Health Services and Outcomes Research Methodology (2019) 19:61–78
https://doi.org/10.1007/s10742-018-00195-9

Causal difference-in-differences estimation for evaluating the impact of semi-continuous medical home scores on health care for children

Bing Han[1] · **Hao Yu**[2]

Received: 2 May 2018 / Revised: 30 November 2018 / Accepted: 10 December 2018 /
Published online: 9 February 2019

Abstract
Difference-in-differences (DID) is a popular approach in observational and quasi-experimental studies to estimate the effects of a treatment with discrete statuses. In many studies, however, the treatment can have a range of dosages or exposure levels. In our paper, "medical homeness" is a semi-continuous score ranging from 0 to 100 to indicate the extent to which a patient-centered medical home model is achieved. We developed a causal DID approach to estimating the effects of a treatment with semi-continuous dosages. The proposed approach allows for mixed-type designs as well as different propensity models. We applied the proposed approach to evaluate the dosage effect of medical homeness scores on the utilization and quality of children's health care. We found that there was a roughly linear effect of medical homeness scores on the annual number of visits to doctor offices when medical homeness scores were below 60 points. The number of office visits did not further increase when medical homeness scores were above 60. A similar relationship was found between medical homeness scores and ratings for health care quality.

Keywords Medical home · Difference-in-differences · Causal inference

1 Introduction

Over the past 25 years, there has been a growing emphasis on the medical home model to improve the pediatric health care delivery system (American Academy of Pediatrics Medical Home Initiatives for Children With Special Needs Project Advisory Committee and others 2004; Dickens et al. 1992; Sia et al. 2004; Fisher 2008; Rittenhouse and Shortell 2009; Rittenhouse et al. 2010; Stange et al. 2010). The American Academy of Pediatrics (AAP) has specified that a medical home for children is a model of delivering primary care

✉ Bing Han
bhan@rand.org

Hao Yu
haoyu@rand.org

[1] RAND Corporation, Santa Monica, CA, USA

[2] RAND Corporation, Pittsburgh, PA, USA

that is accessible, continuous, comprehensive, family-centered, coordinated, compassionate, and culturally effective (American Academy of Pediatrics Medical Home Initiatives for Children With Special Needs Project Advisory Committee and others 2004). There is also an expectation that children with a medical home will have more appropriate health service use, less unmet need for care, greater satisfaction with care, and lower medical expenditures (Homer et al. 2008). Interest in the medical home has increased with the passage of the landmark 2010 Patient Protection and Affordable Care Act, which designated substantial resources for demonstrating and evaluating medical homes across the nation. Nevertheless, the published evaluation studies have generated mixed evidence about the impacts of a medical home on health care utilization, costs and quality (Cohen et al. 1996; Bethell et al. 2004; Strickland et al. 2004; Damiano et al. 2006; Beal et al. 2009; Romaire and Bell 2010; Stevens et al. 2010; Strickland et al. 2011; Han et al. 2017). Most utilization and cost outcomes did not differ significantly between patients with and without a medical home.

Prior studies have applied a summary score, i.e., medical homeness, to indicate the extent to which primary care practices adhere to the ideal medical home model. A dichotomous medical home status was defined using a cutoff on the medical homeness score. A prominent example is Romaire and Bell (2010), who used 22 survey items in the Medical Expenditure Panel Survey (MEPS) to assess the level of medical home implementation in four domains (accessible, family-centered, coordinated, and compassionated care). Each domain consisted of multiple survey items, each of which was scored on a 0 to 100 likert scale. The domain score was the average of all item scores in that domain. The minimum domain score across all four domains was defined as the medical homeness score. A patient had zero-valued medical homeness if he or she did not have a usual source of care, or had all four domain scores equal zero. The dichotomous medical home status was defined by a cutoff of 75 points on the medical homeness score.

Effects of the medical home were often evaluated using data from observational or quasi-experiment studies and based on the dichotomous medical home statuses. The difference-in-differences (DID) method is among the most popular approaches to account for the lack of randomization in these empirical studies. While a dichotomization of treatment dosages makes it straightforward to conduct the DID analyses, it has some notable restrictions. First, the cutoff can mask important differences in the health care delivered. Although the utilization of primary care should be central to a medical home, few significant differences in various utilization outcomes were found between children with and without a medical home (Romaire and Bell 2010; Strickland et al. 2011; Romaire et al. 2012; Han et al. 2017). A potential reason is that the discrete medical home statuses used in these evaluation studies may have led to inaccurate estimates. For example, the actual health care delivered may differ greatly between medical homeness scores of 0 and 70, but both scenarios were considered not to have a medical home in prior studies. MacCallum et al. (2002) argued that dichotomization of quantitative measurements in quantitative psychology is "rarely defensible and often will yield misleading results". Prior studies based on dichotomized medical home statuses may have the same issues. Second, even if a discrete medical home status is potentially proper, empirical evidence to guide the choice of appropriate cutoff values is still lacking. It is unknown whether a different cutoff value may be more appropriate than the current one in use, i.e., 75 points. Further research is needed to ascertain whether there is a nonlinear relationship between medical homeness and health care utilization and quality outcomes. If the nonlinear relationship is confirmed, it may help identify a more appropriate threshold of medical homeness or justify the current one in use.

Despite the popularity of DID methods, using them to evaluate the effect of medical homeness poses new challenges. While many patients can have a medical homeness score 0, e.g., due to no access to a usual source of care, others can have non-zero medical homeness scores ranging between 0 and 100. Parametric methods can apply stringent assumptions on the functional form of the dosage effects, e.g., a linear dosage effect, but may introduce serious biases due to misspecification. The conventional DID methods require the number of study groups to be small with respect to the total sample size, which is known as the discrete treatment assumption (Conley and Taber 2011). The simplest DID study design has only two groups: one treatment and one control group. However, in many studies, the treatment of interest does not have a small number of discrete levels, but rather a wide range of many dosage levels. Conventionally the DID analysis is implemented by ANOVA or ANCOVA models, which do not formally define a causal effect. In the recent literature, there were several developments to incorporate the Rubin Causal Model (RCM) (Rosenbaum and Rubin 1983) in the DID analyses. For example, causal DID methods with two study arms were discussed by Abadie (2005), Fu et al. (2007), Han et al. (2017). However, causal inference methods that can handle semi-continuous treatment dosages in DID studies are still lacking in the literature.

This study aims to fill the gap in the literature by developing a new DID method for analyzing longitudinal data with semi-continuous treatment dosage and applying it to generate empirical evidences for the effects of medical homeness. We first defined and identified the causal effect of a treatment in the DID setting using the RCM framework. Next, by adapting the generalized propensity score (Hirano and Imbens 2005; Imai and Van Dyk 2004), we developed new causal DID methods for estimating the dosage effect of semi-continuous treatment dosages with inflated zeros. We also presented several alternative implementation strategies for the proposed methods. We applied the new methods to analyze the impact of medical homeness score on health care utilization and quality for children. The results revealed (i) the potentially nonlinear effect of medical homeness on health care utilization and quality, and (ii) an alternative threshold of medical homeness, above which improvement in health care outcomes may level off.

The paper is organized as follows. Section 2 gives a brief review of existing DID methods for discrete treatment statuses. Section 3 presents the theoretical basis for identifying the causal effect of a treatment with semi-continuous dosages. Section 4 introduces the technical details of the proposed causal estimation method. Section 5 shows the results of using the proposed methods to analyze the effect of medical homeness scores on the utilization and quality for children's health care. We end the paper with a discussion in Sect. 6. The appendix includes the proofs for the key results in Sect. 3.

2 A brief review of DID under discrete treatment statuses

2.1 Classic model-based approaches

In a DID study with dichotomous treatment statuses, subjects are measured at two synchronized waves in time (denoted as $t = 0, 1$). At the baseline wave ($t = 0$) all subjects are in the same treatment condition; in the follow-up wave ($t = 1$) some subjects change their treatment condition to a new status, and the other subjects' treatment condition remain unchanged. Hereafter, the treatment status in the baseline wave is referred to as the *control* condition, and the new treatment status for a fraction of subjects receiving in the follow-up

wave is referred to as the *treatment* condition. Let D be a binary variable indicating the two groups of subjects: $D = 0$ if a subject remains in the control condition in both waves (i.e., the *control group* or the 0-0 study arm), and $D = 1$ if a subject changes from the control to the treatment status between the two waves (i.e., the *treated group* or the 0-1 study arm). The DID study design is based on the *parallel trajectory* assumption (also known as *parallel paths*, *parallel trends*, and *constant biases*), namely, that the mean outcome of the two groups would have followed similar trajectories over time, had the two groups shared the same (un)change in treatment statuses.

Conventionally, a two-way ANOVA with interaction is applied to implement the DID analysis (Lechner 2011),

$$Y_t = \beta_0 + \beta_1 t + \beta_2 D + \beta_3 Dt + \epsilon_t, \tag{1}$$

where Y_t is the outcome measured at $t = 0, 1$, ϵ_t is the error term and customarily assumed to be correlated within the same subject. The coefficient β_2 reflects the baseline difference between the two groups in the mean outcome. The coefficient β_1 is the slope for the control group. By the parallel trajectory assumption, the treated group would have the slope β_1 had they not received treatment at $t = 1$. Therefore, the interaction coefficient β_3 corresponds to the treatment effect. In a randomized trial, the baseline difference β_2 should not be significant, but controlling for it can enhance the estimation efficiency for the treatment effect β_3 in the ANOVA model (1) (Yang and Tsiatis 2001). In the lack of randomization, the baseline difference β_2 can be significant due to various sample selection mechanisms.

The ANOVA model can be readily expanded to an ANCOVA model to incorporate observed covariates, as well as their interactions with other terms in Eq. (1). The classic DID analysis can use both longitudinal measurements and repeated cross-sectional data (Lechner 2011). A mixture of both data types, i.e., some subjects have only one measurement and others have both measurements over time, is also allowed, provided that the missing data mechanism is ignorable. More than one measurement time in each wave, as well as unsynchronized measurement times, is also allowed by specifying suitable temporal trajectories in (1).

2.2 Semiparametric approaches based on the Rubin Causal Model

For various reasons, the classic model-based approaches can be biased for estimating the causal effect of a treatment. In particular, the mean function in ANOVA or ANCOVA may be misspecified in the presence of multiple pre-treatment covariates. Certain higher order terms or interactions may be omitted. In the lack of randomization, the pre-treatment covariates are usually imbalanced between groups. Covariate imbalance, in conjunction with some misspecification of (1), can lead to serious biases in estimating a treatment effect.

To address these issues, several studies used the Rubin Causal Model (RCM) as a framework to define and identify causal effects in the DID setting (Lechner 2011; Fu et al. 2007; Abadie 2005; Han et al. 2017). By the RCM, we can define the *potential outcome under treatment at time t*, denoted by $Y_t(1)$, as the outcome that would be observed at time t under treatment. Similarly, we can define the *potential outcome under control at time t*, denoted as $Y_t(0)$. The average treatment effect (ATE) at follow-up, ATE$= E[Y_1(1) - Y_1(0)]$, is the causal effect of the treatment by definition. Since at the baseline all subjects are in the control condition, the notation is simplified to $Y_0 = Y_0(0)$. By contrast, the observed outcome at $t = 1$ is $Y_1(D) = DY_1(1) + (1 - D)Y_1(0)$. Since $E[Y(D)|D = 1]$ is usually not equal to the

unconditional mean $E[Y_1(1)]$, the mean observed outcomes cannot directly produce an unbiased estimate for ATE. Further assumptions are needed to identify the unconditional expectation. A basic assumption under the RCM is ignorability. In the classic DID setting, it is

$$Y_1(D^*) - Y_0 \perp D|\mathbf{X}, \text{ for all } D^* \in \{0, 1\}. \tag{2}$$

where $\perp$ stands for statistical independence. Namely, conditioning on pre-treatment covariates $\mathbf{X}$, potential outcomes would follow parallel trajectories between the two groups defined by the observed treatment statuses D. Other standard technical assumptions include: (i) one's potential outcome is not dependent on others' potential outcomes; (ii) the distributions $F(\mathbf{X}|D = 1)$ and $F(\mathbf{X}|D = 0)$ have the same support; (iii) the probability of receiving treatment $P(D = 1|\mathbf{X})$ is bounded away from 0 and 1.

While the classic RCM does not allow for unobserved confounding covariates, a causal DID study allows for certain unobservables related to the potential outcomes. First, time-invariant unobservables are always allowed. Imbalanced time-invariant unobservables between the two treatment groups result in a between-group difference in the baseline outcome Y_0, which cannot be eliminated by conditioning on the observed $\mathbf{X}$. Such a time-invariant difference is removed by differencing the outcomes in (2) over time. Second, time-varying unobservables, if different between the two treatment groups, can violate the assumption (2) and make the trajectories of potential outcomes not parallel no matter how $\mathbf{X}$ is adjusted (Rosenbaum 1984). If there is more than one measurement time in the pre-treatment wave, a "placebo test" can be conducted by inserting a fake intervention time in the pre-treatment wave and conducting a DID analysis for the pre-treatment wave only (Slusky 2017). However, results of placebo tests are not direct evidence for or against (2). Justification for (2) is ultimately based on researchers' substantive knowledge.

The ignorability assumption can be rewritten using the *propensity score*, i.e., the conditional probability of receiving treatment status $P(D = 1|\mathbf{X})$ given pre-treatment covariates,

$$Y_1(D^*) - Y_0 \perp D|P(D = 1|\mathbf{X}), \text{ for all } D^* \in \{0, 1\}. \tag{3}$$

A family of estimators based on propensity scores have been applied to the classic DID study design, such as inverse propensity score weighting (Abadie 2005), stratification and matching (Lechner 2011; Fu et al. 2007), and the doubly robust method (Han et al. 2017).

A weaker version of the identification assumption restricts D^* to 0 in (2) and (3) (Abadie 2005; Lechner 2011; Stuart et al. 2014). This weaker assumption describes the trajectory of the potential outcome in the absence of treatment, and is thus only sufficient to identify the causal effect for the 0-1 arm, which is referred to as the average treatment effect on the treated (ATT). The identification assumption (2) can be seen as assuming a homogeneous effect between the treated and the control groups, after conditioning on covariates. By contrast, the ATT assumes different effects and only focuses on the effects for the treated group. Although neither version of the assumption can be empirically verified, the weaker assumption is preferred by some authors for substantive reasons. Methodologically, the two versions of the assumption result in minor differences in estimating approaches.

3 The average dosage effect and its identification

3.1 Average dosage effect (ADE)

In many studies, a subject's treatment status has a range of dosage levels. Although the total sample size is usually large in a non-experimental study, at any specific dosage level

there are usually a very small number of subjects. Conventionally these dosage levels were combined to two groups by a hard cutoff on the treatment dosages, and an existing DID method for binary treatment status was applied. Conceivably, selecting the cutoff is crucial, because an improper cutoff can suppress any differences among treatment dosages within the same discretized treatment level. Ideally, the choice of the cutoff should be based on theory, empirical evidence, or both. Nevertheless, in many applications, prior knowledge is lacking and the cutoff may have been set somewhat arbitrarily. Sometimes more than one cutoff is applied to discretize treatment dosages into three or more levels to partially address this issue, because with more discretized treatment levels, there is a smaller difference in dosages within each discretized level. The classic DID method can then be applied in a pairwise fashion. A slight extension of the classic DID uses a set of indicators for study arms in the same model. Conley and Taber (2011) discussed asymptotic consistency of the classic DID estimators with more than two arms.

Hereafter we use d to denote the observed continuous dosage level, where d is a scalar random variable. For simplicity in notation, we assume that d takes value in [0, 1], where 0 is absolutely no exposure (e.g., medical homeness score 0) and 1 is full exposure to the ideal treatment statuses (e.g., medical homeness score 100). Although subjects might take any dosage level in either wave of measurement, to define an unambiguous causal parameter we focus on those subjects who had the dosage level of 0 in the baseline wave, i.e., $d = 0$ when $t = 0$. This is usually possible in a quasi-experiment where no one had any opportunity to be exposed to any treatment in the pre-treatment stage. In the follow-up wave, many subjects might remain at $d = 0$, and others could have a non-zero dosage ranging in (0, 1]. Under this setting, d is a semi-continuous random variable with a point mass at 0. The observed outcomes are denoted as Y_0 and $Y_1(d)$. Similar to the classic two-arm DID setting, we can still group the subjects into two arms, denoted as 0-0 and 0-d. The 0-0 arm consists of subjects who remained at $d = 0$ in the follow-up wave, and the 0-d arm has subjects who changed their treatment status to $d > 0$ in the follow-up wave. Nevertheless, the 0-d arm contains many dosage levels.

Under our setting of semi-continuous treatment dosages, there is only one potential outcome Y_0 in the baseline, but infinitely many potential outcomes in the follow-up wave. Let δ denote a hypothetical and fixed dosage level, i.e., δ is a scalar constant in [0, 1]. The potential outcome at follow-up for a certain dosage level δ is denoted as $Y_1(\delta)$. Only one of the many potential outcomes, corresponding to $\delta = d$, is observed.

By the potential outcomes, the average dosage effect (ADE) for a fixed dosage level $\delta, \delta \in [0, 1]$ is defined as

$$\mathrm{ADE}(\delta) = E[Y_1(\delta) - Y_1(0)]. \tag{4}$$

The ADE represents the causal effect of a treatment dosage compared with the zero dosage level at the follow-up wave. By definition, $\mathrm{ADE}(0) = 0$. Same as in the classic setting, the expected value of the observed outcome $E[Y_1(d)|d = \delta]$ is usually not equal to the marginal expectation $E[Y_1(\delta)]$, unless the dosage level d is independent of the potential outcomes.

3.2 Identification of ADE

To identify the ADE, we first adapt the ignorability assumption for continuous treatment dosages from the cross-sectional studies (Hirano and Imbens 2005) to the DID setting

$$Y_1(\delta) - Y_0 \perp d|\mathbf{X}, \text{ for all } \delta \in [0, 1]. \tag{5}$$

The ignorability assumption (5) is analogous to the parallel trajectory assumption (2) for discrete treatment statuses. Essentially, (5) assumes that pre-treatment covariates $\mathbf{X}$ are sufficient to unconfound the trajectory of the potential outcome for a hypothetical dosage level δ. Similar to assumption (2) in the discrete treatment statuses, assumption (5) allows for certain unobservables. First, any time-invariant unobservables are allowed, irrespective of their imbalance across dosage levels. Second, time-varying unobservables have to be either non-existent or balanced across dosage levels. Imbalanced time-varying unobservables across dosage levels will violate (5). As in Sect. 2.2, if there is more than one measurement time in the pre-treatment wave, a placebo test may be applied to test the impact of time-varying observables in the pre-treatment wave. However, such placebo tests cannot directly shed light on (5) which is across both pre- and post-treatment waves.

Next, we define three concepts: the general propensity function (GPF), the generalized propensity score for a fixed dosage δ (GPSf), and the generalized propensity score for the observed dosage (GPSo).

Definition 1 The general propensity function (GPF) is the conditional distribution function of dosage d given covariates $\mathbf{X}$, denoted as $\mathcal{G}$

$$\mathcal{G} = F_d(a|\mathbf{X}) = P(d \leq a|\mathbf{X}), \text{ for all } a \in [0, 1]. \tag{6}$$

Definition 2 The generalized propensity score for a fixed dosage (GPSf) is the density function of the GPF evaluated at a fixed dosage level δ, denoted as r^δ

$$r^\delta = f_d(\delta|\mathbf{X}). \tag{7}$$

Definition 3 The generalized propensity score for the observed dosage (GPSo), denoted as R, is defined as

$$R = f_{d_0}(d|\mathbf{X}), \tag{8}$$

where d_0 is a random variable sharing the same distribution as d, and $d_0 \perp d$.

The essential concept of GPF was originally proposed by Hirano and Imbens (2005) using density functions and called as the generalized propensity score. However, by defining it using the distribution function we aim to improve the technical rigor in both definitions and subsequent theoretical discussions. We also emphasize that given $\mathbf{X}$, GPF is a distribution function rather than a scalar score. By contrast, given $\mathbf{X}$, GPSf is a degenerated constant and GPSo is a scalar random variable. Without conditioning on $\mathbf{X}$, GPF describes a stochastic process indexed by $\mathbf{X}$, and both GPSf and GPSo are univariate random variables.

Note that these definitions encompass the ordinary propensity score for a binary treatment, i.e., $P(D = 1|\mathbf{X})$. Since the probability $P(D = 1|\mathbf{X})$ suffices to describe a Bernoulli distribution, the regular scalar-valued propensity score is equivalent to GPF, GPSf, and GPSo altogether for a binary treatment. However, for continuous dosages these three concepts are distinct.

Some theoretical properties of GPF, GPSf, and GPSo are listed below. The sketch of the proof is in the appendix.

Properties

1. $E(I\{d \leq \delta\}|\mathbf{X}) = \int_0^\delta r^a \mathrm{d}a$;

2. $P(r^{\delta} = R|d = \delta) = 1$;
3. $d \perp \mathbf{X}|\mathcal{G}$;
4. $Y_1(\delta) - Y_0 \perp d|\mathcal{G}$, for any $\delta \in [0, 1]$;
5. $E[Y_1(\delta) - Y_0|r^{\delta} = s] = E[Y_1(d) - Y_0|d = \delta, R = s]$, provided that $s \neq 0$.

Properties 1 to 3 directly follow from the definitions. Property 4 suggests that the GPF $\mathcal{G}$ suffices to unconfound the trajectories of potential outcomes, in the same way as the covariates $\mathbf{X}$. Property 5 further simplifies the identification by equating the mean trajectory of potential outcomes to that of observed outcomes, both for a fixed dosage δ and conditioning on GPSf and GPSo, respectively.

The last two properties result in practically useful estimators for ADE. We state them as the following two observations. The first observation is based on Property 4 and assumes a parametric GPF, cf. Imai and Van Dyk (2004). The second observation is based on Property 5 and does not require a specific distribution family for GPF.

Observation 1 If the GPF belongs to a parametric distribution family, i.e., $\mathcal{G} = F_{\theta(\mathbf{X})}(d)$, where $\theta(\mathbf{X})$ is a p-dimension parameter vector, then the GPF is equivalent to the parameter vector $\theta(\mathbf{X})$. By Property 4, we have

$$Y_1(\delta) - Y_0 \perp d|\theta(\mathbf{X}), \text{ for any } \delta \in [0, 1]. \tag{9}$$

In the literature, the GPF has been assumed to belong to a normal distribution family. If the variance of the normal distributions is identical, then $\theta(\mathbf{X}) = E(d|\mathbf{X})$ consists of only one parameter ($p = 1$). In this simplest case, the GPF is a scalar score, (cf. the propensity score for dichotomous treatment statuses). If the conditional variances $Var(d|\mathbf{X})$ are unequal, two parameters ($p = 2$) are needed, e.g., $\theta(\mathbf{X}) = (E(d|\mathbf{X}), Var(d|\mathbf{X}))$ (Hirano and Imbens 2005; Imai and Van Dyk 2004). Since parametric families usually have a small number of parameters, p is much smaller than the number of covariates $\mathbf{X}$. Thus, Observation 1 greatly reduces the technical difficulty in understanding and operationalizing the GPF. Conditioning on $\theta(\mathbf{X})$, the expected value of an unobserved $Y_1(\delta) - Y_0$ for a dosage of interest δ is equal to that of $Y_1(d) - Y_0$, provided the observed $d = \delta$. The main drawback is the strong parametric assumption on the GPF implied by $\theta(\mathbf{X})$, which is difficult to justify.

To overcome the limitation in Observation 1, in this paper we consider an alternative way for identifying the ADE based on Property 5.

Observation 2 By Property 5 and for a fixed $\delta \in [0, 1]$

$$\begin{aligned} &E[Y_1(\delta) - Y_0] \\ &\quad = E\{E[Y_1(\delta) - Y_0|r^{\delta}]\} \\ &\quad = E\{E[Y_1(d) - Y_0|d = \delta, R = r^{\delta}]\}. \end{aligned} \tag{10}$$

Observation 2 suggests that by iterative expectations, one can estimate the mean trajectory of the potential outcomes by the mean of observed outcomes. Applying Observation 2 for $d = \delta$ and $d = 0$, we have identified $\text{ADE}(\delta)$ by the usual DID form

$$\begin{aligned} \text{ADE}(\delta) &= E[Y_1(\delta) - Y_0] - E[Y_1(0) - Y_0] \\ &= E\{E[Y_1(d) - Y_0|d = \delta, R = r^{d}]\} - E\{E[Y_1(0) - Y_0|d = 0, R = r^{0}]\} \end{aligned} \tag{11}$$

4 Estimation strategy

Observation 2 suggests multiple ways to construct estimators for ADE(δ). We consider the following empirical estimation strategy. First, we can impose a flexible outcome regression model for $g_1(d,R) = E[Y_1(d)|d,R]$ and $g_0(d,R) = E(Y_0|d,R)$. Such outcome models can provide estimations at $d = \delta, R = r^\delta$, denoted as $g_1(\delta, r^\delta), g_0(\delta, r^\delta)$. By Observation 2, $g_1(\delta, r^\delta) - g_0(\delta, r^\delta) = E[Y_1(\delta) - Y_0|r^\delta]$. Next, averaging across the marginal distribution of $\mathbf{X}$ gives the estimate for ADE(δ). Using subscript i for each subject, we can write a general form for the estimator of ADE(δ),

$$\widehat{\text{ADE}}(\delta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{g}_1(\delta, r_i^\delta) - \hat{g}_0(\delta, r_i^\delta) + \hat{g}_0(0, r_i^0) - \hat{g}_1(0, r_i^0)\right\}, \tag{12}$$

where n is the sample size.

Repeating $\widehat{\text{ADE}}(\delta)$ at all δ values yields a complete dosage-effect relationship. Since no restriction is assumed among different dosage levels, the dosage effect is flexible and nonlinear. If g_1, g_0, and R are all continuous in d, then the dosage-effect relationship $\widehat{\text{ADE}}(\delta)$ is also continuous in δ.

Equation (12) can be applied to repeated cross-sectional data. For longitudinal measurements, the estimator can be simplified by first taking within-subject difference $\Delta Y(d) = Y_1(d) - Y_0$. Let $g = E[\Delta Y(d)|d,R]$. The estimator for $ADE(\delta)$ is simplified to

$$\widehat{\text{ADE}}(\delta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\hat{g}(\delta, r_i^\delta) - \hat{g}(0, r_i^0)\right\}. \tag{13}$$

4.1 Modeling the GPF

Before we apply the empirical estimators (12) and (13), we need to first estimate the GPF, which in turn provides estimated GPSf and GPSo. We applied the usual two-part modeling approach for handling semi-continuity, where the dosage d is redefined to (i) a dichotomized status $I\{d \neq 0\}$, and (ii) $d^* = d$ when $d > 0$. The two parts are modeled separately, and the GPF is given by combining the two parts

$$F_d(a|\mathbf{X}) = \begin{cases} P(d = 0|\mathbf{X}), \text{ if } a = 0, \\ P(d \neq 0|\mathbf{X})P(0 < d < a|d \neq 0, \mathbf{X}), \text{ if } a > 0 \end{cases}$$

In the two-part model, the first part is to model the dichotomized $I\{d \neq 0\}$ by a generalized regression model, such as a logistic regression or a generalized additive model. The second part is to model the subsample of d^* by a regression model for the continuous outcomes, such as a linear regression or a smoothing spline. Note that the regression technique applied in the second part should be able to provide the prediction inference for $F(d|\mathbf{X}, d > 0)$. Depending on the specific choice in each step, the GPF can be estimated by parametric, semiparametric, or completely nonparametric tools.

In this paper, we consider a parametric approach and a semi-parametric approach to modeling the GPF. In the parametric approach, we use a regular logistic regression for the binary part. We rescale the non-zero medical homeness score to (0, 1), and apply a beta regression (Ferrari and Cribari-Neto 2004) for the second part, namely,

$$\begin{aligned} &logit(P(d=0|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \\ &logit(E(d^*|\mathbf{X}) = \mathbf{X}\boldsymbol{\gamma}, \end{aligned} \tag{14}$$

where d^* follows a beta distribution with mean $E(d^*|\mathbf{X})$ and variance $\frac{1}{\phi}E(d^*|\mathbf{X})(1-E(d^*|\mathbf{X}))$. This parametric approach fully specifies the GPF.

In the semi-parametric approach, we use the generalized additive model (GAM) (Hastie 2017) for the binary part and the multivariate kernel regression (Li and Racine 2007) for the second part. Many nonparametric regression techniques, including the multivariate kernel, do not specify the prediction distributions for the continuous outcome. Instead, they may provide estimates for the mean and variance functions. We consider a semi-parametric approach, in which we first use a nonparametric regression for estimating the conditional mean and variance of d^* given $\mathbf{X}$. We then assume that d^* follows a beta distribution given $\mathbf{X}$. The mean and variance are sufficient to specify a beta distribution. Namely,

$$\begin{aligned} &logit(P(d=0|\mathbf{X}) = h(\mathbf{X}) \\ &E(d^*|\mathbf{X}) = m(\mathbf{X}), Var(d^*|\mathbf{X}) = v(\mathbf{X}), \end{aligned} \tag{15}$$

and d^* follows a beta distribution with mean $m(\mathbf{X})$ and variance $v(\mathbf{X})$.

4.2 Outcome modeling and recycled prediction

To apply the estimators (12) or (13), we need to fit outcome regressions between an observed outcome and the observed dosage d and GPSo R. In this paper, we apply a pth order polynomial basis regression. A model-selection criterion, such as AIC or cross-validation, can be used to help choose a proper smoothing order p. Other smoothing techniques, such as bivariate kernel, are equally applicable. We use longitudinal data to illustrate the procedure.

$$g(d,R) = E(\Delta Y|d,R) = \sum_{j=0}^{p}\sum_{k=0}^{p}\beta_{j,k}d^jR^k, \tag{16}$$

which can be fitted by applying the ordinary least squares estimator to the data set $(\Delta Y, d, R)$.

Estimator (13) is implemented through recycled prediction using fitted outcome model (16). For each subject $i = 1, \ldots, n$ we first calculate r_i^{δ}, the GPSf for dosage level δ . Using the fitted outcome we can estimate $\hat{g}(\delta, r_i^{\delta})$. Next, plugging in the estimated $\hat{g}(\delta, r_i^{\delta}), i = 1, \ldots, n$ to (13), we can calculate $\widehat{\text{ADE}}(\delta)$. Repeat this process for all other δ levels yields the entire ADE curve. Given the complexity in the estimation strategy, we applied the bootstrapping method to assess the sampling error in $\widehat{\text{ADE}}(\delta)$.

Estimation steps for repeated cross-sectional data are similar. The only difference is that we need to fit $g_1(d,R)$ and $g_0(d,R)$ separately by two outcome models, and recycled predictions are conducted for these two outcome models separately by (12).

5 Evaluating of dosage effect of medical homeness scores

5.1 Medical homeness, covariates, and outcomes

Instead of using the dichotomous medical home status, we aimed to study the continuous dosage effect of medical homeness on health care utilization and quality rating outcomes. We used nine MEPS panels from 2004 to 2012, where each MEPS panel consists of a nationally representative sample followed-up for 2 years, allowing for a DID analysis. The study sample consists of 3466 children, who were 17 years old or younger and had a medical homeness score 0 in the first year of each panel. Among them, 2112 children had zero-valued medical homeness in year 2 (i.e., the 0-0 arm), and the rest 1354 had non-zero medical homeness in year 2 (i.e., the 0-d arm). The mean and standard deviation of medical homeness scores in year 2 of the 0-d arm is 63.1 and 23.6, respectively. The distribution of medical homeness is roughly bell-shaped but a little skewed to the left. Only 597 children, or 44.1% in the 0-d arm, surpassed the conventional cutoff of 75 points in year 2. Therefore, the conventional dichotomous medical home status would have led to very different study arm compositions compared with the arms based on the continuous dosages.

In this analysis, we focused on three outcomes: the annual number of doctor visits, the annual number of emergency room (ER) visits, and the ratings of overall health care quality. The two utilization outcomes are counts, and the quality rating outcome is on a 1 to 10 likert scale and treated as a numeric outcome. The two utilization outcomes have repeated measures on each pediatric patient. The quality rating outcome is of a mixed data type with many patients only reporting in one of the 2 years.

We used the same set of baseline covariates as in Han et al. (2017), who analyzed these outcomes by dichotomized medical home statuses. Selection of covariates were first based on the Andersen's model of health care seeking behavior (Andersen 1995), including (1) predisposing characteristics (age, gender, race/ethnicity), (2) enabling factors (family income, insurance, and urban residence), and (3) need factors (parent-perceived physical and mental health statuses, and whether their child has a special health care need). In addition, we included health insurance status: covered by private insurance, covered by public insurance (Medicaid, CHIP, or Medicare), uninsured, and unknown insurance status. We also included dummy indicators for survey panels to adjust for between-panel differences unrelated to the research question. Table 1 presents the descriptive statistics by dividing the sample into the 0-0 and 0-d arms.

5.2 Estimation and results

Medical homeness scores were rescaled to between 0 and 0.99 (dividing the medical homeness core by 101) to avoid the right edge of a beta distribution where the density is usually 0 and violates the assumption in Property 5. We used the parametric and semiparametric approach described in Sect. 4.1 to fit the GPF. We used the bivariate power series (16) to model the outcome. For the two utilization outcomes, we took within-person difference and fit a single outcome model. For the quality rating outcome, we fitted two outcome models for each wave separately because many patients reported the outcome only in one of the 2 years. We fitted all possible models with $p \leq 4$ and used AIC to select the smoothing order as well as interaction terms in the outcome model. All selected models have the highest polynomial order below 3. Results are very similar between the two sets of GPF fitting. We omitted the results from the semiparametric GPF and only presented the results

Table 1 Descriptive statistics of the MEPS data: numbers are mean(SD) or percentage points

	0-0 arm (n = 2112)	0-d arm (n = 1354)	Total (n = 3466)
Medical homeness score	0 (0)	63.1 (23.6)	24.7 (34.1)
% having binary medical home status	0	44.1	17.2
#doctor visits in year 1	0.8 (3.1)	1.9 (3.9)	1.2 (3.5)
#doctor visits in year 2	0.7 (2.5)	3.2 (5.1)	1.7 (3.9)
# ER visits in year 1	0.10 (0.36)	0.17 (0.52)	0.12 (0.43)
# ER visits in year 2	0.09 (0.36)	0.23 (0.64)	0.14 (0.49)
Quality rating in year 1	8.3 (2.1)	8.1 (2.2)	8.2 (2.1)
Quality rating in year 2	8.4 (1.9)	8.7 (1.5)	8.6 (1.7)
Age in year 1	10.9 (4.9)	6.7 (5.7)	9.3 (5.6)
% males	51.3	51.3	51.3
Race/ethnicity			
Hispanics	39.7	48.5	45.1
Non-Hispanic black	22.5	22.5	22.5
Non-Hispanic Asian	4.1	5.2	4.8
Non-Hispanic others	33.6	23.8	27.6
% poor or near poor	45.6	46.8	46.4
% MSA status in year 1: yes	84.6	87.0	86.1
% Insurance status in year 1			
No insurance	17.2	43.4	33.2
Medicaid	48.0	32.2	38.4
Private insurance	29.8	21.2	24.6
% CSHCN in year 1	12.6	8.3	10.0
% Perceive poor or fair health status in year 1 (round 1)	4.0	3.0	3.4
% Perceive poor or fair mental health status in year 1 (round 1)	3.2	3.0	3.1

based on the parametric GPF. We also applied 200 iterations of bootstrapping the entire estimation procedure including estimation of the GPF and the outcome models to produce pointwise 95% confidence intervals.

Figure 1 presents the estimated dosage effect for the three outcomes. Note that the ADE must be exactly zero with no uncertainty when the dosage is 0 by definition. For the annual number of doctor visits, there is a roughly linear increasing trend when the medical homeness score is approximately below 60 points, and then the effect reaches a plateau of roughly one doctor visit a year for the medical home score between 60 and 100 points. At any medical homeness level, the ADE is always significantly above zero. The findings for quality rating are similar, where a roughly linear effect reaches the peak of 0.4 points around medical homeness scores of 60, and drops slightly afterwards. For the annual number of ER visits, there is no clear dosage effect, and the 95% confidence intervals always contain zero. All confidence intervals are wide near the maximum medical homeness score of 100, likely because very few patients have medical homeness scores that high.

These results agree with the previous findings in the literature that medical homes can have mixed effects in different outcomes. It is reasonable to expect that medical homeness is positively related to quality of care and the number of office visits. However, the most

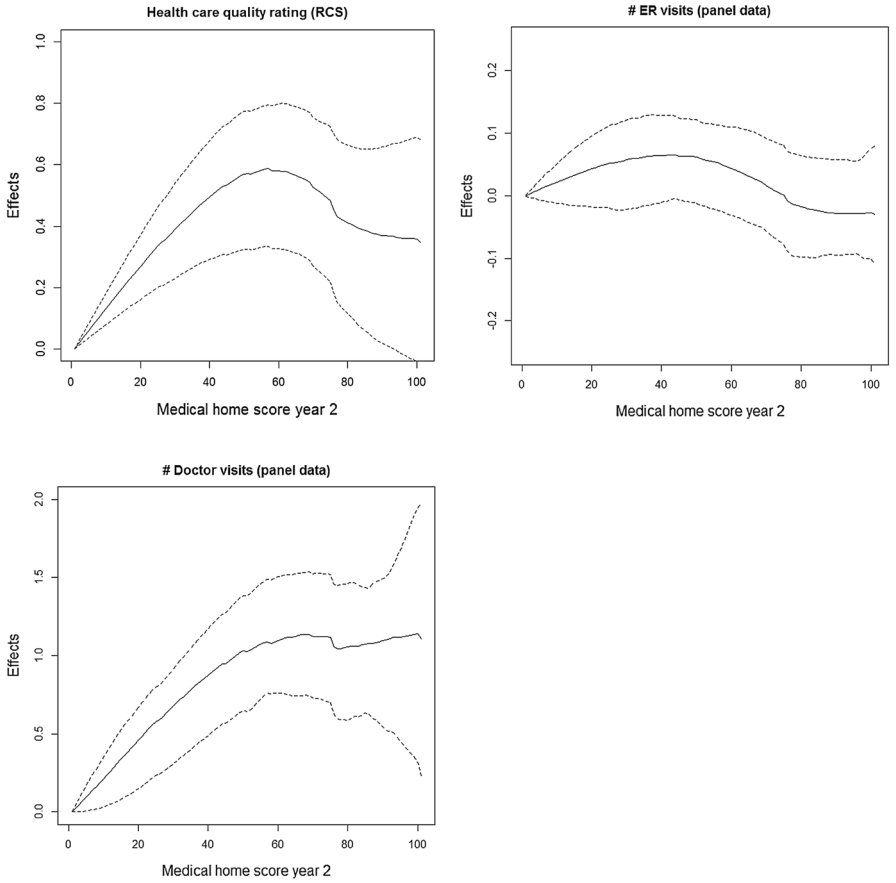

Fig. 1 Estimated average dosage effects for medical homeness on health services outcomes: solide lines are point estimates and dashed lines are pointwise 95% confidence intervals

common reasons for ER visits among children are injury and virus infections, which may not be affected by their primary care.

The dosage effect reveals more details for the two outcomes with significant effects of medical homeness. The linear trend in the lower range of medical homeness scores suggests that any improvement in primary care can improve patients' health services outcomes to a certain extent. In fact, most of the effect of a medical home is realized by changing from having poor access to primary care (e.g., a medical homeness score of 0) to having fair primary care access (e.g., medical homeness scores around 60). The effect of medical homeness has been almost fully reached below the conventional cutoff of 75 points. Therefore, the conventional cutoff of 75 points is acceptable, but perhaps overly stringent.

6 Conclusion and discussion

As a popular method in the empirical literature, the DID approach has been increasingly applied to observational and quasi-experimental data. However, its requirement for discrete treatment statuses may not be met in large-scale policy implementations where the treatment levels are more similar to a continuous dosage. The ongoing efforts to promote the medical home model across the nation aim to improve health care outcomes by increasing medical homeness with a range of numerical dosages. With these efforts comes the need to develop novel DID methods by relaxing the discrete treatment assumption for analyzing the effects of interventions with semi-continuous dosages.

To meet the need, we have extended the conventional DID method to analyzing a continuous-dosage intervention based on the causal inference framework of RCM. We present properties of the GPF, GPSf, and GPSo, which lead to identifications for the causal effect ADE(δ). We also provide several empirical estimators with implementation details. We have applied the proposed methods to an analysis of how medical homeness affects health care outcomes for children. The results revealed a nonlinear relationship between medical homeness and health care utilization and quality. To our knowledge, our study is the first to identify the flexible nonlinear relationship and to confirm a continuous dosage effect on some outcomes. For example, while there is a roughly linear effect of medical homeness score on the annual number of visits to doctor offices if the score is under 60, little change in the number of visits occurs if the score increases between 60 and 100. There is similar relationship between medical homeness and parents' rating of health care quality for children. On the other hand, we did not find a significant relationship between medical homeness and ER visits. Parts of our findings, such as quality improvement associated with medical homeness, are consistent with prior studies (Domino et al. 2009), whereas other parts of our findings are different from previous research. For example, one study by Romaire et al. (2012) found that having a medical home significantly reduced the number of ER visits, compared with no significant relationship found in our study. One possible reason for the different results is that their study used the binary medical home variable, which does not necessarily reflect the exact change in medical homeness. Based on the new DID method, our study not only provided a more accurate depiction of the effect of medical homeness, but also revealed the threshold above which the medical home effect leveled off. Such a threshold can potentially be used as a cutoff value for those who insist on using a binary variable of medical home status.

This study has some limitations. First, theoretically the RCM depends on the critical ignorability assumption. We have used a rich list of covariates in MEPS based on behavioral theory as well as the previous studies in the medical home literature. With only two waves of data in each panel, we were not able to run placebo tests for the pre-treatment wave. However, our analysis sample includes a relatively long study period of nine panels (2004–2012). An unobserved time-varying covariate would need to confound with the medical homeness score over the long study period and across all the children age cohorts to bias the results. It is unclear to us what unobserved covariate could have such persistent impact. Second, the proposed estimation procedure for ADE consists of multiple steps of modeling. In particular, it is difficult to diagnose the two-part model fitting for the GPF. Although the results were not sensitive to the choice of GPF modeling approach in our case study, the influence of potential misspecification in GPF needs further investigation. Third, the bootstrap procedure for evaluating the sampling error is computationally demanding: in each iteration we needed to refit the two-part GPF as well as the outcome models, and redo

the recycled predictions. Fourth, in reality many patients had a non-zero but varying medical homeness over time. One may call this situation as $d_1 - d_2$, where both d_1 and d_2 were non-zero. However, the definition of ADE, as well as the approach to estimating ADE, does not directly apply to the situation of $d_1 - d_2$. Conceivably, a stronger version of the ignorability assumption than (2) is needed for studying $d_1 - d_2$. Fifty, this paper is focused on the ADE, which corresponds to the ATE under the dichotomous treatment paradigm. A concept corresponding to the ATT under the dichotomous treatment paradigm can certainly be based on this work and may require only minor amendments to the method. However, further investigations are needed to verify this conjecture. Lastly, from the substantive perspective, our measurement of medical homeness was based on survey information about parents' reported experience with health care for their children. The survey information is subject to recall bias, and may not reveal the full picture of health care delivery.

Funding This study was funded by the Grant R21HD078881 from Eunice Kennedy Shriver National Institute of Child Health & Human Development and by the Grant R01HS023336 from the Agency for Healthcare Research and Quality.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix

Properties 1 to 3 are simply by the definition of $\mathcal{G}$, r^δ, and R.

Proof of Property 4 For any $\zeta \geq 0$ and by iterative expectations,

$$P(d \leq \zeta | Y_1(\delta) - Y_0, \mathcal{G}) = E\Big\{ E\big[I\{d \leq \zeta\} | Y_1(\delta) - Y_0, \mathbf{X}\big] \Big| Y_1(\delta) - Y_0, \mathcal{G} \Big\}.$$

By the ignorability assumption, the last expression is equal to

$$E\Big\{ E\big[I\{d \leq \zeta\} | \mathbf{X}\big] | Y_1(\delta) - Y_0, \mathcal{G} \Big\} = E\Big\{ \int_0^\zeta r^a \mathrm{d}a \Big| Y_1(\delta) - Y_0, \mathcal{G} \Big\}, \text{ by Property 1.}$$

Since r^a is fully determined by $\mathcal{G}$, the last expression is equal to $\int_0^\zeta r^a \mathrm{d}a = P(d \leq \zeta | \mathbf{X})$. □

Proof of Property 5 By Property 2,

$$E[Y_1(d) - Y_0 | d = \delta, R = s] = E[Y_1(\delta) - Y_0 | d = \delta, R = s] = E[Y_1(\delta) - Y_0 | d = \delta, r^\delta = s]. \tag{17}$$

Then following essentially the same technique in Theorem 2 in Hirano and Imbens (2002), we first show $f(Y_1(\delta) - Y_0 | d = \delta, r^\delta = s) = f(Y_1(\delta) - Y_0 | r^\delta = s)$.

Given $s \neq 0$,

$$f(Y_1(\delta) - Y_0 | d = \delta, r^\delta = s) = \frac{f(d = \delta | Y_1(\delta) - Y_0, r^\delta = s) f(Y_1(\delta) - Y_0 | r^\delta = s)}{f(d = \delta | r^\delta = s)}$$
$$= s^{-1} f(d = \delta | Y_1(\delta) - Y_0, r^\delta = s) f(Y_1(\delta) - Y_0 | r^\delta = s).$$

In addition,

$$\begin{aligned}
&f(d = \delta | Y_1(\delta) - Y_0, r^\delta = s) \\
&= \int f(d = \delta, \mathbf{X} | Y_1(\delta) - Y_0, r^\delta = s) \mathrm{d}\mathbf{X} \\
&= \int f(d = \delta | \mathbf{X}, Y_1(\delta) - Y_0, r^\delta = s) f(\mathbf{X} | Y_1(\delta) - Y_0, r^\delta = s) \mathrm{d}\mathbf{X} \\
&= \int f(d = \delta | \mathbf{X}, r^\delta = s) f(\mathbf{X} | Y_1(\delta) - Y_0, r^\delta = s) \mathrm{d}\mathbf{X}, \quad \text{by ignorability,} \\
&= \int s f(\mathbf{X} | Y_1(\delta) - Y_0, r^\delta = s) \mathrm{d}\mathbf{X} = s.
\end{aligned}$$

Therefore, $f(Y_1(\delta) - Y_0 | d = \delta, r^\delta = s) = f(Y_1(\delta) - Y_0 | r^\delta = s)$.

Revisit the last expression in (17),

$$\begin{aligned}
&E[Y_1(\delta) - Y_0 | d = \delta, r^\delta = s] \\
&= \int (Y_1(\delta) - Y_0) f(Y_1(\delta) - Y_0 | d = \delta, r^\delta = s) \mathrm{d}(Y_1(\delta) - Y_0) \\
&= \int (Y_1(\delta) - Y_0) f(Y_1(\delta) - Y_0 | r^\delta = s) \mathrm{d}(Y_1(\delta) - Y_0) \\
&= E[Y_1(\delta) - Y_0 | r^\delta = s].
\end{aligned}$$

□

References

Abadie, A.: Semiparametric difference-in-differences estimators. Rev. Econ. Stud. **72**(1), 1–19 (2005)

American Academy of Pediatrics Medical Home Initiatives for Children With Special Needs Project Advisory Committee and others. Policy statement: organizational principles to guide and define the child health care system and/or improve the health of all children. Pediatrics **113**(5 Suppl), 1545 (2004)

Andersen, R.M.: Revisiting the behavioral model and access to medical care: does it matter? J. Health Soc. Behav. **36**, 1–10 (1995)

Beal, A., Hernandez, S., Doty, M.: Latino access to the patient-centered medical home. J. Gen. Internal Med. **24**(3), 514 (2009)

Bethell, C.D., Read, D., Brockwood, K.: Using existing population-based data sets to measure the american academy of pediatrics definition of medical home for all children and children with special health care needs. Pediatrics **113**(Supplement 4), 1529–1537 (2004)
Cohen, J.W., Monheit, A.C., Beauregard, K.M., Cohen, S.B., Lefkowitz, D.C., Potter, D., Sommers, J.P., Taylor, A.K., Arnett III, R.H.: The medical expenditure panel survey: a national health information resource. Inquiry **33**, 373–389 (1996)
Conley, T.G., Taber, C.R.: Inference with difference in differences with a small number of policy changes. Rev. Econ. Stat. **93**(1), 113–125 (2011)
Damiano, P.C., Momany, E.T., Tyler, M.C., Penziner, A.J., Lobas, J.G.: Cost of outpatient medical care for children and youth with special health care needs: investigating the impact of the medical home. Pediatrics **118**(4), e1187–e1194 (2006)
Dickens, M.D., Green, J.L., Kohrt, A.E., Pearson, H.A.: The medical home. Pediatrics **90**(5), 774–774 (1992)
Domino, M.E., Humble, C., Lawrence Jr., W.W., Wegner, S.: Enhancing the medical homes model for children with asthma. Med. Care **47**(11), 1113–1120 (2009)
Ferrari, S., Cribari-Neto, F.: Beta regression for modelling rates and proportions. J. Appl. Stat. **31**(7), 799–815 (2004)
Fisher, E.S.: Building a medical neighborhood for the medical home. N. Engl. J. Med. **359**(12), 1202–1205 (2008)
Fu, A.Z., Dow, W.H., Liu, G.G.: Propensity score and difference-in-difference methods: a study of second-generation antidepressant use in patients with bipolar disorder. Health Serv. Outcomes Res. Methodol. **7**(1–2), 23–38 (2007)
Han, B., Yu, H., Friedberg, M.W.: Evaluating the impact of parent-reported medical home status on children's health care utilization, expenditures, and quality: a difference-in-differences analysis with causal inference methods. Health Serv. Res. **52**(2), 786–806 (2017)
Hastie, T.J. (ed.): Generalized additive models. In: Statistical Models in S. Routledge, New York (2017)
Hirano, K., Imbens, G.W.: The propensity score with continuous treatments. In: Shewhart, W.A., Wilks, S.S., Gelman, A., Meng, X. (eds.) Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives (2005)
Homer, C.J., Klatka, K., Romm, D., Kuhlthau, K., Bloom, S., Newacheck, P., Van Cleave, J., Perrin, J.M.: A review of the evidence for the medical home for children with special health care needs. Pediatrics **122**(4), e922–e937 (2008)
Imai, K., Van Dyk, D.A.: Causal inference with general treatment regimes: generalizing the propensity score. J. Am. Stat. Assoc. **99**(467), 854–866 (2004)
Lechner, M.: The estimation of causal effects by difference-in-difference methods. Found. Trends Econom. **4**(3), 165–224 (2011)
Li, Q., Racine, J.S.: Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton (2007)
MacCallum, R.C., Zhang, S., Preacher, K.J., Rucker, D.D.: On the practice of dichotomization of quantitative variables. Psychol. Methods **7**(1), 19 (2002)
Rittenhouse, D.R., Shortell, S.M.: The patient-centered medical home: will it stand the test of health reform? JAMA **301**(19), 2038–2040 (2009)
Rittenhouse, D.R., Thom, D.H., Schmittdiel, J.A.: Developing a policy-relevant research agenda for the patient-centered medical home: a focus on outcomes. J. Gen. Internal Med. **25**(6), 593–600 (2010)
Romaire, M.A., Bell, J.F.: The medical home, preventive care screenings, and counseling for children: evidence from the medical expenditure panel survey. Acad. Pediatrics **10**(5), 338–345 (2010)
Romaire, M.A., Bell, J.F., Grossman, D.C.: Health care use and expenditures associated with access to the medical home for children and youth. Med. Care **50**, 262–269 (2012)
Rosenbaum, P.R.: The consquences of adjustment for a concomitant variable that has been affected by the treatment. J. R. Stat. Soc. Ser. A (General) **147**, 656–666 (1984)
Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika **70**(1), 41–55 (1983)
Sia, C., Tonniges, T.F., Osterhus, E., Taba, S.: History of the medical home concept. Pediatrics **113**(Supplement 4), 1473–1478 (2004)
Slusky, D.J.: Significant placebo results in difference-in-differences analysis: the case of the acas parental mandate. East. Econ. J. **43**(4), 580–603 (2017)
Stange, K.C., Nutting, P.A., Miller, W.L., Jaén, C.R., Crabtree, B.F., Flocke, S.A., Gill, J.M.: Defining and measuring the patient-centered medical home. J. Gen. Internal Med. **25**(6), 601–612 (2010)
Stevens, G.D., Seid, M., Pickering, T.A., Tsai, K.-Y.: National disparities in the quality of a medical home for children. Maternal Child Health J. **14**(4), 580–589 (2010)

Strickland, B., McPherson, M., Weissman, G., Van Dyck, P., Huang, Z.J., Newacheck, P.: Access to the medical home: results of the national survey of children with special health care needs. Pediatrics **113**(Supplement 4), 1485–1492 (2004)

Strickland, B.B., Jones, J.R., Ghandour, R.M., Kogan, M.D., Newacheck, P.W.: The medical home: health care access and impact for children and youth in the United States. Pediatrics **127**, 604–611 (2011)

Stuart, E.A., Huskamp, H.A., Duckworth, K., Simmons, J., Song, Z., Chernew, M.E., Barry, C.L.: Using propensity scores in difference-in-differences models to estimate the effects of a policy change. Health Serv. Outcomes Res. Methodol. **14**(4), 166–182 (2014)

Yang, L., Tsiatis, A.A.: Efficiency study of estimators for a treatment effect in a pretest-posttest trial. Am. Stat. **55**(4), 314–321 (2001)