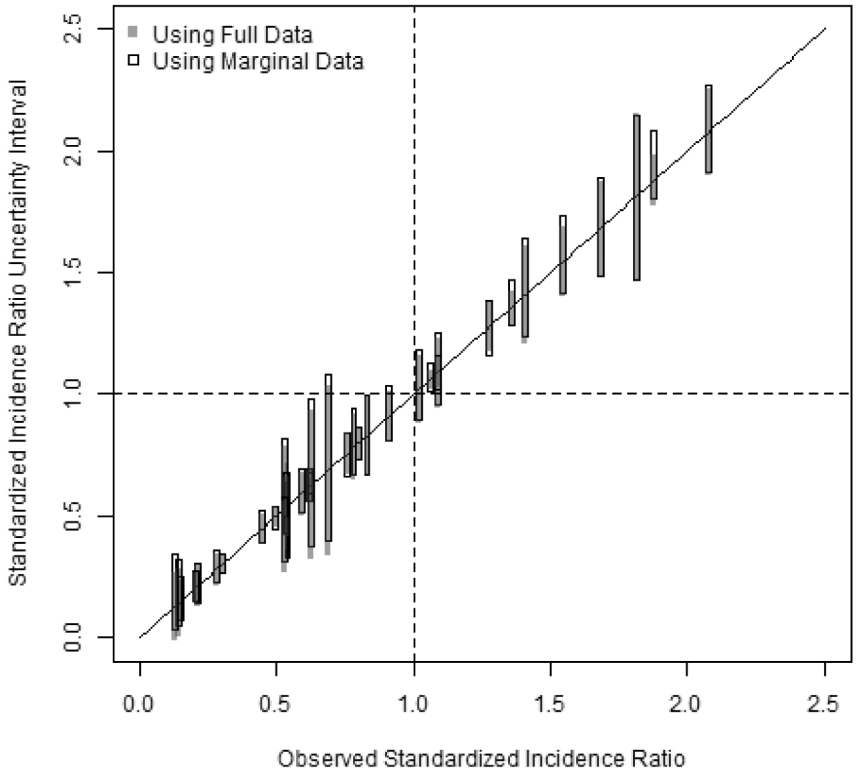

Figure 1:

Results of the Hospital Profiling example, evaluating the incidence of high-dose exams. The y-axis shows Standardized Incidence Ratio (SIR) uncertainty intervals, generated by our algorithm for 37 index hospitals using only covariate marginal distributions (unfilled boxes with black borders), compared to SIR confidence intervals produced using traditional indirect standardization using full joint covariate distributions (grey-filled boxes with no borders). The x-axis shows the observed SIR of the index hospital, given individual-level data. Two hospitals with relatively high SIRs were removed from the plot to improve visualization of the remaining hospitals.

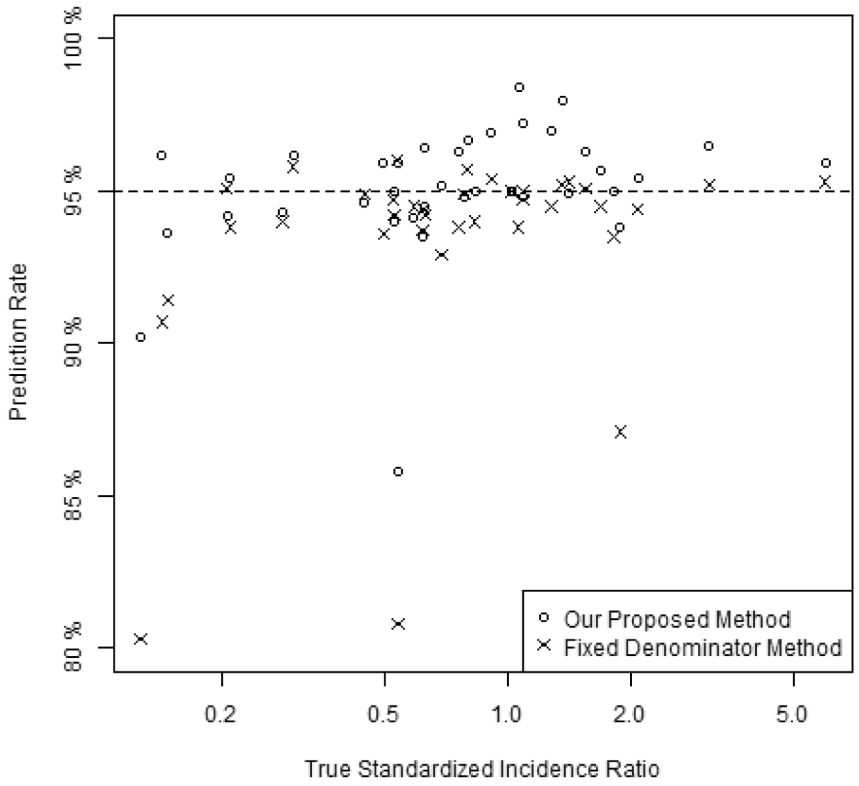

Figure 2:

Results of the simulation study. The y-axis shows the observed coverage rate of uncertainty intervals estimating the Standardized Incidence Ratio (SIR) of each of 37 index hospitals, formed using the 1,000 simulated datasets. The x-axis shows the true SIR of each hospital. Results for both our proposed method (circles) and the fixed denominator method (cross) are shown.

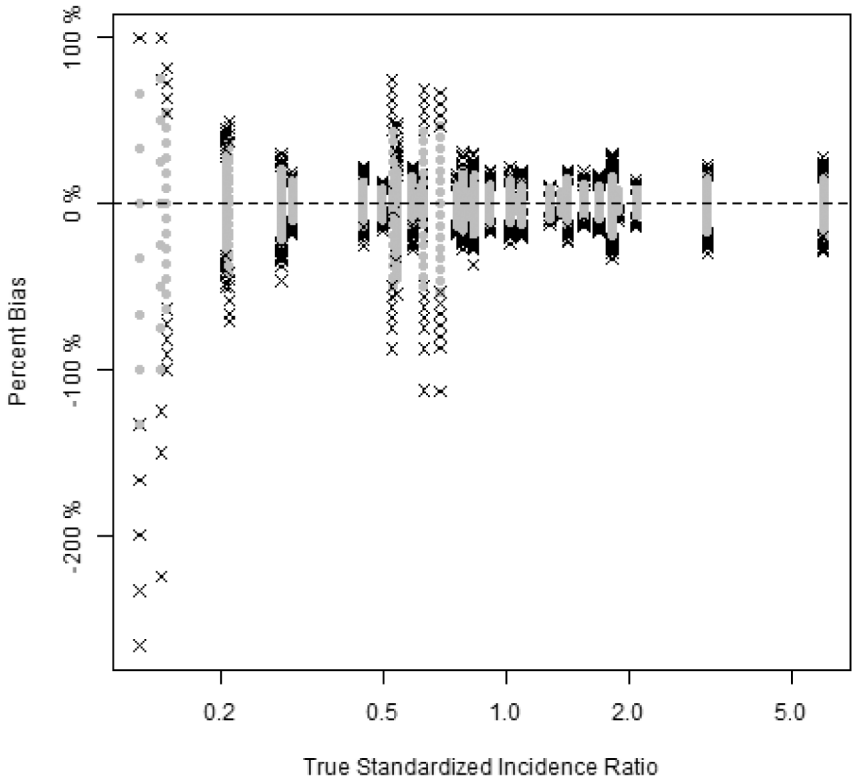

Figure 3:

Results of simulation studies. The y-axis shows the percent bias of the SIRs estimated by our algorithm for 37 simulated index hospitals using only covariate marginal distributions. The x-axis shows the true SIR values from which the datasets were simulated. Grey dots indicate simulated hospitals whose true SIRs were covered by the uncertainty interval estimated using our algorithm, black crosses indicated simulated hospitals whose true SIRs were not.