Hierarchical causal variance decomposition for institution and provider comparisons in healthcare

Bo Chen1 · Kristen McAlpine2 · Keith A. Lawson2 · Antonio Finelli2 · Olli Saarela1

Received: 25 August 2022 / Accepted: 26 November 2022 / Published online: 30 January 2023 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Disease-specific quality indicators are used to compare institutions and health care providers in terms of processes or outcomes relevant to treatment of a particular condition. In the context of surgical cancer treatments, the performance variations can be due to hospital and/or surgeon level differences, creating a hierarchical clustering. We consider how the observed variation in care received at patient level can be decomposed into that causally explained by the hospital performance, surgeon performance within hospital, patient case-mix, and unexplained (residual) variation. For this purpose, we derive a four-way variance decomposition, with particular attention to the causal interpretation of the components. For estimation, we use inputs from a mixed-effect model with nested random hospital/surgeon-specific effects, and a multinomial logistic model for the hospital/surgeon-specific patient populations. We investigate the performance of our methods in a simulation study and demonstrate them through analysis of administrative data on kidney cancer care in Ontario.

Keywords Causal inference · Nested random effects model · Quality indicator · Variance decomposition

1 Introduction
1.1 Background

Data on health care utilization and patient outcomes have multilevel structure, with clusters formed for example by administrative subregions, referral networks, hospitals, and physicians (Daniels and Gatsonis 1999). Quantifying the between-cluster variation in processes of care and outcomes can reveal quality of care related issues. Comparison of outcomes and practice patterns between institutions and physicians is referred to as hospital or provider profiling, with reviews of statistical approaches given for example by Goldstein and Spiegelhalter (1996); Shahian and Normand (2008); Racz and Sedransk (2010). Some

Olli Saarela
olli.saarela@utoronto.ca

1 Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
2 Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada
of the modeling approaches are aimed at identifying outlier clusters (Farrell et al. 2010),
while others focus on quantifying and explaining the sources of variation (Hawley et al.
2006). We will discuss these issues in the context of disease-specific quality indicators (QIs) for surgical care of kidney cancer (Wood et al. 2013; Lawson et al. 2017), where the clusters of interest are surgeons nested within hospitals. The QIs we consider are either process type, capturing variations in care delivered, or outcome type, capturing variation in patient outcomes (Donabedian 1988).

Although it is possible that some surgeons operate in multiple hospitals, a hierarchical clustering can be constructed by considering each hospital-surgeon combination as a separate category. To introduce some notation, let \( Y \in \mathbb{R} \) represent the observed process or outcome experienced by a given patient, used to construct a QI, and \( X = (X_1, \ldots, X_p) \) represent a vector of patient characteristics necessary for case-mix adjustment in the comparisons. Also, let \( Z \in \{1, \ldots, m\} \) indicate the hospital in which the patient was treated, and \( S \in \{1, \ldots, h\} \) the surgeon that operated on the patient in hospital \( z \). The treatment received/outcome of the patient can be modeled through a generalized linear model of the form

\[
E[Y \mid Z = z, S = s, X = x] = g^{-1}(a_0 + \alpha_z + \gamma_{zs} + \beta'x),
\]

where \( g \) is the link function. The model can be made identifiable by setting the fixed effects \( \alpha_0 = 0 \) and \( \gamma_{z1} = 0, \, z = 1, \ldots, m \), or alternatively through random effects by taking them to be IID as \( \alpha_c \sim N(0, \tau^2) \) and \( \gamma_{zs} \sim N(0, \kappa^2) \). The covariance between the hospital and surgeon effects is 0 due to the two-level categories being nested rather than crossed; this model is referred to as a nested random effects model (Norberg 1986; Longford 1987; Rabe-Hesketh et al. 2005). It also involves the additional assumption that the random effects are independent of the individual-level characteristics (Dieleman and Templin 2014; Clarke et al. 2015). The choice between the fixed effect and random effect formulations depends in part on the numbers of clusters at different levels and numbers of observations per cluster. With large number of clusters, some of these small, the random effect model can provide more stable estimation due to the shrinkage effect for the small clusters, which may be desirable even if the distributional assumption on the random effect is violated. Model (1) assumes the absence of interactions between the cluster effects and the individual-level characteristics, but this can be relaxed by allowing for the interactions, which can again be either fixed or random (Bell et al. 2019). While in some contexts models such as (1) are fitted to estimate the effects \( \beta \) of the individual-level characteristics, in the hospital/provider profiling context such a model is typically used to estimate the cluster effects while adjusting for the case-mix factors \( X \); predictions from the model can be used to calculate directly standardized estimates of the hospital/provider specific means. The problem of directly standardized comparisons between hospitals can be framed in a causal inference framework using potential outcomes, as outlined by Varewyck et al. (2014).

Another use for models such as (1) is to quantify how much variation in the outcome is explained by the cluster-level effects. In the present context this answers the question of whether quality of care differences exist in the health care system overall, in particular, adjusted for patient case-mix, whether similar kinds of patients receive different level of care. Demonstrating such variation is often the first step of validating a proposed process or outcome measure as a QI. In the case of identity link, or at the link function scale, an \( X \)-conditional variance decomposition can be directly given in terms of the random effect variance parameters (Merlo et al. 2006). More generally at the outcome scale, variance
decompositions can still be calculated making use of model-based predictions. However, generally there exists several alternative variance decompositions depending on the order of conditioning on the variables (Bowsher and Swain 2012). In Chen et al. (2020) we demonstrated that a certain ordering of the variables results in a variance decomposition where the between-hospital component can be given a causal interpretation. We also observed that the residual variance component was large for several quality indicators for surgical care of kidney cancer, after accounting for hospital-level effects and case-mix variation. This motivates us to further develop a causal variance decomposition and corresponding estimators for hierarchical clusters, for the purpose of quantifying the contribution of hospital and surgeon level effects while adjusting for patient case-mix. This requires introduction of potential outcomes notation for multiple levels of exposures, which can be adapted from instrumental variable (Angrist et al. 1996) and causal mediation analysis (VanderWeele and Vansteelandt 2009; VanderWeele et al. 2014) literature.

1.2 Plan of the paper

The objective motivated in the previous section is to develop a hierarchical causal variance decomposition for nested hospital and surgeon effects, outline an estimation method for it and investigate the properties of the method through simulation. In Sect. 2.1, we adapt potential outcomes notation to represent multiple nested exposure levels, and state the necessary assumptions for unconfounded comparisons between the levels. In Sect. 2.2, we generalize the three-way causal variance decomposition of Chen et al. (2020) to a four-way decomposition capturing variance components due to patient case-mix, hospitals’ performance, surgeons’ performance, and unexplained variation, and discuss its causal interpretation in Sect. 2.3. In Sect. 2.4, we connect the variance decomposition to measures of intra-class correlation. We propose an estimation method based on nested random-effect models in Sect. 3, study its properties in a simulation study in Sect. 4, and illustrate its use in a data analysis in Sect. 5. We end with a discussion on limitations and future research directions in Sect. 6.

2 Proposed measures

2.1 Notation and assumptions

For simplicity, we suppress the individual level index $i$ when it is not needed. To complement the notation for the observed variables introduced in Sect. 1.1, where $Y$, $X$, $Z$ and $S$ stand for the outcome, case-mix variables, hospital indicator and surgeon indicator, respectively, we let $Y(z,s) \in \mathbb{R}$ represent the potential/counterfactual outcomes of the same patient received care via surgeon $s \in \{1, \ldots, h_z\}$ operating in a hospital $z \in \{1, \ldots, m\}$. Further, let $S(z) \in \{1, \ldots, h_z\}$ indicate the surgeon that potentially operates the patient, if referred to hospital $z$. The observed variables $Y$ and $S$ are linked to their potential counterparts under the counterfactual consistency/stable unit treatment value assumption (SUTVA), by $Y = Y(Z, S(Z))$ and $S = S(Z)$. 
Causal inferences on the hospital and surgeon effects are possible under the assumption of strong ignorability of the joint hospital and surgeon assignment mechanism, which states that $0 < P(Z = z, S = z \mid X = x) < 1$ for all $z \in \{1, \ldots, m\}$, $s \in \{1, \ldots, h_z\}$ and $x$ (positivity) and $Y(z, s) \perp (Z, S) \mid X$ (conditional exchangeability) (Rubin and Rosenbaum 1983; Hernán and Robins 2006). We note that positivity is assumed only over the observed hospital/surgeon combinations, to make the levels nested rather than crossed. We introduce the shorthand notations $g(s; z, X) \equiv P(S = s \mid Z = z, X)$ and $e(z; X) = P(Z = z \mid X)$ for the corresponding assignment probabilities. The hypothesized causal relationships are illustrated in the directed acyclic graph (DAG) in Fig. 1. Here $I$ and $H$ represent additional factors that influence the hospital/surgeon assignment and the patient characteristics, but are not confounders. In particular, adjustment for the instrumental variables $I$ would likely lead to positivity violations.

We note that since $Y(z, S(z)) = Y(z)$, the above notation reduces to the hospital-level potential outcome notation used by, for example, Varewyck et al. (2014) and Chen et al. (2020). For this one-level clustering, in Chen et al. (2020) we derived a causal variance decomposition

\[
V(Y) = V_X \left\{ \sum_z E(Y(z) \mid X)e(z; X) \right\} \\
+ E_X \left\{ \sum_z E(Y(z) \mid X) - \sum_{z'} E(Y(z') \mid X)e(z'; X) \right\}^2 e(z; X) \\
+ E_X \left\{ \sum_z V(Y(z) \mid X)e(z; X) \right\} \\
= \text{variance explained by the patient case-mix} \\
+ \text{average variance causally explained by the between-hospital differences in performance conditional on case-mix} \\
+ \text{residual variance.}
\]

Fig. 1 Causal mechanism for hospital ($Z$) assignment, surgeon ($S$) assignment and a process of care ($Y$). $X$ represents a vector of potential confounders relevant to the case-mix adjustment, while $I$ represents instrumental variables that predict the hospital assignment but are not confounders. $H$ represents latent history that can influence $I$ and $X$ but is not in itself confounder.
Here the second variance component captures the average squared differences from the average level of care for similar patients between the hospitals. However, it does not capture between-provider variation within the hospitals, which is included in the residual variance. In the following we derive a four-way decomposition that introduces a new term to capture the within-hospital between-provider variation.

### 2.2 Four-way decomposition for observed variation in care received

In this section, we derive a connection between the observed variance $V[Y]$ and causally interpretable variance components expressed in terms of potential outcome variables. Readers less interested in the details of the derivation may want to skip directly to result (7) and its interpretation. Under counterfactual consistency/SUTVA, we have $V[Y] = V[Y(Z, S(Z))]$. We begin with the two-way variance decomposition

$$V[Y(Z, S(Z))] = V_X[E(Y(Z, S(Z)) \mid X)] + E_X[V(Y(Z, S(Z)) \mid X)].$$

(3)

In Equation (3), the first term can be further written as

$$V_X[E(Y(Z, S(Z)) \mid X)] = V_X[E_{Z\mid X}[E(Y(Z, S(Z)) \mid Z, X)]]$$

$$= V_X[E_{Z\mid X}[E_{S(Z)\mid Z,X}[E(Y(Z, S(Z)) \mid S(Z), Z, X)]]]$$

(4)

In the latter term, we can further write

$$E_X[V(Y(Z, S(Z)) \mid X)]$$

$$= E_X[V_{Z\mid X}[E(Y(Z, S(Z)) \mid Z, X)]]$$

$$+ E_X[E_{Z\mid X}[V(Y(Z, S(Z)) \mid Z, X)]]$$

$$= E_X[V_{Z\mid X}[E_{S(Z)\mid Z,X}[E(Y(Z, S(Z)) \mid S(Z), Z, X)]]]$$

$$+ E_X[E_{Z\mid X}[V_{S(Z)\mid Z,X}[E(Y(Z, S(Z)) \mid S(Z), Z, X)]]]$$

(5)

Substituting Equations (4) and (5) into Equation (3), we obtain

$$V[Y(Z, S(Z))] = V_X[E_{Z\mid X}[E_{S(Z)\mid Z,X}[E(Y(Z, S(Z)) \mid S(Z), Z, X)]]]$$

$$+ E_X[V_{Z\mid X}[E_{S(Z)\mid Z,X}[E(Y(Z, S(Z)) \mid S(Z), Z, X)]]]$$

$$+ E_X[E_{Z\mid X}[V_{S(Z)\mid Z,X}[E(Y(Z, S(Z)) \mid S(Z), Z, X)]]]$$

$$+ E_X[E_{Z\mid X}[E_{S(Z)\mid Z,X}[V(Y(Z, S(Z)) \mid S(Z), Z, X)]]].$$

(6)

Due to the strong ignorability assumption, we have

$$E(Y(z, s) \mid S = s, Z = z, X) = E(Y(z, s) \mid X)$$

and

$$V(Y(z, s) \mid S = s, Z = z, X) = V(Y(z, s) \mid X).$$

Hence, we can obtain the following result in terms of the potential outcomes:
The interpretation is

\[
V[Y] = V_X\left\{ \sum_{z} \sum_{s} E(Y(z, s) \mid X)g(s; z, X)e(z; X) \right\}
\]
\[
+ E_X\left\{ \sum_{z} \left[ \sum_{s} E(Y(z, s) \mid X)g(s; z, X) \right] \right\}
\]
\[
- \sum_{z'} \sum_{s} E(Y(z', s) \mid X)g(s; z', X)e(z'; X) \right)^2 e(z; X) \right\}
\]
\[
+ E_X\left\{ \sum_{z} \left[ \sum_{s} \left( E(Y(z, s) \mid X) \right) \right] \right\}
\]
\[
- \sum_{s'} E(Y(z, s') \mid X)g(s'; z, X) \right)^2 g(s; z, X) \right] e(z; X) \right\}
\]
\[
+ E_X\left\{ \sum_{z} \sum_{s} V(Y(z, s) \mid X)g(s; z, X)e(z; X) \right\}. \tag{7}
\]

The interpretation is

\[
V[Y] = \text{total observed variance in care received}
\]
\[
= \text{variance explained by the patient case-mix}
\]
\[
+ \text{average variance causally explained by the between-hospital differences in performance conditional on case-mix}
\]
\[
+ \text{average variance causally explained by the surgeon performance conditioning on patient case-mix and hospital performance}
\]
\[
+ \text{unexplained (residual) variance.}
\]

We note that the first and second term in (7) are equivalent to the first and second term in (2). This also implies that the third, residual, variance component in (2) is equivalent to the sum of the third and fourth components in (7), meaning that the additional term in (7) is a result of splitting the residual variance in (2). We will consider the second and the third terms as the causal quantities of interest. These have a causal interpretation, as further discussed in Sect. 2.3, and can be linked to the observable quantities under the causal assumptions, working backwards from (7). In the second term, the hospital performance is compared to the average level across all the hospitals for a patient with characteristics \(X\) and then averaged over the patient population. In the third term, the surgeon performance is compared to the average level of the surgeons in the same hospital for a patient with characteristics \(X\), and then averaged over the hospitals and the patient population.

There are two possible approaches to estimate the variance components in (7). In the first approach, directly based on the factorization in (7), we can estimate them based on modeling \(E[Y \mid S, Z, X]\), \(P(S \mid Z, X)\), \(P(Z \mid X)\) and using the empirical distribution of \(X\). Because the four-way variance decomposition can be also expressed as

\[
V[Y] = V_X[\mathbb{E}(Y \mid X)]
\]
\[
+ E_{Z,X}\left\{ [\mathbb{E}(Y \mid Z, X) - \mathbb{E}(Y \mid X)]^2 \right\}
\]
\[
+ E_{S,Z,X}\left\{ [\mathbb{E}(Y \mid S, Z, X) - \mathbb{E}(Y \mid Z, X)]^2 \right\}
\]
\[
+ E_{S,Z,X}[V(Y \mid S, Z, X)], \tag{8}
\]

\(\copyright\) Springer
the variance components can be also estimated based on modeling $E[Y \mid S(Z), Z, X]$, $E[Y \mid Z, X]$, $E[Y \mid X]$ and using the empirical distribution of $(S, Z, X)$. We will discuss both approaches in Sect. 3; the former is based on factorization of the likelihood and can be used to construct an approximate Bayesian inference procedure.

### 2.3 Causal interpretation of the decomposition

To better understand the causal interpretation of the variance decomposition, we considered a special case with two hospitals with two surgeons each. The interpretation of the case-mix and between-hospital components is unchanged and was already discussed by Chen et al. (2020), so we focus on the interpretation of the within-hospital between-surgeon component. With two hospitals with indexed by $z = 1, 2$, and two surgeons in each indexed by $s = 1, 2$, we denote the hospital assignment probability with $P(Z = 1 \mid X) = e(1; X) = e(X)$, with $P(Z = 2 \mid X) = 1 - e(X)$, and the surgeon assignment probability within hospital with $P(S = 1 \mid Z = z, X) = g(1; z, X) = g(z, X)$, with $P(S = 2 \mid Z = z, X) = 1 - g(z, X)$. Now the third term in (7) becomes

$$
E_{X}\left\{ e(X)g(1, X)(1 - g(1, X))[E(Y(1, 1) \mid X) - E(Y(1, 2) \mid X)]^2 \\
+ (1 - e(X))g(2, X)(1 - g(2, X))[E(Y(2, 1) \mid X) - E(Y(2, 2) \mid X)]^2 \right\} 
= E_{X}\left\{ e(X)V(S \mid Z = 1, X)[E(Y(1, 1) \mid X) - E(Y(1, 2) \mid X)]^2 \\
+ (1 - e(X))V(S \mid Z = 2, X)[E(Y(2, 1) \mid X) - E(Y(2, 2) \mid X)]^2 \right\}\tag{9}$$

We consider three scenarios to illustrate the causal interpretation of (9) via the relationship of $X, Z, S$ and $Y$ in Fig. 1.

**Scenario 1.** In the absence of the arrow $X \rightarrow Y$ (the case-mix variables are not prognostic of the outcome and thus are not confounders), which implies $Y \perp X \mid (Z, S)$ and $E(Y(z, s) \mid X) = E(Y(z, s))$, (9) becomes

$$
[E(Y(1, 1)) - E(Y(1, 2))]^2 E_{X}[e(X)V(S \mid Z = 1, X)] \\
+ [E(Y(2, 1)) - E(Y(2, 2))]^2 E_{X}[(1 - e(X))V(S \mid Z = 2, X)].
$$

The first multiplicative terms represent squared pairwise causal contrasts. This is multiplied by the second terms, the magnitude of which depends on the volume of patients of type $X$ in each hospital, as well as the variation in the surgeon assignment for patients of type $X$. The latter is maximized when both surgeons treat similar patient populations. If the surgeons specialize on treatment of specific kinds of patients, so that there is no overlap in the patient populations treated by the two surgeons, the positivity assumption is violated and the corresponding component is equal to 0. Thus, for between-surgeon performance differences to manifest through this variance component, there must be some overlap in the patient population they treat.
Scenario 2. In the absence of the arrows $X \rightarrow S$, $X \rightarrow Z$, $I \rightarrow S$, $I \rightarrow Z$ (all hospitals and surgeons treat similar patient population, essentially randomized assignment), which implies $(Z, S) \perp X$, we have $e(X) = e$, $g(1, X) = g(1)$, and $g(2, X) = g(2)$. Now (9) becomes

$$
e V(S \mid Z = 1)E_X \left\{ \left[ E(Y(1, 1) \mid X) - E(Y(1, 2) \mid X) \right]^2 \right\}$$

$$+ (1 - e)V(S \mid Z = 2)E_X \left\{ \left[ E(Y(2, 1) \mid X) - E(Y(2, 2) \mid X) \right]^2 \right\}.$$  

We note that under this completely randomized setting the magnitude of the causal effects are proportional to the terms

$$E_X \left\{ \left[ E(Y(z, 1) \mid X) - E(Y(z, 2) \mid X) \right]^2 \right\}$$

$$= E_X \left\{ E(Y(z, 1) - Y(z, 2) \mid X) \right\}^2$$

$$= V_X[E(Y(z, 1) - Y(z, 2) \mid X)]$$

that is, proportional to the sum of the variance of the covariate conditional causal effects and the squared population average causal effect. The former captures effect modification by the patient characteristics, showing that any effect modification adds to the measure, rather than canceling out, and the latter captures the overall performance difference. The resulting variance component, as expressed for two levels being compared, has similarities to the causal interpretation recently given to the model reliance metric used to measure variable importance in machine learning contexts (Fisher et al. 2019). Our results show that a similar kind of effect measure can be derived through a variance decomposition argument under a randomized assignment, and generalize this from two to multiple exposure levels being compared. Since the variance component under a randomized assignment mechanism may be of interest in itself as a causal quantity, in Sect. 2.4 we show that it can always be defined and estimated under a hypothetical target assignment mechanism of interest regardless of the actual mechanism that assigns patients for hospitals and surgeons.

Scenario 3. In the absence of the arrow $S \rightarrow Y$ (no surgeon effect on the outcome), which implies $Y \perp S \mid (Z, X)$, we have $E(Y(1, 1) \mid X) = E(Y(1, 2) \mid X)$ and $E(Y(2, 1) \mid X) = E(Y(2, 2) \mid X)$, and the between-surgeon component is zero, as it should it the absence of individual-level causal effects.

2.4 Stochastic interventions

The variance decomposition (7) was derived for the observed marginal variance of the outcome, which is why it depends on the mechanism that assigns patients to hospitals and surgeons, including the hospital and surgeon volume. Alternatively, we can derive a variance decomposition under a hypothetical “randomized” assignment, where for example each hospital/surgeon treats similar kind of patient population, and/or similar patient volume. This resembles the stochastic interventions considered by Díaz and van der Laan (2012). Compared to a deterministic intervention where the level of treatment is set to a fixed value for the entire population, in a stochastic intervention it is given a
chosen probability distribution. This will also allow us to derive a connection between the causal variance decomposition and well-known intra-class correlation measures.

Let \( A \in \{1, \ldots, m\} \) and \( B(A) \in s \in \{1, \ldots, h_a\} \) be hospital and surgeon assignments randomly drawn with specified probabilities \( \tilde{e}(a; X) = P(A = a \mid X) \) and \( \tilde{g}(b; a, X) = P(B = b | A = a, X) \), chosen such that \( (A, B) \perp (Z, S) \mid X \), \( 0 < P(A = a, B = b \mid X = x) < 1 \) and \( Y(a, b) \perp (A, B) \mid X \). Here choosing for example \( \tilde{e}(a; X) = P(Z = a) \) and \( \tilde{g}(b; a, X) = P(S = b | Z = a) \) would correspond to a mechanism where each hospital/surgeon treats similar patient population, but retaining the original patient volumes. Choosing \( \tilde{e}(a; X) \equiv 1/m \) and \( \tilde{g}(b; a, X) = 1/h_a \) would also mean setting the volumes to be the same. While such interventions will not be implemented in reality, these will allow us to formulate more “pure” causal quantities of interest that remove some of the non-causal variation from the decomposition, as explained in Scenario 2 in the previous section.

For the variance under the hypothetical assignment mechanism we get the decomposition

\[
V[Y(A, B(A))] = V_X \left\{ \sum_a \sum_b E(Y(a, b) \mid X) \tilde{g}(b; z, X) \tilde{e}(a; X) \right\} \\
+ E_X \left\{ \sum_a \sum_b \left[ \sum_b E(Y(a, b) \mid X) \tilde{g}(b; a, X) \right]^2 \tilde{e}(a; X) \right\} \\
- \sum_{a'} \sum_b E(Y(a', b) \mid X) \tilde{g}(b; a', X) \tilde{e}(a'; X) \right\}^2 \tilde{e}(a; X) \right\} \\
+ E_X \left\{ \sum_a \sum_b V(Y(a, b) \mid X) \tilde{g}(b; a, X) \tilde{e}(a; X) \right\}.
\]

Because under the above assumptions, we have

\[
E[Y(a, b) \mid A = a, B = b, X] = E[Y(a, b) \mid X] = E[Y(a, b) \mid Z = a, S = b, X] = E[Y \mid Z = a, S = b, X],
\]

and because \( \tilde{e}(a; X) \) and \( \tilde{g}(b; a, X) \) are fixed quantities, this is also estimable from the observed data on \( (Y, Z, S, X) \).

Under the special case of \( \tilde{e}(a; X) \equiv 1/m \) and \( \tilde{g}(b; a, X) = 1/h_a \) and the linear mixed-effect model

\[
E[Y(a, b) \mid Z = a, S = b, X] = E[Y(a, b) \mid X] = \alpha_0 + \alpha_a + \gamma_{ab} + \beta'X,
\]

where the nested hospital and surgeon random effects are IID as \( \alpha_a \sim N(0, \tau^2) \) and \( \gamma_{ab} \sim N(0, \kappa^2) \) and residuals distributed as \( Y - E[Y \mid Z, S, X] \sim N(0, \sigma^2) \), the second term in (10) can be written as
\[
\frac{1}{m} \sum_a \left\{ \left( \alpha_a - \frac{1}{m} \sum_{a'} \alpha_{a'} \right) + \left( \frac{1}{h_a} \sum_b \gamma_{ab} - \frac{1}{m} \sum_{a'} \frac{1}{h_{a'}} \sum_b \gamma_{a'b} \right) \right\}^2.
\]

Keeping \( m \) fixed and letting \( h_a \to \infty \) for all \( a \in \{1, \ldots, m\} \), the terms \( \frac{1}{h_a} \sum_b \gamma_{ab} \) converge to \( E(\gamma_{ab}) = 0 \), and the second term in becomes (10)

\[
\frac{1}{m} \sum_a \left\{ \alpha_a - \frac{1}{m} \sum_{a'} \alpha_{a'} \right\}^2.
\]

Letting \( m \to \infty \), (11) in turn converges to \( V(\alpha_a) = \tau^2 \), the variance of the between-hospital effects. Hence, for this variance component we obtain the same result as for the three-way causal variance decomposition proposed by Chen et al. (2020).

The third term in (10) can be written as

\[
\frac{1}{m} \sum_a \frac{1}{h_a} \sum_b \left( \alpha_{ab} - \frac{1}{h_a} \sum_{b'} \gamma_{ab'} \right)^2.
\]

Keeping \( m \) fixed and letting \( h_a \to \infty \) for all \( a \in \{1, \ldots, m\} \), (12) converges to \( V(\alpha_{ab}) = \kappa^2 \), the variance of the within-hospital between-surgeon effects. Therefore, under this special case, the X-conditional causal variance decomposition is

\[
V(Y(A, B(A)) \mid X) = V_{A|X}[E_{B(A)|A,X}[E(Y(A, B(A)) \mid B(A), A, X)]
\]

\[
+ E_{A|X}[V_{B(A)|A,X}[E(Y(A, B(A)) \mid B(A), A, X)]
\]

\[
+ E_{A|X}[E_{B(A)|A,X}[V(Y(A, B(A)) \mid B(A), A, X)]
\]

\[
\to \tau^2 + \kappa^2 + \sigma^2 \quad \text{when} \quad h_a \to \infty \ \forall a \in \{1, \ldots, m\} \ \text{and} \ \ m \to \infty,
\]

that is, asymptotically equivalent to the variance decomposition obtained through the model-based random effect and residual variances. Here \((\tau^2 + \kappa^2)/(\tau^2 + \kappa^2 + \sigma^2)\) would correspond to the within-hospital within-surgeon intra-class correlation coefficient.

Under the stochastic intervention, we can also see that the causal variance decomposition is closely connected to quantities considered in direct standardization. Taking again \( \tilde{e}(a; X) = 1/m \) and \( \tilde{g}(b; a, X) = 1/h_a \), the third term of (10) can be written as

\[
\frac{1}{m} \sum_a \left\{ \frac{1}{h_a} \sum_b E_X \left[ E(Y(a, b) \mid X) - \frac{1}{h_a} \sum_{b'} E(Y(a, b') \mid X) \right] \right\}^2,
\]

which shows that the between-surgeon variance component is an average of squared differences of the surgeon performance from the average performance in the hospital, taken over all the hospitals, surgeons and the overall patient population. In contrast, ranking of surgeons within a hospital could be based on directly standardized quantities \( E_X[E(Y(a, b) \mid X)] \), \( b = 1, \ldots, h_a \), while outlier detection could be based on the differences \( E_X[E(Y(a, b) \mid X)] - \frac{1}{h_a} \sum_{b'} E_X[E(Y(a, b') \mid X)] \) from the average performance in hospital \( a \).
3 Estimators

3.1 Point estimation

We outline estimators based on the decomposition (7), which requires fitting hospital, surgeon and case-mix conditional outcome model and case-mix conditional assignment model. The same estimation approach works also for decomposition (9), but substituting fixed target assignment/stochastic intervention probabilities in place of the observed ones.

To model the outcomes, we use a generalized linear mixed model

\[
E[Y(z,s) \mid X; \theta] = E[Y \mid Z = z, S = s, X; \theta] = g^{-1} (\alpha_0 + \alpha_z + \gamma_z + \beta X). \tag{13}
\]

where \( \theta = (\alpha_0, \alpha_z, \gamma_z, \beta) \) and where the nested hospital and surgeon random effects are taken to be IID as \( \alpha_z \sim N(0, \tau^2) \) and \( \gamma_z \sim N(0, \kappa^2) \). We use the mixed model for the purpose of stabilizing the estimation of the hospital/surgeon effects in the presence of small volume hospitals and surgeons, that is, it is considered as an approximation, rather than a data generating mechanism. When considering a fixed number of hospitals and surgeons in a given health care system, such as in our application in Sect. 5, there is no sampling of hospitals and surgeons from a superpopulation, and it makes more sense to consider a fixed effect model as the underlying data generating mechanism, which we will do in our simulation study in Sect. 4.

For the joint hospital/surgeon assignment mechanism, we fit a multinomial logistic regression model

\[
P(Z = z, S = s \mid X; \eta) = \begin{cases} 
1 + \sum_{a=1}^{m_a} \exp(\psi_{ab} + \phi_{ab} X) + \sum_{b=1}^{m_b} \exp(\psi_{ab} + \phi_{ab} X) + \sum_{a=1}^{m_a} \sum_{b=1}^{m_b} \exp(\psi_{ab} + \phi_{ab} X) 
& \text{if } z = 1, s = 1 \\
\exp(\psi_{ab} + \phi_{ab} X) 
& \text{if } z \neq 1, s = 1 \\
1 + \sum_{a=1}^{m_a} \exp(\psi_{ab} + \phi_{ab} X) + \sum_{b=1}^{m_b} \exp(\psi_{ab} + \phi_{ab} X) + \sum_{a=1}^{m_a} \sum_{b=1}^{m_b} \exp(\psi_{ab} + \phi_{ab} X) 
& \text{if } z = 1, s \neq 1 \\
\exp(\psi_{ab} + \phi_{ab} X) 
& \text{if } z \neq 1, s \neq 1 
\end{cases} \tag{14}
\]

where \( \eta = \{(\psi_{ab}, \phi_{ab}) : (a, b) \neq (1, 1)\} \). Hence, we can obtain

\[
e(z; X; \eta) = P(Z = z \mid X; \eta) = \sum_{s=1}^{h_z} P(Z = z, S = s \mid X; \eta)
\]

and

\[
g(s; z, X; \eta) = P(S = s \mid Z = z, X; \eta) = \frac{P(Z = z, S = s \mid X; \eta)}{P(Z = z \mid X; \eta)}.
\]
Alternatively, a multinomial logistic assignment model can be first fitted at hospital level to estimate the quantities $e(z; X, \eta)$, after which surgeon level multinomial assignment models are fitted conditionally on each hospital to estimate $g(s; z, X, \eta)$.

We denote the fitted values for the expected outcomes $\mu_s(z, s; \theta) = E(Y_i | Z_i = z, S_i = s, x_i; \theta)$. Under the mixed-effects model these are obtained by using empirical Bayes prediction for the random hospital and surgeon-level intercepts (e.g., Skrondal and Rabe-Hesketh, 2004, Chapter 7). Further, we denote $V[Y; \theta, \eta] = \omega_1(\theta, \eta) + \omega_2(\theta, \eta) + \omega_3(\theta, \eta) + \omega_4(\theta, \eta)$ for the four terms in the parameterized version of the variance decomposition (7). The first (case-mix) component can now be estimated by

$$\omega_1(\hat{\theta}, \hat{\eta}) = \frac{1}{n - 1} \sum_{i=1}^{n} \left\{ \sum_z \sum_s \mu_i(z, s; \hat{\theta}) g(s; z, x_i, \hat{\eta}) e(z; x_i, \hat{\eta}) \right\}^2.$$  

The second (between-hospital) component can be estimated by

$$\omega_2(\hat{\theta}, \hat{\eta}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_z \left[ \sum_s \mu_i(z, s; \hat{\theta}) g(s; z, x_i, \hat{\eta}) \right]^2 e(z; x_i, \hat{\eta}) \right\} - \left[ \sum_z \sum_s \mu_i(z, s; \hat{\theta}) g(s; z, x_i, \hat{\eta}) e(z; x_i, \hat{\eta}) \right]^2 \right\}.$$  

The second (between-surgeon) component can be estimated by

$$\omega_3(\hat{\theta}, \hat{\eta}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_z \left[ \sum_s \mu_i(z, s; \hat{\theta})^2 g(s; z, x_i, \hat{\eta}) \right]^2 \right\} - \left[ \sum_s \mu_i(z, s; \hat{\theta}) g(s; z, x_i, \hat{\eta}) \right]^2 e(z; x_i, \hat{\eta}) \right\}.$$  

The fourth (residual) variance can be estimated by subtracting the sum of the above three components from the empirical marginal variance, or alternatively, based on the distributional assumption in the outcome model. In particular, for a binary outcome we have $V(Y_i | Z_i = z, S_i = s, x_i; \theta) = \mu_i(z, s; \theta)[1 - \mu_i(z, s; \theta)]$, and the residual variance component is given by

$$\omega_4(\hat{\theta}, \hat{\eta}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_z \sum_s \mu_i(z, s; \hat{\theta}) [1 - \mu_i(z, s; \hat{\theta})] g(s; z, x_i, \hat{\eta}) e(z; x_i, \hat{\eta}) \right\}.$$  

If the parameters can be estimated consistently such that $\hat{\theta} \overset{P}{\rightarrow} \theta$ and $\hat{\eta} \overset{P}{\rightarrow} \eta$, the component estimators $\omega_j(\hat{\theta}, \hat{\eta})$, $j \in \{1, 2, 3, 4\}$, are also consistent by the continuous mapping theorem and applying the law of large numbers for the sample averages over the empirical covariate distribution. We will briefly investigate their asymptotic normality through simulation in Sect. 4, but the variance estimation approach we propose in 3.2 does not make use of asymptotic normality.

Alternatively to the above model-based estimators, as noted in Sect. 2.2, a semi-parametric estimation procedure is suggested by decomposition (8), based on three different
outcome models. The first model is the fully conditional model (13). In addition, we need to specify and fit an outcome model marginal over the surgeon indices, such as

\[ E[Y \mid Z = z, X; \theta^*] = g^{-1}(\alpha_0^* + \alpha_z^* + \beta^*X), \]

where the hospital effects \( \alpha_z^* \) can again be specified either as fixed or random effects. Further, we need to specify a model marginal over both the hospital and surgeon indices, such as

\[ E[Y \mid X; \theta^+] = g^{-1}\left(\alpha_0^* + \beta^*X\right). \]

Fitting these models avoids the need to model the assignment mechanism. The semi-parametric approach is applicable to the decomposition of the empirical marginal variance, while the model-based approach can also be used to estimate decomposition under hypothetical assignment mechanisms. We will briefly compare the two approaches in the simulation study of Sect. 4.

### 3.2 Variance estimation

The point estimators suggested in Sect. 3.1, parametrized in terms of \( \theta \) and \( \eta \), are entirely model-based, conditional on the empirical covariate distribution. Because the model components correspond to the factorization of the likelihood, we can evaluate the uncertainty in the variance component estimates via approximate Bayesian inference. This is based on drawing samples from the joint posterior distribution of the parameters \( \theta \) and \( \eta \), given by

\[
f(\theta, \eta \mid Y, Z, S, X) = \frac{f(Y, Z, S \mid X, \theta, \eta)f(\theta \mid X)f(\eta \mid X)}{f(Y \mid Z, S, X)} = \frac{f(Y \mid \theta, Z, S, X)f(\theta \mid X)f(Z, S \mid \eta, X)f(\eta \mid X)}{f(Y \mid Z, S, X)} = \frac{f(\theta \mid Y, Z, S, X)f(\eta \mid Z, S, X)}{f(Y \mid Z, S, X)}.
\]

Posterior samples for the variance components can be obtained by sampling \( \theta \) and \( \eta \) from their posterior distributions, and recalculating \( \omega_1(\theta, \eta), \omega_2(\theta, \eta), \omega_3(\theta, \eta) \) and \( \omega_4(\theta, \eta) \) for each draw. The posteriors of \( \theta \) and \( \eta \) could be obtained via Markov chain Monte Carlo, but for computational convenience, here we use an approximate Bayesian procedure. The outline procedure allows us to use normal approximations for the posteriors of the model parameters, but avoid having to use normal approximation directly on the distributions of the variance components, which are likely to be more non-normal in small samples. For the outcome model parameters \( \theta \), we approximated the posterior using the parametric bootstrap, by resampling outcomes from the fitted model, refitting the model and calculating new fitted values \( \mu_i(z,s; \theta) \). For the assignment model parameters, we used the normal approximation to sample the \( \eta \) from a multivariate normal distribution \( MVN(\hat{\eta}, V(\hat{\eta})) \), where \( \hat{\eta} \) is the maximal likelihood estimator and \( V(\hat{\eta}) \) is the asymptotic variance-covariance matrix from the fitted multinomial logistic regression.
4 Simulation study

4.1 Generating mechanism

We used simulation to study the properties of the methods proposed in Sect. 3. The objectives for the simulation study were to (a) study the asymptotic properties (consistency, asymptotic normality) of the proposed point estimators, (b) to demonstrate that the new four-way decomposition is consistent with our previously proposed three-way decomposition, and (c) to compare the model-based estimators to the alternative semiparametric decomposition to verify that both are estimating the same quantity. We used a data-generating mechanism similar to Fig. 1, omitting the variables $I$ and $H$ for simplicity. The asymptotic behavior of the estimators was studied by varying the total number of hospitals $m$, the total number of surgeons $q$ (split equally between the hospitals, so that $h_1 = \ldots = h_m = q/m$), and the total number of patients $n$. We generated two patient case-mix factors, $X_1 \sim N(0, 1)$ and $X_2 \sim \text{Bernoulli}(0.5)$. The hospital ($Z$) and surgeon ($S$) assignments were generated based on multinomial logistic model (14), where the intercepts were generated from $N(0, 0.25)$ and coefficients from $N(0, 0.5)$. Outcomes were generated from the mean structure

$$E[Y(z, s) \mid X] = \alpha_z + \gamma_{zs} + X_1 + 2X_2$$

where the hospital’s effects $\alpha_z$ and surgeon effects $\gamma_{zs}$ were generated independently from $N(0, 2)$. The continuous outcomes were generated by taking $Y(z, s) = E[Y(z, s) \mid X] + \epsilon$, where $\epsilon \sim \text{Logistic}(0, 1)$, and the binary outcomes by dichotomizing these as $1_{[Y(z,s) \geq 0]}$. The observed outcomes were taken to be $Y = Y(Z, S)$. For estimation, we fitted mixed effect logistic models with nested random effects as in (13) and multinomial assignment models as in (14). The resulting estimates for the variance components were compared to the true values calculated under the above specified parameter values.

4.2 Results

The bars in Fig. 2 show the simulated sampling distribution means for the three variance components for the binary outcomes under different combinations of $n$ (total number of patients), $m$ (total number of hospitals), and $q$ (total number of surgeons), based on 1000 replications. The 95% quantile interval for the sampling distribution is represented by the black error bar. The 95% confidence interval for the mean is represented by the blue error bar; this reflects the Monte Carlo error in the estimated mean of the sampling distribution. The red dots indicate the true values of the variance components. From the results, we can observe that the between-hospital and case-mix components are well estimated under all scenarios. Accurate estimation of the within-hospital between-surgeon component requires sufficient surgeon-specific patient volumes, which are the largest under the scenario $n = 5000$, $m = 5$ and $q = 25$, explaining the more precise estimate for this variance component therein.

Figure 3 shows density plots for the simulated sampling distributions of the three variance components with varying $n$ and fixed $m$ and $q$. These are fairly normal-shaped, and demonstrate decreasing variability with increasing number of patients. With the two sample sizes, the hospital and patient variance components are already quite centered around the true values (reflected by the vertical red lines), while the surgeon component estimates appear closer to the true value with $n = 5000$ (200 per surgeon). As noted in
Sect. 3.1, we expect the proposed estimators to be consistent, but as the variance component estimators are complicated non-linear functions of the model parameters, they are not necessarily unbiased in small samples; the top right panel of Fig. 3 reflects this.

The gray bars in Fig. 4a show the simulated sampling distribution means for the estimated components of the three-way decomposition (2), and the white bars show the corresponding components obtained through the four-way decomposition (7) by adding up the third (between-surgeon) and fourth (residual) variance components. The estimates are similar, demonstrating that the new between-surgeon variance component is part of the residual variance in the three-way decomposition.

The gray bars in Fig. 4b show the simulated sampling distribution means for the three variance components using the model-based formulation (7) and the white bars the alternative semi-parametric formulation (8). The point estimates, as well as their

![Fig. 2](image)

**Fig. 2** Simulated sampling distribution means for the three variance components (without residual variance) under the random-effect model for the binary outcomes under different combinations of $n$ (total number of patients), $m$ (total number of hospitals), and $q$ (total number of surgeons), based on 1000 replications. The red dots indicate the true variances. The 95% quantile interval of the sampling distribution is represented by the black error bar. The 95% confidence interval for the mean is represented by the blue error bar, reflecting the Monte Carlo error in the estimated mean of the sampling distribution.
variability are similar under both approaches, demonstrating that both are appropriate for the point estimation.

5 Illustration in real data

In Chen et al. (2020) we analyzed hospital-level variation in the quality of surgical care of kidney cancer in Ontario, Canada. Using our previously proposed three-way causal variance decomposition, we analyzed four proposed quality indicators; 1) the proportion of partial (versus radical) nephrectomies performed for patients with pT1a renal masses (Partial), 2) the proportion of partial (versus radical) nephrectomies performed for patients with pT1a renal masses who have chronic kidney disease (CKD) or at least one of its risk factors (hypertension, diabetes) (Partial-CKD), 3) the proportion of nephrectomies performed using a minimally-invasive (versus open) approach for patients with a pT1-2 renal mass (MIS), and 4) readmission proportion within 30 days of nephrectomy for patients with a pT1-4 renal mass (Readmission). The first three of these are process type, while readmission is an outcome. While we found significant between-hospital and case-mix variation in several of the indicators, the residual variances were also large, motivating the question of how much between-surgeon variation is captured by these indicators. We will use the proposed four-way decomposition to assess this, while also including two additional outcome type quality indicators; 5) hospital length of stay following surgery (LOS), and 6)

![Density plots for the simulated sampling distributions of the case-mix, between-hospital, and between-surgeon variance components with a binary outcome, based on 1000 replications. The vertical red lines correspond to the true values of the variance components](image-url)
proportion with complications within 30 days of nephrectomy for patients with a T1-T4 renal mass (Comps). High proportions of partial and minimally invasive surgeries for early stage patients, as well as low proportions of unplanned readmissions, complications and shorter average length of stay for all surgical patients are considered to reflect good surgical quality of care. However, the variance decomposition analysis can reflect how much performance related variation can be captured by these measures. The data to evaluate the six disease-specific QIs come from population-level Ontario datasets housed at the Institute for Clinical Evaluative Sciences (ICES). The kidney cancer dataset comprises more than 10,000 nephrectomy patients in Ontario identified from the hospital Discharge Abstract Database (DAD) from 1995-2014, cross-linked to kidney cancer diagnoses in the Ontario Cancer Registry and corresponding Ontario Health Insurance Plan (OHIP) billings, as well as with data on tumor size, stage, histology, and grade available from abstracted pathology reports, needed for case-mix adjustment.

To estimate the variance decomposition, we first fitted multinomial logistic assignment models in two steps, first at hospital level and then surgeon level conditional on hospital. We used the R package VGAM (Yee et al. 2015) to fit the model to deal with empty covariate categories for small volume hospitals and surgeons. We only estimated an intercept term for the hospitals that had treated greater than 35 (40 for MIS) patients over the study period, and determined the cutoff for surgeons based on the criterion of the full-ranked design matrix in the model fit. For the processes/outcomes, we fitted generalized linear mixed models (linear for LOS, logistic for the others) with nested random intercepts for
hospitals and surgeons using the glmer function of R package lme4 (Bates et al. 2015). The models were adjusted for patient case-mix, including age, sex, income quintile, Charlson comorbidity score, ACG comorbidity score, days from diagnosis to treatment, year of diagnosis, tumor size, presence of CKD risk factors (for indicators other than Partial-CKD), and T stage (for indicators other than Partial and Partial-CKD).

Table 1 represents a summary of the effects of case-mix factors for the processes/outcomes. The covariate effects are reported as log-odds ratios (z-scores) for the binary variables (Partial, Partial-CKD, MIS, Readmission, and Comps) and regression coefficients (t-values) for the one continuous variable (LOS). In the table, n, m, and q represent the number of patients, hospitals, and surgeons included for each indicator, respectively. Further, χ₁², χ₂², and χ₃² represent likelihood ratio test statistics for the within-hospital between-surgeon effects (one degree of freedom), between-hospital effects (one degree of freedom), and aggregated between-surgeon and between-hospital effects (two degrees of freedom), respectively. The likelihood ratio tests show that there are statistically significant surgeon effects on all processes/outcomes at the significance level of 0.05, except for Readmission. The surgeon level variation is strongest for MIS. Of the case-mix factors, tumor size predicts strongly the process type measures due as it is related to the indication behind the treatment choice. The associations with year of diagnosis reveals that partial and minimally invasive nephrectomies have become more common over time, while average length of stay has decreased. Charlson comorbidity score predicts both process and outcome type measures.

Table 2 shows the four-way causal variance decomposition for the observed variance in the processes/outcomes, along with a 95% credible interval for each component via the approximate Bayesian method proposed in Sect. 3.2. The table reports both absolute and proportional variances, the latter being more comparable across the indicators. We note that the summation of the four variance components is equal to the total/marginal variance via a binomial calculation for the binary variables. Overall the three outcome type measures capture relatively little performance related variation, compared to the processes. Of the latter, MIS reveals substantial within-hospital between-surgeon variation, which may be explained by surgeons specializing on either minimally invasive or open technique. What is notable is that this variation exists for similar kinds of patients, and is not explained by different indication, as discussed on the interpretation of the variance components in Sect. 2.3. Thus, the variance decomposition analysis reveals a source of performance related variation, where patients might benefits from quality improvement.

For comparison, in Table 3 we also show the causal variance decomposition under a hypothetical randomized assignment mechanism, discussed in Sect. 2.4, where each hospital and surgeon treats a similar patient population, but using their actual marginal patient volumes. We note that the results are very similar to Table 2, though for most of the indicators the case-mix variance is now slightly lower, which is in line with the interpretation given to the case-mix term in Chen et al. (2020), according to which this term captures the indirect effect of the pathway $X \rightarrow (Z, S) \rightarrow Y$ (the effect of different hospitals/surgeons treating different kinds of patient populations), which is broken under the hypothetical assignment mechanism which removes the connection $X \rightarrow (A, B)$. For completeness, we have included also the results for the corresponding three-way decomposition (Table 4).
Table 1  A summary table of the effects of case-mix factors for the six different QIs. The numbers are log-odds ratios (z-scores) for the binary QIs (Partial, Partial-CKD, MIS, Readmission, and Comps) and linear estimated effects (t-values) for the continuous QI (Length of Stay), while n, m and s represent the number of patients, hospitals, and surgeons, respectively. \( \chi^2_1 \), \( \chi^2_2 \), and \( \chi^2_3 \) represent likelihood ratio test for the between-surgeon effects (one degree of freedom), between-hospital effects (one degree of freedom), and aggregated between-surgeon and between-hospital effects (two degrees of freedom), respectively. dx stands for diagnosis and tx for treatment (surgery).

| Process/outcome | Partial | Partial-CKD | MIS | Readmission | Comps | Length of stay |
|-----------------|---------|-------------|-----|-------------|-------|---------------|
| Covariate       |         |             |     |             |       |               |
| n = 4738        | n = 3028| n = 4100    | n = 7362 | n = 7479 | n = 7292 |
| m = 112         | m = 110 | m = 72      | m = 115 | m = 116 | m = 115   |
| s = 295         | s = 282 | s = 232     | s = 314 | s = 315 | s = 314   |
| \( \chi^2_1 \) = 161 | \( \chi^2_2 \) = 71 | \( \chi^2_3 \) = 600 | \( \chi^2_1 \) = 3 | \( \chi^2_1 \) = 31 | \( \chi^2_1 \) = 126 |
| \( \chi^2_2 \) = 377 | \( \chi^2_2 \) = 216 | \( \chi^2_3 \) = 825 | \( \chi^2_2 \) = 16 | \( \chi^2_2 \) = 29 | \( \chi^2_2 \) = 225 |
| \( \chi^2_3 \) = 538 | \( \chi^2_3 \) = 288 | \( \chi^2_3 \) = 1425 | \( \chi^2_3 \) = 20 | \( \chi^2_3 \) = 60 | \( \chi^2_3 \) = 351 |
| Male sex        | 0.21 (2.72) | 0.23 (2.35) | 0.07 (0.79) | -0.02 (−0.17) | 0.08 (1.21) | -0.02 (−2.18) |
| Age             | -0.01 (−3.95) | -0.01 (−2.71) | -0.01 (−1.34) | 0.01 (2.74) | 0.02 (6.57) | 0.01 (14.81) |
| Income quintile | 0.03 (1.03) | 0.04 (1.25) | 0.01 (0.44) | <0.01 (−0.03) | -0.02 (−0.96) | -0.02 (−4.47) |
| Rural vs urban  | 0.17 (1.34) | 0.17 (1.07) | 0.03 (0.22) | <0.01 (−0.13) | -0.03 (−0.30) | 0.01 (0.89) |
| Charlson score  | -0.15 (−4.70) | -0.16 (−4.68) | -0.14 (−4.82) | 0.13 (6.58) | 0.14 (10.03) | 0.04 (15.71) |
| ACG score       | <0.01 (−1.24) | <0.01 (−0.41) | <0.01 (−0.14) | 0.02 (4.71) | 0.01 (2.07) | <0.01 (2.51) |
| log(dx to tx days+1) | 0.01 (0.41) | 0.02 (0.90) | 0.02 (0.56) | 0.09 (3.72) | 0.03 (1.65) | 0.02 (6.99) |
| Year of dx      | 0.19 (19.24) | 0.19 (15.80) | 0.18 (10.80) | -0.01 (−1.34) | -0.04 (−5.59) | -0.03 (−23.31) |
| Tumor size (cm) | -0.89 (−17.32) | -0.82 (−13.13) | -0.18 (−6.91) | 0.02 (1.37) | 0.08 (7.10) | 0.02 (9.23) |
| CKD risk factors| 0.03 (0.35) | - | -0.01 (−0.07) | 0.15 (1.31) | 0.06 (0.80) | 0.02 (1.44) |
| T2 vs T1 stage  | - | - | -0.44 (−2.71) | -0.03 (−0.18) | -0.09 (−0.83) | 0.05 (2.66) |
| T3 or 4 vs T1 stage | - | - | - | 0.19 (1.65) | 0.46 (5.67) | 0.11 (7.54) |
| Source of variation | Process/outcome | Patient case-mix | Between-hospital | Between-surgeon | Residual |
|---------------------|----------------|-----------------|-----------------|----------------|---------|
| Empirical proportion| Variance       | 0.062           | 0.019           | 0.005          | 0.161   |
| Total/marginal variance| 95% CI       | (0.057, 0.068)  | (0.015, 0.022)  | (0.003, 0.007) | (0.155, 0.167) |
| Partial | Partial 0.447 | 0.447 × (1 − 0.447) = 0.247 | 25.1% | 7.6% | 2.0% | 65.3% |
|         | 95% CI        | (23.0%, 27.5%)  | (6.2%, 9.1%)    | (1.3%, 2.8%)   | (62.7%, 67.7%) |
| Partial-CKD | Variance       | 0.057           | 0.017           | 0.003          | 0.170   |
|            | 95% CI        | (0.050, 0.064)  | (0.013, 0.021)  | (0.001, 0.005) | (0.162, 0.177) |
| MIS | MIS 0.438 | 0.438 × (1 − 0.438) = 0.246 | 23.1% | 6.8% | 1.2% | 68.9% |
|            | 95% CI        | (20.5%, 25.8%)  | (5.2%, 8.7%)    | (0.6%, 2.1%)   | (65.9%, 72.1%) |
| Readmission | Variance       | 0.021           | 0.054           | 0.027          | 0.129   |
|            | 95% CI        | (0.017, 0.024)  | (0.049, 0.060)  | (0.023, 0.031) | (0.123, 0.135) |
| Comps | Comps 0.637 | 0.637 × (1 − 0.637) = 0.231 | 8.9% | 23.5% | 11.6% | 56.0% |
|            | 95% CI        | (7.5%, 10.6%)   | (21.3%, 25.8%)  | (9.8%, 13.4%)  | (53.1%, 58.5%) |
| Length of Stay | Variance       | 0.0019          | 0.0001          | < 0.0001       | 0.0679  |
|            | 95% CI        | (0.0012, 0.0026) | (< 0.0001, 0.0004) | (< 0.0001, 0.0001) | (0.0671, 0.0686) |
| Readmission | Proportion     | 2.69%           | 0.16%           | 0.01%          | 97.14%  |
|            | 95% CI        | (1.78%, 3.76%)  | (< 0.01%, 0.56%) | (< 0.01%, 0.15%) | (95.98%, 98.04%) |
| Comps | Variance       | 0.0146          | 0.0007          | 0.0002         | 0.1527  |
|            | 95% CI        | (0.0124, 0.0171) | (0.0002, 0.0013) | (< 0.0001, 0.0007) | (0.1501, 0.1550) |
| Length of Stay | Proportion     | 8.70%           | 0.41%           | 0.15%          | 90.74%  |
|            | 95% CI        | (7.36%, 10.15%) | (0.13%, 0.79%)  | (0.01%, 0.40%)  | (89.18%, 92.13%) |
| – | Variance       | 0.056           | 0.010           | 0.002          | 0.188   |
|            | 95% CI        | (0.051, 0.061)  | (0.008, 0.012)  | (0.001, 0.003)  | (0.182, 0.193) |
| 0.256 | Proportion     | 21.8%           | 4.0%            | 0.8%           | 73.4%   |
|            | 95% CI        | (20.0%, 23.7%)  | (3.2%, 4.8%)    | (0.5%, 1.2%)   | (71.3%, 75.4%) |

95% CI 95% credible interval
6 Discussion

In this manuscript, we have proposed causal variance decompositions for nested comparisons of hospital/physician quality of care. We considered two different decompositions, one for the observed variance of the quality of care outcome, and another for the variance under a hypothetical intervention on the hospital/physician assignment, the latter aiming for a more “pure” causal interpretation. We also considered two different model-based approaches for estimation, one requiring modeling the outcome and hospital/physician assignment, and another modeling the outcome marginally and conditionally. The methods in the present paper are aimed at helping to assess the usefulness of a given process or outcome in constructing a disease-specific quality indicator for identifying performance related between-hospital and between-surgeon variation. Establishing statistically significant between-provider variation while accounting for patient case-mix would support using the QI for performance outlier detection or ranking of providers using direct or indirect standardization methods. Alternatively, small between-provider variation relative to case-mix variation could suggest redefining the QI in a narrower patient population.

Although here we focused on two levels of hierarchical clustering due to the motivating application, the proposed four-way variance decomposition proposed here could be generalized to arbitrary number of levels of hierarchical clustering, by introducing further conditioning variables. The additional variance components will come out of the residual variation for lower level clusters introduced, as we observed for the surgeons within hospitals. While in the present context there are no more lower level clusters to introduce, we could introduce higher level clusters such as the Local Health Integration Networks (LHINs) which are health administrative subregions in Ontario.

For estimation of the variance decompositions, we used nested random effect models, as these can easily accommodate large number of small categories without identifiability problems that would be present with corresponding fixed effect models. For simplicity, we also omitted hospital-case-mix and surgeon-case-mix interaction terms from the models; however, in principle these can be easily incorporated, either through fixed or random effects, as the interpretation of the variance decompositions is separate from the parametrization of the hospital and surgeon effects. In fact, the formulation of the causal variance decompositions does not dictate what kind of models are used to estimate the predictive means/probabilities needed for calculation of the decomposition. Instead of parametric models, predictions derived through machine learning algorithms might be useful as well, especially as similar types of measures have been suggested for measuring variable importance in machine learning (Fisher et al. 2019). In the context of multlcategotry categorical exposures, the components in the causal variance decomposition can be seen as a way to concisely summarize a large number of pairwise causal contrasts. In principle the same approach could also be extended to other types of exposures, including continuous and function valued exposures, which in one further research direction we are pursuing.

We borrowed nested potential outcome notation from causal mediation analysis and instrumental variable estimation literature to represent the nested exposure levels. However, although the path hospital \rightarrow surgeon \rightarrow outcome in the causal diagram 1 resembles mediation, we note that the current problem is not a mediation problem, due to the surgeons being nested within the hospitals by definition in our analysis. Because of this, the surgeon effect is separate from the hospital effect, rather than a component of it,
### Table 3
A four-way variance decomposition for the potential outcome variance in the different QIs, assuming all the hospitals/surgeons are treating similar patient populations, but using the actual marginal patient volumes.

| Process/outcome | Source of variation | Total | Patient case-mix | Between-hospital | Between-surgeon | Residual |
|-----------------|---------------------|-------|------------------|------------------|-----------------|---------|
| Partial         | Variance            | 0.246 | 0.060            | 0.019            | 0.006           | 0.161   |
|                 | 95% CI              | (0.245, 0.248) | (0.055, 0.066) | (0.016, 0.023) | (0.004, 0.009) | (0.155, 0.166) |
| Proportion      | 100%                | 24.5% | 7.7%             | 2.6%             | 65.2%           |         |
|                 | 95% CI              | (22.2%, 26.8%) | (6.3%, 9.1%) | (1.8%, 3.6%) | (62.8%, 67.5%) |         |
| Partial-CKD     | Variance            | 0.245 | 0.056            | 0.017            | 0.004           | 0.168   |
|                 | 95% CI              | (0.243, 0.247) | (0.050, 0.062) | (0.013, 0.021) | (0.002, 0.006) | (0.162, 0.175) |
| Proportion      | 100%                | 22.8% | 6.9%             | 1.6%             | 68.7%           |         |
|                 | 95% CI              | (20.3%, 25.4%) | (5.2%, 8.7%) | (0.7%, 2.6%) | (65.9%, 71.4%) |         |
| MIS             | Variance            | 0.232 | 0.018            | 0.046            | 0.035           | 0.133   |
|                 | 95% CI              | (0.229, 0.235) | (0.015, 0.021) | (0.042, 0.051) | (0.031, 0.040) | (0.127, 0.140) |
| Proportion      | 100%                | 7.7%  | 20.0%            | 15.1%            | 57.2%           |         |
|                 | 95% CI              | (6.4%, 9.1%) | (18.2%, 21.8%) | (13.1%, 17.2%) | (54.7%, 59.8%) |         |
| Readmission     | Variance            | 0.0673 | 0.0019          | 0.0001           | < 0.0001        | 0.0653  |
|                 | 95% CI              | (0.0623, 0.0726) | (0.0013, 0.0027) | (< 0.0001, 0.0004) | (< 0.0001, 0.0001) | (0.0605, 0.0702) |
| Proportion      | 100%                | 2.79% | 0.17%            | 0.01%            | 97.03%          |         |
|                 | 95% CI              | (1.93%, 3.83%) | (< 0.01%, 0.61%) | (< 0.01%, 0.22%) | (95.93%, 97.88%) |         |
| Comps           | Variance            | 0.1660 | 0.0140          | 0.0006           | 0.0003          | 0.1511  |
|                 | 95% CI              | (0.1606, 0.1708) | (0.0116, 0.0164) | (0.0002, 0.0012) | (< 0.0001, 0.0009) | (0.1455, 0.1559) |
| Proportion      | 100%                | 8.43% | 0.36%            | 0.19%            | 91.02%          |         |
|                 | 95% CI              | (7.07%, 9.85%) | (0.12%, 0.75%) | (0.02%, 0.52%) | (89.49%, 92.37%) |         |
| Length of Stay  | Variance            | 0.244 | 0.053            | 0.010            | 0.005           | 0.178   |
|                 | 95% CI              | (0.237, 0.252) | (0.049, 0.058) | (0.008, 0.012) | (0.002, 0.004) | (0.172, 0.184) |
| Proportion      | 100%                | 22.0% | 4.1%             | 1.1%             | 72.8%           |         |
|                 | 95% CI              | (20.3%, 23.5%) | (3.3%, 5.0%) | (0.6%, 1.7%) | (71.1%, 74.5%) |         |

95% CI 95% credible interval
and thus not a mediator on the causal pathway from the hospital to the outcomes. In Daignault et al. (2019) we proposed methodology for mediation analysis in the quality of care context, with the aim of quantifying how much of between-hospital differences in an outcome type measure could be accounted for by of between-hospital differences in a process type measure considered as a mediator, using the hospital → MIS → LOS pathway as an example. The mediation analysis approach could be used to validate process type QIs by linking them causally to relevant outcomes. In Chen et al. (2022) we approached this problem from the variance decomposition point of view, decomposing the causal between-hospital variance component into direct and indirect effects, which themselves have a variance interpretation. We envision that this approach can accommodate the present hierarchical exposures, with a separate mediation analysis decomposition for the between-hospital and between-surgeon variances; this is also a direction for future research.

### Table 4 A three-way variance decomposition for the potential outcome variance in the different QIs. The hospital assignment probabilities were calculated based on the actual patient volume of the hospital

| Source of variation | Process/outcome | Total | Patient case-mix | Between-hospital | Residual |
|---------------------|-----------------|-------|------------------|------------------|----------|
|                     |                 | 0.246 | 0.058            | 0.018            | 0.170    |
|                     | 95% CI          | (0.245, 0.248) | (0.052, 0.063) | (0.015, 0.022) | (0.165, 0.175) |
|                     | Proportion      | 100%  | 23.5%            | 7.5%             | 69.0%    |
|                     | 95% CI          | –     | (21.3%, 25.5%)   | (6.1%, 8.9%)     | (66.9%, 71.0%) |
| Partial-CKD         |                 | 0.245 | 0.053            | 0.016            | 0.176    |
|                     | 95% CI          | (0.243, 0.247) | (0.046, 0.060) | (0.012, 0.021) | (0.170, 0.183) |
|                     | Proportion      | 100%  | 21.4%            | 6.7%             | 71.9%    |
|                     | 95% CI          | –     | (19.0%, 24.2%)   | (4.9%, 8.5%)     | (69.1%, 74.6%) |
| MIS                 |                 | 0.232 | 0.022            | 0.045            | 0.165    |
|                     | 95% CI          | (0.228, 0.235) | (0.018, 0.026) | (0.040, 0.050) | (0.160, 0.171) |
|                     | Proportion      | 100%  | 9.4%             | 19.3%            | 71.3%    |
|                     | 95% CI          | –     | (7.9%, 11.1%)    | (17.2%, 21.4%)   | 69.0%, 73.8% |
| Readmission         |                 | 0.0686| 0.0019           | 0.0001           | 0.0666   |
|                     | 95% CI          | (0.0636, 0.0733) | (0.0012, 0.0027) | < 0.0001, 0.0004 | (0.0618, 0.0714) |
|                     | Proportion      | 100%  | 2.75%            | 0.17%            | 97.08%   |
|                     | 95% CI          | –     | (1.89%, 3.83%)   | < 0.01%, 0.58%   | (95.93%, 97.92%) |
| Comps               |                 | 0.1674| 0.0139           | 0.0005           | 0.1530   |
|                     | 95% CI          | (0.1623, 0.1719) | (0.0114, 0.0163) | (0.0001, 0.0011) | (0.1482, 0.1580) |
|                     | Proportion      | 100%  | 8.29%            | 0.33%            | 91.38%   |
|                     | 95% CI          | –     | (6.88%, 9.62%)   | (0.07%, 0.67%)   | (90.04%, 92.75%) |
| Length of Stay      |                 | 0.252 | 0.055            | 0.010            | 0.187    |
|                     | 95% CI          | (0.245, 0.260) | (0.051, 0.060) | (0.008, 0.012) | (0.181, 0.193) |
|                     | Proportion      | 100%  | 21.9%            | 3.9%             | 74.2%    |
|                     | 95% CI          | –     | (20.5%, 23.5%)   | (3.1%, 4.8%)     | (72.5%, 75.8%) |

95% CI 95% credible interval
Acknowledgements This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (to OS), a Catalyst Grant in Health Services and Economics Research from the Canadian Institutes of Health Research (to AF, KAL and OS) and the Ontario Institute for Cancer Research through funding provided by the Government of Ontario (to BC). This study contracted ICES Data & Analytic Services (DAS) and used de-identified data from the ICES Data Repository, which is managed by ICES with support from its funders and partners: Canada’s Strategy for Patient-Oriented Research (SPOR), the Ontario SPOR Support Unit, the Canadian Institutes of Health Research and the Government of Ontario. The opinions, results and conclusions reported are those of the authors. No endorsement by ICES or any of its funders or partners is intended or should be inferred. Parts of this material are based on data and information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed herein are those of the author, and not necessarily those of CIHI. Parts of this material are based on data and information provided by Cancer Care Ontario (CCO). The opinions, results, view, and conclusions reported in this paper are those of the authors and do not necessarily reflect those of CCO. No endorsement by CCO is intended or should be inferred.

Author contributions The methodological research problem was devised by OS; the methods were formulated and implemented by BC, who wrote the first draft of the manuscript. KM, KAL and AF devised the applied research problem, and provided input into the creation of the analysis dataset, data analysis and interpretation of the results. All authors reviewed the manuscript.

Data availability The data that support the findings of this study are available from ICES (https://www.ices.on.ca/). Restrictions apply to the access to these data, which were used under agreement for this study.

Code availability R code to reproduce the simulation study will be made available at the Journal website.

Declarations

Conflict of interest The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval Our study was approved by the Research Ethics Board of the University Health Network, Toronto, Ontario.

Consent for publication Please see https://www.ices.on.ca/Data-and-Privacy/ICES-data/Working-with-ICES-Data.

References

Angrist, J.D., Imbens, G.W., Rubin, D.B.: Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91(434), 444–455 (1996)
Bates, D., Mächler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67(1), 1–48 (2015). https://doi.org/10.18637/jss.v067.i01
Bell, A., Fairbrother, M., Jones, K.: Fixed and random effects models: making an informed choice. Qual. Quant. 53(2), 1051–1074 (2019)
Bowsher, C.G., Swain, P.S.: Identifying sources of variation and the flow of information in biochemical networks. Proc. Natl. Acad. Sci. 109(20), E1320–E1328 (2012)
Chen, B., Lawson, K.A., Finelli, A., Saarela, O.: Causal variance decompositions for institutional comparisons in healthcare. Stat. Methods Med. Res. 29(7), 1972–1986 (2020)
Chen, B., Lawson, K.A., Finelli, A., Saarela, O.: Causal mediation analysis decomposition of between-hospital variance. Health Serv. Outcomes Res. Method. 22(1), 118–144 (2022)
Clarke, P., Crawford, C., Steele, F., Vignoles, A.: Revisiting fixed-and random-effects models: some considerations for policy-relevant education research. Educ. Econ. 23(3), 259–277 (2015)
Daignault, K., Lawson, K.A., Finelli, A., Saarela, O.: Causal mediation analysis for standardized mortality ratios. Epidemiology 30(4), 532–540 (2019)
Daniels, M.J., Gatsonis, C.: Hierarchical generalized linear models in the analysis of variations in health care utilization. J. Am. Stat. Assoc. 94(445), 29–42 (1999)
Díaz, I., van der Laan, M.: Population intervention causal effects based on stochastic interventions. Biometrics \textbf{68}(2), 541–549 (2012)

Dieleman, J.L., Templin, T.: Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: A simulation study. PLoS One \textbf{9}(10), e110257 (2014)

Donabedian, A.: The quality of care. How can it be assessed? JAMA \textbf{260}(12), 1743–1748 (1988)

Farrell, P.J., Groshen, S., MacGibbon, B., Tomberlin, T.J.: Outlier detection for a hierarchical bayes model in a study of hospital variation in surgical procedures. Stat. Methods Med. Res. \textbf{19}(6), 601–619 (2010)

Fisher, A., Rudin, C., Dominici, F.: All models are wrong, but many are useful: learning a variable’s importance by estimating an entire class of prediction models simultaneously. J. Mach. Learn. Res. \textbf{20}(177), 1–81 (2019)

Goldstein, H., Spiegelhalter, D.J.: League tables and their limitations: statistical issues in comparisons of institutional performance. J. R. Stat. Soc. Ser. A \textbf{159}, 385–443 (1996)

Hawley, S.T., Hofer, T.P., Janz, N.K., Fagerlin, A., Schwartz, K., Liu, L., Deapen, D., Morrow, M., Katz, S.J.: Correlates of between-surgeon variation in breast cancer treatments. Med. Care \textbf{44}, 609–616 (2006)

Hernán, M.A., Robins, J.M.: Estimating causal effects from epidemiological data. J. Epidemiol. Community Health \textbf{60}, 578–586 (2006)

Lawson, K.A., Saarela, O., Abouassaly, R., Kim, S.P., Breau, R.H., Finelli, A.: The impact of quality variations on patients undergoing surgery for renal cell carcinoma: a national cancer database study. Eur. Urol. \textbf{72}(3), 379–386 (2017)

Longford, N.T.: A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Biometrika \textbf{74}(4), 817–827 (1987)

Merlo, J., Chaix, B., Ohlsson, H., Beekman, A., Johnell, K., Hjerpe, P., Råstam, L., Larsen, K.: A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J. Epidemiol. Community Health \textbf{60}(4), 290–297 (2006)

Norberg, R.: Hierarchical credibility: analysis of a random effect linear model with nested classification. Scand. Actuar. J. \textbf{1986}(3–4), 204–222 (1986)

Rabe-Hesketh, S., Skrondal, A., Pickles, A.: Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. J. Econom. \textbf{128}(2), 301–323 (2005)

Racz, M.J., Sedransk, J.: Bayesian and frequentist methods for provider profiling using risk-adjusted assessments of medical outcomes. J. Am. Stat. Assoc. \textbf{105}, 48–58 (2010)

Rubin, D.B., Rosenbaum, P.R.: The central role of the propensity score in observational studies for causal effects. Biometrika \textbf{70}(1), 41–55 (1983)

Shahian, D.M., Normand, S.L.T.: Comparison of risk-adjusted hospital outcomes. Circ. J. Am. Heart Assoc. \textbf{117}, 1955–1963 (2008)

Skrondal, A., Rabe-Hesketh, S.: Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Crc Press, Florida (2004)

VanderWeele, T., Vansteelandt, S.: Conceptual issues concerning mediation, interventions and composition. Stat Interface \textbf{2}, 457–468 (2009)

VanderWeele, T., Vansteelandt, S., Robins, J.: Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology \textbf{25}(2), 300–306 (2014)

Varewyck, M., Goetgebeur, E., Eriksson, M., Vansteelandt, S.: On shrinkage and model extrapolation in the evaluation of clinical center performance. Biostatistics \textbf{15}(4), 651–664 (2014)

Wood, L., Bjarnason, G.A., Black, P.C., Cagiannos, I., Heng, D.Y., Kapoor, A., Kollmannsberger, C.K., Mohammadzadeh, F., Moore, R.B., Rendon, R.A., Soulieres, D., Tanguay, S., Venner, P., Jewett, M., Finelli, A.: Using the Delphi technique to improve clinical outcomes through the development of quality indicators in renal cell carcinoma. J. Oncol. Pract. \textbf{9}(5), e262-267 (2013)

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.