Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

Identification of the Differential Effect of City-level on the Gini Coefficient of Healthcare Service Delivery in Online Health Community

Hai-Yan Yu[1,2]; Jing-Jing Chen[1]; Jying-Nan Wang[3]; Ya-Ling Chiu[4]; Hang Qiu[5]; Li-Ya Wang[5]

[1] School of Economics and Management, Chongqing University of Posts and Telecommunications, Chongqing, China
[2] Department of Statistics, The Pennsylvania State University, University Park, USA
3 College of International Finance and Trade, Zhejiang Yuexiu University of Foreign Languages, China
[4] College of International Business, Zhejiang Yuexiu University of Foreign Languages, China
[5] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

Hai-Yan Yu email: huy170@psu.edu
Jing-jing Chen email: chenjingjing_361@163.com
Jying-Nan Wang email: jyingnan@gmail.com
Ya-Ling Chiu, email: ylchiu0105@gmail.com
Hang Qiu email: qiuhang@uestc.edu.cn
Li-Ya Wang email: hbigdata@uestc.edu.cn

*Corresponding author
Dr. Hai-Yan Yu
Department of Statistics,
The Pennsylvania State University, University Park,
PA16801, USA
email: huy170@psu.edu
Tel:+1-814-8528826
Fax: +1-814-8528826

Peer-reviewed version available at *Int. J. Environ. Res. Public Health* **2019**, *16*, 2314; doi:10.3390/ijerph16132314

Identification of the Differential Effect of City-level on Gini Coefficient of Healthcare Service Delivery in Online Health Community

ABSTRACT

Inequality of health service for different specialty categories not only occurs in different areas inequality of health service for different specialty categories in the world, but also happens in the online service platform. In the online health community (OHC), health service was often of inequality for different specialty categories, including both online views and medical consultation for offline registered service. Moreover, how the factor city-level impacts the inequality of health service in OHC is still unknown. We designed a causal inference study with data on distributions of serviced patients and online views in over 100 distinct specialty categories on one largest OHC in China. To derive the causal effect of the city-levels (two levels inducing 1 and 0) on the Gini coefficient, we matched the focus cases in cities of rich healthcare resources with the potential control cities. For the Gini coefficient of serviced patients in over 100 specialty categories, the average treatment effect of level-1 cities is 0.470, which is 0.029 higher than that of the matched group. Similarly, for the Gini coefficient of online views, the average treatment effect of Level-1 cities is 0.573, which is 0.016 higher than that of the matched group. For each of the specialty categories, we also estimated the average treatment effect the specialty category's Gini coefficient (SCGini) with the balanced covariates. The results support the argument that the total Gini coefficient of all the doctors in OHC shows that the inequality of health service is still very serious. This study contributes to the development of the theoretically grounded understanding of the causal effect of city-level on the inequality of health service in an online to offline healthcare service setting.

Keywords: Gini coefficient; online health community; medical service delivery; Lorenz curve; inequality of health service; differential Effect

INTRODUCTION

Background

The development of health services takes place not only, of course, within a national but also, third, a world setting[1]. Inequality of health service for different specialty categories not only occurs in different areas inequality of health service for different specialty categories in the world, but also happens in the online service platform, i.e., rural-urban health disparities[2]. More importantly, substantial inequalities remain in the geographical distribution of medical resources (as illustrated in Figure 1); in particular, provinces in western China have the lowest levels of resources[3]. With its potential to mitigate the low levels of medical resources in rural areas, the online health community is no longer merely a site for the public to share physician reviews; it has also become a physician-patient communication platform in China [4].Up to 500,000 people with chronic diseases have used PatientsLikeMe [5], the online healthcare servicer in America, according to a report of the Economist[6]. However, few studies focused on the inequality of the online health service, especially in the inequality of health service for different city-levels. As our previous studies suggested[4], physicians with more past physician online contribution, with higher review ratings, and not coming from cities of rich healthcare resource, were more willing to participate in activities of online health community(OHC). The city-level (or state level) has been studied in other areas, i.e., equity in health[3], public capital[7], and public health[8]. However, the causal effect of the city-level on the inequality of health service is still unknown, especially for the online healthcare community.

Insert Figure 1 about here

As our previous findings[4] suggested that, in various specialty areas, the average levels of physician online contribution were different. Even after the associated characteristics with the potential outcomes are controlled for differences in observed characteristics, there are reasons to believe that the treated and untreated differ in unobservable characteristics[9]. In this scenario, the treated and untreated may not be directly comparable, even after adjusting for observed characteristics. The city-level is an important factor that aggregative the information of geographical distribution and other related resources distribution [10, 11]. Can we still identify and estimate the causal effects of the characteristics (city-level) on the inequality of health service between online views and offline serviced patients for specialty categories? To find a solution to those issues, we design a causal inference study to examine the average treatment effect of the city-level, identifying the difference of inequality of health service between online views and offline serviced patients for specialty categories.

Research Issues

Although the facility of OHC can mitigate the low levels of medical resources in rural areas, few studies focused on the inequality of the online health service, especially in inequality of health service for different specialty categories. The OHC platform can be regarded as an O2O system that provides both communication channels (interaction) for online medical service and records (or feedback) for offline medical service. Although many pieces of research have suggested a long tail phenomenon exists in the online product sale platform, seldom of them simultaneously took both the inequalities in online views and in the offline service (patients' consultation) into consideration. This study attempts to bridge this gap in our knowledge. We examine whether the online health community reduces the inequality of health service for different specialty categories through a retrospective study of the Lorenz curve of doctors' service diversity. Our motivation is trying to answer the following issues: (1) What kind of patterns are the distributions of medical service delivery in distinct specialty categories in the online health community? (2) How does the factor 'city-level' impact the inequality of health service in OHC? (3) How to identify the difference of the response of the Gini coefficient with the treatment variable of the city-level and other confounding variables?

Literature Review

The OHC platform can be regarded as an online-to-offline (O2O) system which provides both communication channels (interaction) for online medical service and records (or feedback) for offline medical service[12]. Among these users, the three types of services with the highest usage rate[4] are medical information inquiry (10.8%), online registration (10.4%), and online consultation services (6.4%). Meanwhile, the online health community can also have the facilities, including guiding the patients to go to hospitals for necessary conditions and multiple virtual visits with their doctors for saving time, travel costs and environmental pollutants[13]. As the posters of Good Doctor (the OHC with the largest population of registered doctors in China) online platform says "based on patients' self-introduction of their conditions, those comments presented by doctors can only be deemed as references rather than direct guidelines for diagnosis and treatment". Since patients often seek information (doctor's outpatient time, their personal introduction and review rating, etc.) of doctors on the OHC, they

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

also revisit the community to give feedback (i.e., rating, online registration, thanks-letters, and gifts) to their doctors after the face to face medical service. Although many pieces of researches have suggested a long tail phenomenon existed in online product sale platforms[14, 15] and online and offline prices similar[16], seldom of them took the inequalities of doctors' service delivery (online or offline service) into consideration.

How far are health-care values and practices shaped by the general structure of inequality in society? On the inequality of the online sales, the study [15] investigated the recommender systems and associated the average influence of the network on each category with the inequality in the distribution of its demand and revenue, quantifying this inequality using the Gini coefficient derived from the category's Lorenz curve. For information cascade [17], they estimated the relationship between a category's Gini coefficient (RevenueGini) and the average PageRank of its books (AvgPageRank) using ordinary least-squares regression. This paper is among the first to measure the concentration of healthcare service delivery in OHC.

The Lorenz curve is a graphical statistic that was first introduced in 1905 as a tool for exhibiting the concentration of wealth in a population [18]. In this context, one can then select any quantile to characterize concentration using a statistic such as 'Y percent of the wealth is owned by X percent of the population.' Alternatively, a summary index of concentration, the Gini coefficient[1], is frequently used. Gini coefficient was originally proposed as methods for studying the concentration of income in a population and had been applied to many problems. Both the Lorenz curve and Gini coefficient have been primarily utilized in the economic and social sciences over the last century. In recent years, however, these methods have also seen applications in other areas such as medical and health services research. For example, the Lorenz curve has been used to describe patterns of drug use. The Lorenz curve and Gini coefficient have also been used to explore the distribution of health professionals in relation to the population distribution of patients. Thus the estimation of both the Lorenz curve and the Gini coefficient involves ranking the units of observation on the basis of some quantity of interest and then estimating cumulative proportions.

A number of approaches are capable of revealing the associative relationship between the outcomes and the related independent variables at a significant statistic level. The causal inference method takes the advantages of non- significant related covariates, which assigns treatment experiments on different units. However, challenges lie in the identification of the causal effect of the treatment variables on the dependent variables. Average treatment effect (ATE) is a measure used to compare treatments (or interventions) in randomized experiments[19]. Although the term 'treatment effect' originated in the medical literature concerned with the causal effects of binary, yes-or-no 'treatments', such as an experimental drug or a new surgical procedure, the term is now used much more generally, such as evaluation of policy interventions and social networks. In a randomized trial (i.e., an experimental study), the average treatment effect can be estimated from a sample using a comparison in mean outcomes for treated and untreated units. However, the ATE is generally understood as a causal parameter (i.e., an estimate or property of a population) that a researcher desires to know, defined without reference to the study design or estimation procedure. Both observational studies and experimental study designs with random assignment may enable one to estimate an ATE in a variety of ways. The difference between these two averages is the ATE, which is an estimate of the central tendency of the distribution of unobservable individual-level treatment effects[20]. If a sample is randomly constituted from a population, the ATE from the sample (the SATE) is also an estimate of the population ATE (or PATE)[21]. The primary goal of causal analysis becomes the investigation of selected effects of a particular cause, rather than the search for all possible causes of a particular outcome along with the comprehensive estimation of all of their relative effects. The rise of the counterfactual model to prominence

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

has increased the popularity of data analysis routines that are most clearly useful for estimating the effects of causes. If a saturated regression model is fit to the data, the lack of overlap in the distribution of covariates will be revealed to the analyst when the regression routine drops the coefficient for the zero cells. However, if a constrained version of the model were fit, such as if covariates were entered as a simple linear term interacted with treatment, the regression would yield seemingly reasonable coefficients. Although using the propensity score to find the region of overlap may not capture all dimensions of the common support (as there may be interior spaces in the joint distribution defined by covariates), subsequent matching is then expected to finish the job [22]. When estimating causal effects using observational data, it is desirable to replicate a randomized experiment as closely as possible by obtaining treatment and control groups with similar covariate distributions. This goal can often be achieved by choosing well-matched samples of the original treatment and control groups, thereby reducing bias due to the covariates. When estimating causal effects using observational data, it is desirable to replicate a randomized experiment as closely as possible by obtaining treatment and control groups with similar covariate distributions. This goal can often be achieved by choosing well-matched samples of the original treatment and control groups, thereby reducing bias due to the covariates [23]. Estimation of average treatment effects under unconfoundedness or exogenous treatment assignment is often hampered by a lack of overlap in the covariate distributions. This lack of overlap can lead to imprecise estimates and can make commonly used estimators sensitive to the choice of specification. In such cases, researchers have often used informal methods for trimming the sample[24].

METHODS

Research Models

In the research design, the treatment variable (city-level) represents the doctor's location status at a specific time. Second, the mean and variance of the number of doctors' articles across the specialty categories, mean in the degree of voted diversity, mean of doctors' review rating and mean in doctors' online contribution as independent variables are considered as the covariates. Based on this framework, we can verify whether doctors' average treatment effect of cities with rich healthcare resources on the inequality of health service is the same for online service (online reviews) and offline service delivery (serviced patients) in different specialty categories.

Gini coefficient[25] was introduced to reveal the distributions (patterns) within categories in a way that is comparable across doctors' specialty areas by calculating the Gini coefficient of each category of the doctors' online service. In applications, the Gini coefficient frequently accompanies a graphical presentation of the Lorenz curve. To comparative analyses of the inequalities in service delivery of online service and in the offline service delivery, we defined two concepts with the Gini coefficient, Gini coefficient of service delivery and Gini coefficient of patient reviews.

The difference of Gini coefficients (of serviced patients or online views) was the dependent variables of interest, and the average number of articles, average breadth of service diversity, average doctor review rating and average doctor online contribution are set as the covariate variables and the city-level (T_i) as the treatment variable. The treatment variable is a binary (0–1) variable, which represents the doctors staying the cities of rich healthcare resources or not at the data acquisition time. The treatment variable is employed to test the average treatment effects of their status. For example, for all the specialty categories, the statistical analysis is designed and conducted for those doctors from cities of rich healthcare resources (i.e., Beijing and Shanghai) $T_i = 1$ and (other cities in China) $T_i = 0$, respectively. The reason why we choose Beijing and Shanghai as the treatment lies in two aspects. First, the healthcare resources in those two cities are much richer than those in other cities or even provinces in

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

China. Approximately 22% of the physicians are working in Beijing or Shanghai, the two largest cities in China. This naturally reflects the relative inequality of the health service of medical resources in large cities. In all the 31 regions, Shanghai ranked first on the perspective of health care institutions (number per 10, 000 km^2), health technical personnel, beds in health care institutions and health investment, while Beijing got the second place[26]. Second, those two cities are often formally treatment as special cases, comparing to any other cities in China. The study [27]revealed that Shanghai with the highest level of economic development had more advanced computed tomography and magnetic resonance imaging machines, and higher government subsidies on these two types of equipment.

The average treatment effects study has many strengths. First, this model will avoid selection bias in the estimation of treatment effects. The bias problem is critical for analyzing the imbalanced data, i.e., the distribution of numbers of owning $T_i = 1$ is not overlapped with that of owning $T_i = 0$. Second, although other independent variables may attract the readers on the topic of this area, the average treatment effects of city-level (T_i) on the inequality of health service attract the most important concerns in the stakeholders of OHC.

The definitions and measurements of all variables are demonstrated in Table 1.

Insert Table 1 about here

With the two dependent variables, we can estimate the doctors' average treatment effect of cities of rich healthcare resources on the inequality of health service in different specialty categories separately and compared them between online views and offline service (patients).

Data Collection

Through web crawler technology, data from the Good Doctor website were collected (on July 26, 2017) and filtered for the purposes of the study. Since 140,344 doctors with personal homepages were commonly considered to be genuinely involved in this OHC. The collected data set contained all the values of this study as well as the doctor's identity document (personal web site) and other de-identified information. The following filtering criterion was set to design an observational retrospective study. (a) Amount of served patients for doctor i's is larger than 0, and the volume of patient online reviews for doctor i's is larger than 0. (b) The number of doctors' articles is larger than 0, number of reviews rating larger than 0, doctor i's online contributions larger than 0 and the number of patients' votes larger than 0.

After filtering, 9,644 samples of doctors remained from the original data set. Meanwhile, 114 specialty categories were filtered from the original 132 categories. The data acquired and filtering process is illustrated in Figure 2.

Insert Figure 2 about here

The filtered samples have the following characteristics. First, our samples were from a large heterogeneous population with diverse backgrounds. The 9, 644 doctors came from 127 different specialty categories, 1,338 different hospitals widely distributed in China. Second, the number of service delivery and the number of patient reviews were collected for the retrieved doctors on the OHC. Although their usage time was different, the corresponding values of the independent variables were also collected during the same period for their usage time. Third, the number of doctors' articles were collected without distinguishing between the original articles and reprinted long articles (not the communication posts with patients). We also collected the doctors' review ratings (regarded as online word-of-mouth) from the stars labeled on the OHC. The average score of these ratings is 2.756 for all the sample data on a scale from 1 (the lowest) to 5 (the highest). Moreover, despite the association with the post articles

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

on the website, the contribution scores of the doctors were also impacted by many other factors, including the post articles communicating with the patients online on the website. The other values we collected were the patient votes, which were different from the doctors' review votes for the word-of-mouth rating and the case records of doctors' accumulated clinical experience. Finally, the values of the location of hospitals were also collected for those doctors clustered in the samples. 2585(26.8%) of all the doctors from Beijing or Shanghai, which are China's two largest developed cities (municipalities). Moreover, 7001 (72.6%) of all the doctors hold the clinic title of the chief or associate chief physician, and 9302(96.4%) of the doctors come from tertiary hospitals. Thus, a causal inference study can be designed with those collected and filtered data samples.

MEASURES

Before examining the OHC platform' effects, it is necessary to distinguish between service delivery and service diversity. Service diversity typically measures how many different services a doctor offers. It is a supply-side measure of breadth. In contrast, we use the diversity of service delivery to describe the concentration of market shares conditional on doctors' assortment decisions[28].

Gini Coefficient: Quantifying the Distribution of Service Inequality

To identify the causal effect of cities of rich healthcare resources on service inequality, our research framework is designed as a retrospective observational study. We aim to investigate the outcomes from two aspects: (a) Gini coefficient of service delivery: offline registered patients, and (b) Gini coefficient of patient reviews: online service. Thus, the dependent variable will be used to reveal the patterns (i.e., inequality phenomena) of the doctors' online service and reveal the relationship between specialty category's Gini coefficient ($SCGini$) and doctors' endorsement on a diversity of specialty categories.

Let $L(p)$ be the Lorenz curve denoting the percentage of the provider's service delivery generated by the lowest $(100 \times p)$% of doctors clustered in the same specialty area during a fixed time period. In our analysis, the Lorenz Curve $L(p)$ is drawn inside a square box with the x-axis being a cumulative percentage of doctors' serviced patients (service delivery) and the y-axis being the cumulative percentage of service delivery for doctors clustered in the same specialty area during a fixed time period. The Lorenz curve of a category's service delivery ranks the services (online medical consultation) in increasing order of the amount of past served patients, then plots the cumulative fraction $L(p)$ of amount of service delivery (served patients) associated with each ascending rank percentile p, where $0 < p < 1$.

This study on the total amount of doctor i's past served patients online will provide evidence to factors of success on which the potential customers select an online doctor and reveal the evolving mechanism of clinical acceptance of telemedicine. SP_i is measured as the cumulative size of the served patients (referring to the doctors' service delivery) in the past. Therefore, the volume of service delivery for doctors clustered in the same specialty area during a fixed time period, SP_j, is calculated by summing the total amount of past served patients SP_i(j) of all the doctors in the same specialty area.

$$SP_j = \sum_{i=1}^{N_j} SP_i(j)\,,$$

where $SP_i(j)$ is the total amount of doctor i's past served patients online in the specialty category (discipline) j, N_j is the number of doctors clustered in the specialty category j.

Thus, the Gini coefficient of distribution of service delivery $SCGini$ is defined by [15]. The Gini coefficient $SCGini$ measures the distributional inequality of the amount of service delivery (serviced patients). $SCGini$ of serviced patients for the specialty category j is defined as

$$SCGini_j(SP) = \frac{Area(SC_j, 45°)}{0.5}, \tag{1}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 May 2019 doi:10.20944/preprints201905.0230.v1

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

$$Area(SP_j, 45°) = \int_0^1 (p - L(p))\, dp,$$

where $Area(SP_j, 45°)$ is the area between the Lorenz Curve of service delivery and a 45-degree line. Thus, $SCGini$ measures how much $L(p)$ deviates from the 45° line, $SCGini \in [0,1]$. A value $SCGini = 0$ reflects diversity (all services have equal service delivery), whereas values near one represent concentration (a small number of services account for most of the service delivery).

When service delivery is perfectly evenly distributed among products, the Lorenz Curve $L(p)$ coincides with a 45-degree line and the Gini Coefficient $SCGini$ equals zero. As the distribution becomes more concentrated, the $L(p)$ curves away from a 45-degree line and the $SCGini$ increases. Thus, $SCGini$ is an aggregate inequality measure and vary anywhere from 0 (perfect equality) to 1 (perfect inequality). Perfect equality in our case illustrates that all the doctors in that category (specialty area) have the same number of service delivery, and perfect inequality illustrates one doctor in the category service all the patients in that specialty area and all other doctors in the category have zero of served patients.

Similar to the definition of $SCGini_j(SP)$, the Gini coefficient $SCGini$ measures the distributional inequality of the number of patient reviews for the doctors in the sociality category.

First, the volume of patient online reviews for doctors clustered in the same specialty area during a fixed time period,OR_j, is calculated by summing the total amount of past online reviews OR_i(j) of all the doctors in the same specialty area.

$$OR_j = \sum_{i=1}^{N_j} PR_i(j),$$

where $PR_i(j)$ is the total amount of doctor i's past patients reviews for doctor i in the specialty category (discipline) j, N_j is the number of doctors clustered in the specialty category j.

$SCGini$ of patient reviews for the specialty category j is defined as

$$SCGini_j(OR) = \frac{Area(OR_j,45°)}{0.5}, \qquad (2)$$

A value $SCGini$ (OR)= 0 reflects diversity (all doctors have equal online reviews), whereas values near one represent concentration (a small number of doctors account for most of the online reviews).

Measure of Doctors' Endorsement

To test this main conjecture, we use the mean and variance of the number of doctors' articles across the specialty categories, mean in the degree of voted diversity, mean of doctors' review rating and mean in doctors' online contribution as independent variables.

(a) mean of the number of Doctors' articles

In this study, we measured the number of doctors' articles through a cumulative count of the articles of each doctor listed on the Good Doctor website. $NDAMea_j$ is measured as the mean of the number of doctors' articles for doctors clustered in the specialty category j.

$$NDAMea_j = \frac{\sum_{i=1}^{N_j} NDA_i(j)}{N_j} \qquad (1)$$

where $NDA_i(j)$ is the number of doctors' articles of the doctor i clustered in the specialty category j, N_j is the number of doctors clustered in the specialty category j.

(b) degree of voted diversity

Given the voting states $(S_i, \# Votes(S_i))$, $S_i = \{S_{i1}, S_{i2}, \ldots, S_{im}\}$ is the vector of doctor i's service specialty labeled by the serviced patients in specialty category j, and $\# Votes(S_i)$ is the corresponding volume vector of their votes. The total amount of doctor i's service specialties labeled by the serviced patients

$$BVS_i(j) = \sum_{j=1}^{m} 1_{(\# Votes(S_i)>0)}$$

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

$BVSMea_j$ is measured as the average breadth of the voted specialties (from patient votes) of all the doctors clustered in specialty category j.

$$\mathrm{BV}SMea_j = \frac{\sum_{i=1}^{N_j} BVS_i(j)}{N_j} \tag{4}$$

where $BVS_i(j)$ is the breadth of the voted specialties (from patient votes) of the doctor i in specialty category j, N_j is the number of doctors clustered in the specialty category j.

(c) mean of the doctors' review rating

In this study, we measured the physicians' ratings in user reviews through the star scores listed on the Good Doctor website. $DRRMea_j$ is measured as the mean of the ratings in user reviews of the doctors clustered in the specialty category j.

$$DRRMea_j = \frac{\sum_{i=1}^{N_j} DRR_i(j)}{N_j} \tag{5}$$

where $DRR_i(j)$ is the ratings in user reviews of the doctor i clustered in the specialty category j, N_j is the number of doctors clustered in the specialty category j.

(d)mean of the doctors' online contribution

Essentially, the existence of online contributions means that members are involved in community-related activities, such as sharing information actively, responding positively to other members' questions, and intuitively interacting with other members[16, 19]. In this study, we measured the physicians' online contribution through the contribution scores listed on the Good Doctor website. There are three principal ways in which the contribution score can change. First, when physicians update their personal information, such as outpatient information and consultation range, in a timely manner, their contribution scores can be increased through the OHC administrator's audit. Second, physicians are encouraged to post medical articles for patients on the website. After the article is referenced by the Good Doctor website, the contribution score is updated. Third, if a physician can answer a patient's question online, his/her contribution score will be increased.

$DOCMea_j$ is measured as the mean of the contribution score for the doctors clustered in the specialty category j.

$$DOCMea_j = \frac{\sum_{i=1}^{N_j} DOC_i(j)}{N_j} \tag{6}$$

where $DOC_i(j)$ is the contribution score for the doctor i clustered in the specialty category j, N_j is the number of doctors clustered in the specialty category j.

Propensity Score: Measure of the Likelihood Being Treated

The propensity score is often employed to reduce the dimensionality of the causal influence problem. The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates [29].

Let $p(X_j)$ be the probability of unit i having been assigned to treatment, and the propensity score was defined as [30]

$$p(X_i) = Pr(T_i|X_i) = E((T_i|X_i)).$$

where $Pr(T_i|X_i)$ is the probability of being assigned to the treatment given X_j, and $E(\cdot)$ is the expectation operator. Here X_i denotes the covariates, i.e., NDA_i, BVS_i, DRR_i, and DOC_i.

Usually, the propensity score was estimated by training the logistic regression.

$$T_i = logit(\beta_0 + \beta_1 NDA_i + \beta_2 BVS_i + \beta_3 DRR_i + \beta_4 DOC_i + \varepsilon_t) \tag{7}$$

where β_0 is the coefficient of the constant term and β_j, j=1,..., 4, are the coefficients of control variables as detailed in Table 1. The error term ε_i obeys normal distribution with mean 0 and variance σ^2.

To achieve a balanced control-treatment case dataset, matching on pre-treatment covariates is one popular method. We match control-treatment cases on pre-treatment

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

covariates with the propensity score. In the matching process, the scalar can be preset for the number of matches which should be found, i.e., the default value 1 is for one-to-one matching. More similar units are more likely to experience more similar trends so the parallel path assumption may be more plausible. Finally, we run the causal effect regression model with the matched data-set.

STATISTICAL ANALYSIS

Having defined our two main variables—service diversity and Gini—we now turn to motivate our empirical analysis. To test the main conjecture of whether doctors' patient votes will affect service usage, it's easy to think about the associative relationship between the covariates and the outcomes. We first fit these data for ten specialty areas by examining how an increase in its influence might enhance or diminish the long tail of medical service demand, rather than fit the size of serviced patients and scale of vote data for the individual doctors. However, we are not only investigating the associative relationship of main effects but also revealing the causal effect of the treatment variable on the outcome, the inequality of health service for different specialty categories.

The above regression model reveals the associative relationship between the main effects. To further reveal the causal effect, the statistical analysis is designed and conducted for those doctors, respectively. The term 'treatment effect' refers to the causal effect of a binary (0–1) variable on an outcome variable of interest. The results are compared for this pair of values in the control variable.

$$ATE(SCGini_j, T) = \mathbb{E}\left(SCGini_j(T = 1) - SCGini_j(T = 0)\right) \tag{8}$$

For all the specialty categories, the $SCGini_j$ consists of two aspects, the specialty category's Gini coefficient of serviced patients and the specialty category's Gini coefficient of online reviews. Those results will be employed to verify the effectiveness of online service and offline service.

In the form of regression [31], the causal effect α can be a model with the linear model:

$$Y_j = \mu + \alpha T_j + \beta X_j + \varepsilon_j$$

where Y_j denotes the outcomes of the j–th units, namely, the Gini coefficient of the j–th categories; T_j the indicator of treatment variable, and X_j the covariates and ε_j the error for unit j.

The coefficient for the treatment indicator α still represents the average treatment effect, but controlling for covariates can improve the efficiency of the estimate. More generally, the regression can control for multiple covariate predictors. As the covariates can be substituted by the observational variables, the causal inference using regression on the treatment variable can be formed as

$$Ln(SCGini_j) = \mu + \alpha T_j + \begin{bmatrix} \beta_1 Ln(NDAMea_j) + \beta_2 Ln(BVSMea_j) + \\ \beta_3 Ln(DRRMea_j) + \beta_4 Ln(DOCMea_j) \end{bmatrix} + \varepsilon_j \tag{9}$$

where Y_j is substituted by $Ln(SCGini_j)$, the logarithm transform of the Gini coefficient of patients or views.

RESULTS

Overlap of the Confounding Variables

With the propensity score matching theory[32], we analyzed the experimental data using logistic regression (10) with one main effect (on treatment) for each covariate. The nearest neighbor method was implemented to achieve control cases to the focus cases.

First, as the literature usually did[23, 33], graphical diagnostics are helpful for quickly assessing the covariate balance. And the histogram distributions of propensity scores in the original and matched groups are also useful for assessing common support. Although the densities of raw treatment and matched treatment cases did not change, those of raw control and match controls took significantly changes. The results show an adequate overlap of the propensity scores, with a good control match for each treatment unit.

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

Second, plots in Figure 3 (left) can show the dots with their size proportional to their weight, which is also useful for weighting or subclassification. Meanwhile, the absolute standardized difference is helpful for comparing the mean of continuous variables between treatment groups, illustrated in Figure 3 (right).

Insert Figure 3 about here

To diagnose the balance of the control-case data, we also compared the focus cases and matched control cases. Table 2 demonstrated the statistics of the selected matched patient characteristics. The results provided empirical evidence that no statistically significant difference exists between those two groups of cases.

Insert Table 2 about here

Lorenz Curve of the Inequality Service

The OHC system associated the average influence of the reputation award on the doctors' serviced patients and online views in each category, with the inequality measure (Gini coefficient) derived from the category's Lorenz curve.

To diagnosis the difference of the cases in those two groups, we examined the data with Welch two sample t-test, as demonstrated in Table 3. Before matching, the means of patients are 1698.112 for the group control and 2680.151 for the focus cases. Since the null hypothesis is rejected, the alternative hypothesis is the true difference in means is not equal to 0. The results show that the mean of focus cases and that of the matched cases is significantly different.

Insert Table 3 about here

With the cases of control-case matching, the Gini coefficients of the empirical experimental data were compared among focus cases, control cases after matching and those before matching. We also compared the Gini of all the cases after matching and those of all the cases before matching, shown as in Table 4. And figure 4 deploys the Lorenz curve of the empirical experimental data on patients and views after matching and before matching.

Insert Table 4 about here

The results in table 4 show three essential facts. First, the number of views shows much higher inequality than that of patients for all the cases, the focus cases and the controls (no matter before matching or after matching). Second, the number of patients of focus cases shows higher inequality than those of controls, but the number of views of focus cases shows lower inequality than those of controls (both before matching and after matching). On patients, the difference of Gini coefficients between focus cases and controls after matching is 0.006, and that between focus cases and controls before matching is 0.031. On views, the difference of Gini coefficients between focus cases and controls after matching is -0.031, and that between focus cases and controls before matching is -0.022. Third, the number of patients of all the cases after matching show higher inequality than that of before matching, but the number of views of all the cases after matching show lower inequality than that of before matching. Moreover, the difference of inequality of health service between online views and offline

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

serviced patients is 0.161 before matching in the 9644 cases, and 0.142 after matching for the 5206 cases.

Insert Figure 4 about here

Causal Effects of City-level on Services Inequality

We first identified the causal effects of cities of rich healthcare resources on online service and offline service with eq. (8). Here we deduced the causal effect with the definition, which is different from the identification process of average treatment effect using regression. This is because the experimental data were provided with complete observations (not counterfactual) on the covariates. For Gini coefficients the specialty categories, 101 entities remained after filtering the NA values in the Gini coefficient table. The distribution of those Gini coefficients was deployed by the Gini coefficient of serviced patients and the views. For the Gini coefficient of serviced patients, 95% quantile of $SCGini_j$(SP) of focus cases is 0.721, which is 0.052 higher than that of the matched group. The 50% quantile of $SCGini_j$(SP) of focus cases is 0. 531, which is 0.025 higher than that of the matched group. And the average treatment effect of level-1 cities (the mean of $SCGini_j$(SP) of focus cases) is 0.470, which is 0.029 higher than that of the matched group. Similarly, for the Gini coefficient of online views, the 95% quantile of $SCGini_j$(OR) of focus cases is 0.840, which is 0.035 higher than that of the matched group. The 50% quantile of $SCGini_j$(OR) of focus cases is 0.642, which is 0.015 higher than that of the matched group. And the average treatment effect of level-1 cities (the mean of $SCGini_j$(OR) of focus cases) is 0.573, which is 0.016 higher than that of the matched group. Moreover, the difference between the average treatment effect of online views and that of offline serviced patients is 0.103 for the 101 specialties categories. In total, the results support the argument that the inequality of health service in level-1 cities is much higher (more serious) than that outside of those level-1 cities for different specialty categories. It also provides evidence that the patients are more likely to be aggregated in level-1 cities, and they are more likely to be served by the doctors.

DISCUSSION

Confounding Effect of the Covariates

Although this paper is designed as a causal inference about the inequality of health service between online views and offline serviced patients for specialty categories, we also analyzed the associative relationship between those covariates and the (Gini) responses of inequality of health service. With the cases before matching, we estimated the correlation between specialty category's Gini coefficients and the other predictors (covariates), including the mean of the number of doctors' articles across the specialty categories, mean in the degree of voted diversity, mean of doctors' review rating and mean in doctors' online contribution. The correlation between the Gini of the coefficient of serviced patients ($SCGini_j$(SP)) and the logarithm of $NDAMea_j$, $BVSMea_j$, $DRRMea_j$, and $DOCMea_j$ are relatively low (0.03, 0.05, 0.17 and 0.10, respectively). Similar results are depicted for the correlation between the logarithm of $SCGini_j$(OR) and the covariates. Based on these correlations, the variation of the response variable ($SCGini_j$(SP) and $SCGini_j$(OR)) may not be mainly explained by the covariates. As the results show, their R-Squared values are very low (R^2=2.5% and R^2=3.8%, respectively), illustrating that the model using ordinary regression is not interpretable to a substantial amount of variance in the dependent variable. The results support our argument that when the associative relationship with the constraints of strong related independent variables is not statistically significant, the causal inference method takes the advantages of non- significant related covariates by assigning treatment

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

experiments on different units. The larger NDAMea, DRRMea, and BVSMea are, the inequality of $SCGini_j$(SP) would be lower. But a larger DOCMea would increase the inequality of $SCGini_j$(SP).

Principal Results

In the original data, the top four specialty categories of doctors' serviced patients are gynecologic and pediatrics, five senses of Chinese traditional medicine (CTM), occupational disease and prosthodontics with their average doctors' serviced patients over 5,000. However, surgery of CTM, plant medicines, infectious medicine of CTM, osteoporosis, and periodontitis are the lowest 5 specialty categories with their average doctors' serviced patients under 800. The Gini coefficient of serviced patients ranges from 0.136 to 0.759 with a mean 0.564, which suggested that the inequality of health service in the online health community is relatively serious for the specialty categories. The total Gini coefficients of all the doctors in OHC are 0.632 for serviced patients and 0.774 for online views after control-case matching, and the Gini coefficient in level-1 cities is much higher (0.006 for serviced patients and -0.031 for online views) than those in the other cities.

Essentially, we should first realize that our empirical results cannot be used to explain all of the doctors' specialties to serve patients but to interpret the causal effect of the city-level on the inequality of health service. As shown in Tables 3 and 4, the causal effect of the city location on Gini coefficient was driven with the matched cases, which are the focus cases in level-1 cities with the potential control cities in the covariates of with the covariates as number of articles, breadth of service diversity, doctor's review rating , doctor's online contribution. Our findings show that, in various specialty areas, the average treatment effect of level-1 cities are different for doctors' specialty categories. Figure 6 indicates that, for the Gini coefficient of serviced patients in over 100 specialty categories, the average treatment effect of level-1 cities is 0.470, which is 0.029 higher than that of the matched group. Similarly, for the Gini coefficient of online views, the average treatment effect of level-1 cities is 0.573, which is 0.016 higher than that of the matched group.

Finally, we make specific recommendations for the OHC managers to reduce the inequality in the distribution of doctors' service delivery among specialty categories based on our findings. For example, the platform should managers should make an effort to reduce the service inequality, improving the referral system and assigning the patients to the matched doctors with the appropriate service diversity. Holding average influence constant, the association between the influence of the specificity diversity and the distributions service delivery was enhanced when the influence was spread more evenly across the doctors in the clinical title, rather than concentrated on a few doctors within the clinical title. For example, when the doctor encountered a not well-experienced disease case (with low votes for a few voted specialties), she/he may directly refuse to provide the online medical consultation service and suggested the patient to referral to another doctor or go to the hospital.

Limitations

Although the difference of inequalities between the units of cases from the level-1 cities and the others in OHC were reflected, more investigations need to be designed on the causality and policy evaluation. In the future, heterogeneous of the results would be considered for distinct groups of doctors who devoted different combinations of online contributions and online attendance. According to the scholarly commonsense of the coauthors, the samples may be grouped by the mean online contributions and online attendance values. As the samples did not completely conform to the standard normal distributions but were nevertheless supported, the mean value was used to represent the entire data set.

First, the number of doctors' articles was collected at a specific time for this study. To further investigate the contribution of doctors' articles, more properties of doctors' articles

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

could be abstracted in the future from the website, including the number of doctors' articles written by themselves, number of doctors' articles copies from others, the average count of words in a doctor' articles, the average times of reviewing for a doctor' articles, etc. Second, the measure of serviced patients used to rank experimental units when estimating the empirical Lorenz curve, and the corresponding Gini coefficient was subject to random error. This error could also lead to an incorrect ranking of experimental units that inevitably results in a curve that exaggerates the degree of diversity (variation) among doctors. Furthermore, all the data were collected from one single OHC, the Good Doctor website. Since the size of each individual doctor' specialty was calculated in the patient voting process from August 26, 2017, to August 27, 2017, there exists a bias in the measurement time interval. Moreover, propensity score matching (PSM)[34] in this study only accounted for observed (and observable) covariates. But the unobserved factors may influence assignment to treatment and outcomes while they cannot be accounted for in the matching procedure[35]. As PSM only controls for observed variables, there can still be hidden biases caused by latent variables after matching[36]. In the worst case, hidden bias may increase because matching on observed variables can unleash bias due to dormant unobserved confounders[37].

CONCLUSIONS

The causal inference method takes the advantages of non- significant related covariates, which assigns treatment experiments on different units. The research design in this paper avoids selection bias in the estimation of treatment effects. The Lorenz curve has been documented for a number of service diversities enrolled in OHC. The distribution of the online service delivery (of patient virtual visits) across the physicians in specialty category j was characterized by a Lorenz curve in which the cumulative proportion of the volume of service delivery was plotted against the cumulative proportion of physicians in the same specialty category in the OHC. We designed a causal inference study with data on distributions of serviced patients and online views in over 100 distinct specialty categories on one largest OHC in China. For the Gini coefficient of serviced patients in over 100 specialty categories, the average treatment effect of level-1 cities is 0.470, which is 0.029 higher than that of the matched group. Similarly, for the Gini coefficient of online views, the average treatment effect of Level-1 cities is 0.573, which is 0.016 higher than that of the matched group. The results support the argument that the total Gini coefficient of all the doctors in OHC shows that the inequality of health service is still very serious. The inequality of health service in level-1 cities is much higher (more serious) than that outside of those level-1 cities for different specialty categories. It also provides evidence that the patients are more likely to be aggregated in level-1 cities, and they are more likely to be served by the doctors.

 doi:10.20944/preprints201905.0230.v1

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

List of Abbreviations

ATE	Average treatment effect
OHC	online health community
SCGini	the specialty category's Gini coefficient
O2O	online-to-offline
SP	served patients
OR	online reviews
NDA	the mean of the number of Doctors' articles
BVS	the breadth of the voted specialties
DRR	the ratings in user reviews of the doctors
DOC	the contribution score for the doctors
CTM	Chinese traditional medicine
PSM	propensity score matching

Declarations

Ethics approval and consent to participate

This study used existing records to conduct a retrospective study. Requirement for individual doctor consent was waived as the study did not impact clinical care and all data were de-identified. None of the data collected for the study are related to private information about the physicians.

Consent for publication

The author(s) declare(s) that the manuscript does not contain any individual person's data. So this paper requires no consent to publish.

Availability of data and material

Our sample data are public on the platform (Haodf.com) for all the users, even without registration. The study is reduplicate with the availability of the public data acquired from the website. We executed statistical tasks in which all-individual information was not involved.

Competing interests

None declared.

Funding

This article presents independent research funded by the National Natural Science Foundation of China (No.71601026) and Chongqing Science and Technology Commission (No. cstc2017zdcy-zdzxX0013).

Authors' contributions

All co-authors are justifiably credited with authorship, according to the authorship criteria. Final approval is given by each co-author. In detail: H.Y. Yu led the research, designed and performed all data analysis and interpretation of results. J.J. Chen, J.N. Wang, H Qiu, and Y.L. Chiu participated in the conception, design and implementation of the

Peer-reviewed version available at Int. J. Environ. Res. Public Health 2019, 16, 2314; doi:10.3390/ijerph16132314

study. J. N. Wang and Y. L. Chiu made substantial contributions to data acquisition. H Y Yu and J.J Chen drafted the manuscript. L.Y. Wang attributed to data analysis and visualization.

Acknowledgements
This work has been supported by the National Natural Science Foundation of China (No.71601026, No. 61601331, No. 71571105), China Postdoctoral Science Foundation (No.2016M602676), Key Research and Development Program of Sichuan Province (No. 2018SZ0114, No. 2019YFS0271) and Chongqing Science and Technology Commission (No. cstc2017zdcy-zdzxX0013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The author(s) also thank(s) the platform (Haodf.com) for the availability of the data.

REFERENCES

1. Balarajan, Y.; Selvaraj, S.; Subramanian, S., Health care and equity in India. *The Lancet* **2011,** 377, (9764), 505-515.
2. Mein Goh, J.; Gao, G.; Agarwal, R., The creation of social value: Can an online health community reduce rural–urban health disparities? *MIS Quarterly* **2016,** 40, (1).
3. Pan, J.; Shallcross, D., Geographic distribution of hospital beds throughout China: a county-level econometric analysis. *International journal for equity in health* **2016,** 15, (1), 179.
4. Wang, J.-N.; Chiu, Y.-L.; Yu, H.; Hsu, Y.-T., Understanding a nonlinear causal relationship between rewards and physicians' contributions in Online Health Care Communities: Longitudinal Study. *Journal of medical Internet research* **2017,** 19, (12), e427.
5. Eichler, G. S.; Cochin, E.; Han, J.; Hu, S.; Vaughan, T. E.; Wicks, P.; Barr, C.; Devenport, J., Exploring concordance of patient-reported information on PatientsLikeMe and medical claims data at the patient level. *Journal of medical Internet research* **2016,** 18, (5), e110.
6. The.Economist.Briefing, China's audacious and inventive new generation of entrepreneur. In The Economist: 2017.
7. Garcia-Mila, T.; McGuire, T. J.; Porter, R., The Effect of Public Capital in State-Level Production Functions Reconsidered. *The Review of Economics and Statistics* **1996,** 78, (1), 177-80.
8. Boyce, J. K.; Klemer, A. R.; Templet, P. H.; Willis, C. E., Power distribution, the environment, and public health: A state-level analysis. *Ecological Economics* **1999,** 29, (1), 127-140.
9. Neumark, D.; Wascher, W., Employment effects of minimum and subminimum wages: panel data on state minimum wage laws. *ILR Review* **1992,** 46, (1), 55-81.
10. Qiu, H.; Yu, H.; Wang, L.; Zhu, X.; Chen, M.; Zhou, L.; Deng, R.; Zhang, Y.; Pu, X.; Pan, J., The burden of overall and cause-specific respiratory morbidity due to ambient air pollution in Sichuan Basin, China: A multi-city time-series analysis. *Environmental research* **2018,** 167, 428-436.
11. Qiu, H.; Tan, K.; Long, F.; Wang, L.; Yu, H.; Deng, R.; Long, H.; Zhang, Y.; Pan, J., The burden of COPD morbidity attributable to the interaction between ambient air pollution and temperature in Chengdu, China. *International journal of*

environmental research and public health **2018,** 15, (3), 492.

12. Heywood, J.; Wicks, P., Systems and methods for encouragement of data submission in online communities. In Google Patents: 2017.
13. Dullet, N. W.; Geraghty, E. M.; Kaufman, T.; Kissee, J. L.; King, J.; Dharmar, M.; Smith, A. C.; Marcin, J. P., Impact of a university-based outpatient telemedicine program on time savings, travel costs, and environmental pollutants. *Value in Health* **2017,** 20, (4), 542-546.
14. Brynjolfsson, E.; Hu, Y.; Simester, D., Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* **2011,** 57, (8), 1373-1386.
15. Oestreicher-Singer, G.; Sundararajan, A., RECOMMENDATION NETWORKS AND THE LONG TAIL OF ELECTRONIC COMMERCE. *MIS Quarterly* **2012,** 36, (1), 65-83.
16. Cavallo, A., Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers. *Nber Working Papers* **2016,** 107, (1), 283-303.
17. Easley, D.; Kleinberg, J., Networks, crowds, and markets: Reasoning about a highly connected world. *Significance* **2012,** 9, 43-44.
18. Moskowitz, C. S.; Seshan, V. E.; Riedel, E. R.; Begg, C. B., Estimating the empirical Lorenz curve and Gini coefficient in the presence of error with nested data. *Statistics in medicine* **2008,** 27, (16), 3191-3208.
19. Imbens, G. W., Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics* **2004,** 86, (1), 4-29.
20. Holland, P. W., Statistics and Causal Inference. *Publications of the American Statistical Association* **1986,** 81, (396), 945-960.
21. Imai, K.; Stuart, K. E. A., Misunderstandings between Experimentalists and Observationalists about Causal Inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*.
22. Morgan, S. L.; Harding, D. J., Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods & Research* **2006,** 35, (1), 3-60.
23. Stuart, E. A., Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* **2010,** 25, (1), 1.
24. Crump, R. K.; Hotz, V. J.; Imbens, G. W.; Mitnik, O. A., Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **2009,** 96, (1), 187-199.
25. Gini, C., Measurement of inequality of incomes. *The Economic Journal* **1921,** 31, (121), 124-126.
26. Liu, W.; Liu, Y.; Twum, P.; Li, S., National equity of health resource allocation in China: data from 2009 to 2013. *International journal for equity in health* **2016,** 15, (1), 68.
27. He, D.; Yu, H.; Chen, Y., Equity in the distribution of CT and MRI in China: a panel analysis. *International journal for equity in health* **2013,** 12, (1), 39.
28. Fleder, D.; Hosanagar, K., Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science* **2009,** 55, (5), 697-712.
29. ROSENBAUM; Paul, R.; RUBIN; Donald, B., The central role of the propoensity score in observational studies for causal effects. *Biometrika* **1983**.
30. Dehejia, R. H.; Wahba, S., Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics* **2002,** 84, (1),

151-161.

31. Greenland, S.; Robins, J. M.; Pearl, J., Confounding and collapsibility in causal inference. *Statistical science* **1999,** 14, (1), 29-46.
32. Ho, D. E.; Imai, K.; King, G.; Stuart, E. A., Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis* **2007,** 15, (3), 199-236.
33. Keller, B.; Tipton, E., Propensity score analysis in R: a software review. *Journal of Educational and Behavioral Statistics* **2016,** 41, (3), 326-348.
34. Becker, T. E.; Atinc, G.; Breaugh, J. A.; Carlson, K. D.; Edwards, J. R.; Spector, P. E., Statistical control in correlational studies: 10 essential recommendations for organizational researchers. *Journal of Organizational Behavior* **2016,** 37, (2), 157-167.
35. Garrido, M. M.; Kelley, A. S.; Paris, J.; Roza, K.; Meier, D. E.; Morrison, R. S.; Aldridge, M. D., Methods for constructing and assessing propensity scores. *Health services research* **2014,** 49, (5), 1701-1720.
36. Cook, T. D.; Campbell, D. T.; Shadish, W., *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston, MA: 2002.
37. Pearl, J., *Causality*. Cambridge university press: 2009.