Supplementary Appendix A: Algorithm 1

Algorithm 1 Causality discover for potential factors affecting the doctor consultation

Input: The doctor set with corresponding profiles and historical interactions.

Output: The DAG consisting of observed doctor attributes.

#Pairwise relationship discovery

Determine the relevant attributes of doctors regarding to the consultation number by analyzing the collected data.

Leverage the determined doctor attributes to construct the objective function of GLasso for estimating the inverse covariance matrix $\hat{\Theta}$ as Eq. (1).

Transform Eq. (1) into the dual optimization function as Eq. (2).

while not reach convergence **do**

    Iteratively update $\hat{\beta}_j$ as Eq. (4).

Obtain the estimated inverse covariance by $\hat{\Theta} = \mathrm{W}^{-1}$.

Generate the data dependency graph G based on $\hat{\Theta}$ to reflect the dependency relationship between multiple features of doctors.

#Causal-effect direction determination

for each variable x_i **do**

    for other variables x_j **do**

        Determine the causal relationship between x_i and x_j by Eq. (5).

Add the causal directions into G and generate DAG.

return The DAG consisting of observed doctor features.

Supplementary Appendix B: Proof of unbiased stratified estimation

Proof. As the theorem discussed in Rubin (1974), when both SUTVA and conditionally random treatment assignment hold, the basic stratified estimation with covariates X_i that take values in a discrete space $X_i \in \mathcal{X}$ is unbiased, and this is manifested as follows:

$$\sqrt{n}(\hat{\tau}_{STRAT} - \tau) \Rightarrow \mathcal{N}(0, V_{STRAT}),$$
$$V_{STRAT} = \mathrm{Var}[\tau(X_i)] + \mathbb{E}\Big[\frac{\sigma^2_{(1)}(X_i)}{e(X_i)} + \frac{\sigma^2_{(0)}(X_i)}{1 - e(X_i)}\Big],$$

where $e(x) = \mathbb{P}[W_i = 1 | X_i = x]$ and $e(x)$ satisfies $0 < e(x) < 1$ for all x. The complicated variables of competency and service level consist of multiple base variables as discrimination conditions, where the base variables jointly form a larger discrete space.

Then, benefited from its excellent property of the asymptotic variance, the usage of $\hat{\tau}_{STRAT}$ does not depend on the number of the groups, i.e. $|\mathcal{X}| = p < \infty$. As a result, the stratified estimation on two complicated variables in this paper is unbiased as classical stratified estimation. Therefore, the proof of Theorem 2 is completed. □

Supplementary Appendix C: The sensitivity analysis of the E-value

The analysis of the constructed DAGs in causality discovery leads to the proposal of Assumption 1, which elucidates a specific treatment-outcome relationship between service quality and online consultation orders, attributing it to the concern of popularity bias. Inevitably, while the causality discovery incorporates a set of measured confounders, the presence of hidden unmeasured confounders that potentially impact the proposed causal effect is imperatively considered. Inspired by the concept of the E-value in VanderWeele and Ding (2017), it employs a quantitative approach for sensitivity analysis to explore the influence of unmeasured confounders on the association. In other word, the E-value refers to the risk ratio that completely eliminates unmeasured confounding effects under the condition of controlling measured confounding factors (Cheng et al. 2024, Lonati and Wulff 2024, VanderWeele and Ding 2017).

In this study, the exposure denoted as E represents the professional competency y_c, the outcome denoted as D signifies the status of order quantity y_o (where $y_o = 1$ if the order quantity exceeds the average doctor's order, and $y_o = 0$ otherwise), and the measured confounding factor denoted as C corresponds to the service quality y_s. As detailed in Table F.3, the attributes *Gender* and *Waiting*, labeled as U $(U = 0, 1, \cdots, k, \cdots, K-1)$, are identified as potential unmeasured confounding factors. To begin the analysis, the risk ratio (RR) between the exposure and the outcome in the observed data $RR^{obs}_{ED|c}$ is computed by controlling the measured confounders as follows:

$$RR^{obs}_{ED|c} = \frac{P(D=1|E=1, C=c)}{P(D=1|E=0, C=c)}. \tag{1}$$

And RR between unmeasured confounding factors and exposure when $U = k$ is determined as the maximal association impact, i.e. $RR_{EU|c} = max_k\ RR_{EU,k|c}$, where $RR_{EU,k|c} = \frac{P(U=k|E=1,C=c)}{P(U=k|E=0,C=c)}$. Then, the association impact between the unmeasured confounders and outcome $RR_{UD|c}$ is necessary to determine as follows:

$$RR_{UD|c} = max(RR_{UD|E=1,c}, RR_{UD|E=0,c}), \tag{2}$$

where $RR_{UD|E=1,c} = \frac{max_k P(D=1|E=0,C=c,U=k)}{min_k P(D=1|E=0,C=c,U=k)}$ and $RR_{UD|E=0,c} = \frac{max_k P(D=1|E=1,C=c,U=k)}{min_k P(D=1|E=1,C=c,U=k)}$, respectively measuring the impact in the exposure group and non-exposure group. Finally, the true causal RR can be inferred as follows:

$$RR^{true}_{ED|c} = \frac{\Sigma_{k=0}^{K-1} P(D=1 \mid E=1, C=c, U=k) P(U=k \mid C=c)}{\Sigma_{k=0}^{K-1} P(D=1 \mid E=0, C=c, U=k) P(U=k \mid C=c)}. \tag{3}$$

Based on the whole data in this study, we can obtain $RR^{true}_{ED|c} = 1.614$, $RR^{obs}_{ED|c} = 1.087$, $RR_{EU|c} = 5.768$ and $RR_{UD|c} = 5.333$. Following to the discrimination formula in VanderWeele and Ding (2017) under the condition of $RR^{true}_{ED|c} > 1$, the association impact of the gender and waiting time insufficiently eliminates the causal effect between competency and orders with the service level controlled when the following inequality holds:

$$RR^{true}_{ED|c} \geq RR^{obs}_{ED|c} / \frac{RR_{EU|c} RR_{UD|c}}{RR_{EU|c} + RR_{UD|c} - 1}. \tag{4}$$

From the computation of the above inequality, it holds in this context and demonstrates the feasibility of Assumption 1. Moreover, the E-value can be computed as follows:

$$E - value = RR^{obs}_{ED|c} + \sqrt{RR^{obs}_{ED|c} \times (RR^{obs}_{ED|c} - 1)}. \tag{5}$$

$E - value = 2.609$ can be derived, which indicates that the observed RR can be completely explained by the unmeasured confounders when both $RR_{EU|c}$ and $RR_{UD|c}$ is more than 2.609 and the measured confounders have been controlled.

Supplementary Appendix D: Algorithm 2

Algorithm 2 Pre-training process for competency representation of doctors

Input: The specialty texts of doctors, the nearest neighbors K, the number of negative edges M and relevant hyper-parameter settings.

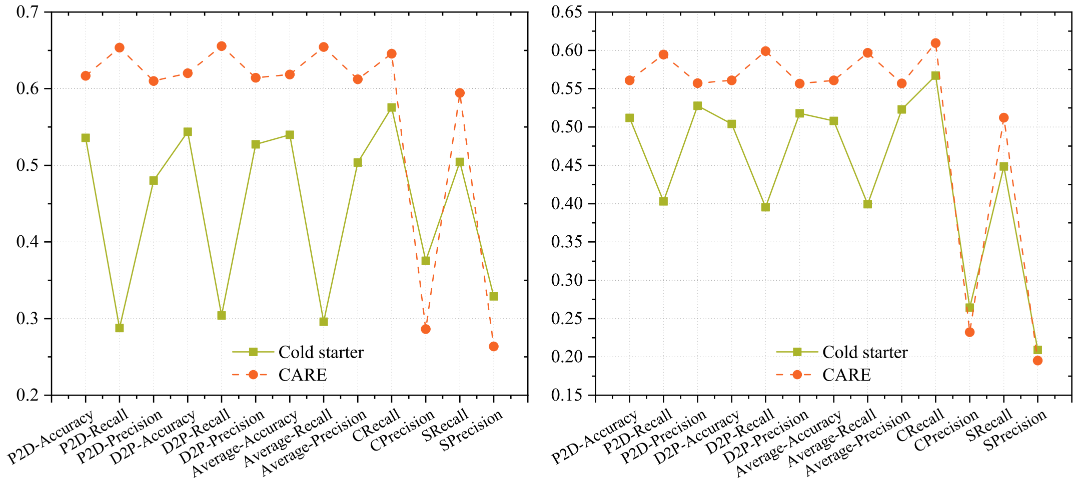
Output: The competency embedding $\hat{X}$ of doctors.

```
#Graph construction
for each doctor i ∈ D do
    Segment the specialty text into word level by pkuseg.
Adopt word2vec method to train word segmentation for doctor embedding R^{D×h_d}.
for each doctor i ∈ D do
    Calculate the similarity of i with other doctors.
    Select the doctors with top-k similarity to link as edges.
The competency graph G = {D, E} where nodes are doctors is constructed.
#Competency pre-training
Initialize the initial embedding of doctors X ∈ R^{D×h_c}.
Incorporate the information of neighbors and update embedding by GraphTransformer in Eq. (8).
for each link e(u,v) ∈ E do
    Generate M negative edges for u.
    Generate M negative edges for v.
Optimize the pre-train loss function of Eq. (9) to determine final competency embedding of doctors.
return The final competency embedding of doctors X̂ ∈ R^{D×h_c}.
```

Supplementary Appendix E: Case study on doctor-patient matching

Beyond the above experimental results, we further provide some detailed case study for clarification of explainable CARE in this section, as well as its extension on cold start issue. As we known, the concern of cold start usually occurs in information systems and is a part of platform governance (Liang et al. 2024), reflecting as the unavailable user profiles of new entrants and subsequent poor model performance. In this section, to validate the effectiveness of CARE, we select the doctors whose historical consultation is less than 50 as cold starters, and compare their evaluation to the results of all doctors. The comparison on two datasets is depicted as Fig. E.1.

In Fig. E.1, data insufficiency of new entrants inevitably impedes the matching prediction with inferior performance in most of metrics when compared with CARE. The heterogeneity related to different diseases can be clearly seen that Recall declines more strongly in Skin disease than that of Dizziness. This demonstrates that the dermatologists are easily affected by cold start, and patients prefer to select experienced doctors. Besides, Recall changes most compared with Accuracy and Precision, illustrating that more pairs are predicted to not match for new entrants. Moreover, some specific positive and negative pairs that are correctly predicted are presented in Tables E.1 and E.2, where each table contains two cases.



(a) Skin disease (b) Dizziness

Figure E.1 The matching performance for doctors who are under cold start condition

Table E.1 A sample of positive matching on Skin disease and Dizziness of Haodf dataset

CASE 1-Skin disease				
Doctor Information				
Doctor ID	Province	Professional Competence	Specialty	Historical consultation
8347418324	Beijing	Grade A tertiary hospital Pediatric dermatology physician-in-charge Article: 0	Dermatitis, eczema, atopic dermatitis, papular urticaria, urticaria and other allergic skin diseases, psoriasis in children, alopecia areata and other common diseases, hemangioma in children, port wine stains, coffee spots and nevus of Ota	Popularity: 3.5 Consultation: 2201 Viewed: 252733 Patients after consultation: 11 Vote: 6 Thanks: 5 Gift: 30
Patient Information				
Patient ID	Disease	Province	Disease detail	Product type
9573301640	Skin disease	Hebei	Case level: 1 Time: 2021/7/22 9:23	Type 6
9573301640	Skin disease	Hebei	Case level: 1 Time: 2021/12/16 21:15	Type 5
CASE 2-Dizziness				
Doctor Information				
Doctor ID	Province	Professional Competence	Specialty	Historical consultation
1049803759	Guangzhou Guangdong	Grade A tertiary hospital Neurology associate chief physician Article: 16	Motor neuron disease (also known as frostbite, amyotrophic lateral sclerosis), mitochondrial disease, myasthenia gravis, congenital myopathy, muscular dystrophy, Guillain Barré syndrome, Charcot Marie Tooth disease, amyloidogenic peripheral neuropathy and other myopathies, peripheral neuropathy and nervous system suspicious rare diseases Available patients: muscle weakness, muscular atrophy, muscle jump, electromyography suggesting neurogenic or myogenic lesions	Popularity: 4.5 Consultation: 727 Viewed: 167356 Patients after consultation: 320 Vote: 97 Thanks: 86 Gift: 85
Patient Information				
Patient ID	Disease	Province	Disease detail	Product type
10405351869	Dizziness	Gansu	Case level: 2 Time: 2021/9/8 12:40	Type 6

Table E.2 compares the basic information of negative doctors and real-matching doctors to illustrate the potential reason why real-matching doctors are chosen by patients. There are three obvious trends in online medical consultation: (1) Patients living in economy-developed regions tend to select the doctors of nearby grade A tertiary hospitals.

Table E.2 A sample of negative matching on Skin disease and Dizziness of Haodf dataset

<table>
<tr><th colspan="5">CASE 1-Skin disease</th></tr>
<tr><td colspan="5">Information of Doctor to be Matched</td></tr>
<tr><td>Doctor ID</td><td>Province</td><td>Professional Competence</td><td>Specialty</td><td>Historical consultation</td></tr>
<tr><td>5732433124</td><td>Jiangsu
Nanjing</td><td>Grade A tertiary hospital
Pediatric dermatology
associate chief physician
Article: 0</td><td>He works in the Department of Dermatology,
children’s Hospital of Nanjing Medical University.
He has been engaged in pediatric dermatology
for more than ten years and
has rich clinical experience in common skin diseases in children.
Such as urticaria, atopic dermatitis, contact dermatitis,
drug eruption and other allergic diseases,
varicella, herpes simplex and other infectious diseases,
as well as other diseases such as hemangioma,
vitiligo and so on.</td><td>Popularity: 3.5
Consultation: 69
Viewed: 19704
Patients after consultation: 3
Vote: 2
Thanks: 2
Gift: 7</td></tr>
<tr><td colspan="5">Real-matching Doctor Information</td></tr>
<tr><td>Doctor ID</td><td>Province</td><td>Professional Competence</td><td>Specialty</td><td>Historical consultation</td></tr>
<tr><td>6962350694</td><td>Jiangsu
Nanjing</td><td>Grade A tertiary hospital
Dermatology
associate chief physician
Article: 5</td><td>1: Surgery and photodynamic therapy of skin malignant tumors
2: comprehensive treatment of scars and keloids
3: double eyelids, pouch, face rejuvenation, fat filling
4: fine resection and suture treatment of benign tumors
such as nevus, lipoma, sebaceous nevus
5: skin grafting for vitiligo, circumcision, axillary osmidrosis,
paronychia, condyloma acuminatum and other clinical treatments</td><td>Popularity: 4.5
Consultation: 1319
Viewed: 182158
Patients after consultation: 268
Vote: 37
Thanks: 35
Gift: 55</td></tr>
<tr><th colspan="5">CASE 2-Dizziness</th></tr>
<tr><td colspan="5">Information of Doctor to be Matched</td></tr>
<tr><td>Doctor ID</td><td>Province</td><td>Professional Competence</td><td>Specialty</td><td>Historical consultation</td></tr>
<tr><td>1048723848</td><td>Beijing</td><td>Grade A tertiary hospital
Gastroenterology
chief physician
Article: 0</td><td>Diagnosis, treatment, and
prevention of Helicobacter pylori infection;
Especially the eradication of refractory Helicobacter pylori;
Chronic gastritis, peptic ulcer, indigestion,
gastroesophageal reflux disease, chronic diarrhea, constipation,
irritable bowel syndrome, fiberoptic gastroscopy
and colonoscopy examination and partial endoscopic treatment</td><td>Popularity: 3.7
Consultation: 639
Viewed: 117694
Patients after consultation: 7
Vote: 0
Thanks: 0
Gift: 10</td></tr>
<tr><td colspan="5">Real-matching Doctor Information</td></tr>
<tr><td>Doctor ID</td><td>Province</td><td>Professional Competence</td><td>Specialty</td><td>Historical consultation</td></tr>
<tr><td>8175699939</td><td>Beijing</td><td>Grade A tertiary hospital
Neurology
associate chief physician
Article: 26</td><td>Specific symptoms of peripheral neuropathy
and cerebrovascular disease (cerebral atrophy, cerebral infarction,
stroke, cerebral blood supply insufficiency, atherosclerosis)
treated with integrated traditional Chinese and western medicine:
1. prevention of cerebrovascular disease and treatment of sequelae
2. dizziness, headache and memory impairment
related to cerebral arteriosclerosis
3. numbness, pain and weakness of limbs</td><td>Popularity: 4.2
Consultation: 26
Viewed: 92708
Patients after consultation: 57
Vote: 64
Thanks: 53
Gift: 14</td></tr>
</table>

It is convenient to offline conduct outpatient examination, and receive long-term drug treatment. (2) Online medical platforms are beneficial to citizens living in underdeveloped areas with scarce medical resources. These patients can be accessible to high-quality medical services as Case 2 of Table E.1, which fully highlights the strength of online medicine. Moreover, the doctors who are both capable and willing can offer high-quality medical consultations, motivated by self-actualization and monetary rewards. This not only allows for the realization of the scientific allocation of medical resources but also promotes the efficient operation of the medical service system. (3) The weak reciprocity of online medical platforms is also empirically analyzed in real datasets. From these cases, the selection of doctors is largely influenced by the filtering and display manipulated by platforms. In the presence of filtering, the patients incorporate their preference into the selection process to seek more suitable doctors as shown in Table E.2, such as doctor competency, experience and registration fee. By pre-training designed based on the specialty text and competency graph, the CARE method is useful to identify medical professional competence of the doctors to achieve effective matching. In short, the above-mentioned results have revealed some meaningful insights for platforms to design and adjust their two-sided matching mechanism, improving reciprocal performance and satisfaction of the users.

Supplementary Appendix F: Features of empirical data

The performance evaluation of CARE method is examined in two genuine datasets corresponding to two diseases, where they are extracted from a large-scale dataset of Haodf online healthcare platform. Both the two datasets encapsulates dual facets: basic profile of doctors, and matching pair instance, whose feature lists are presented in Tables F.3 and F.4. In these feature lists, the features with abbreviations, data types, and their definitions are described in accordance to the data detail of Haodf platform [1]. Note that three features, satisfaction, waiting time and stopped outpatient information of the most doctors is missing within dataset in Table F.3.

Table F.3 Feature list of basic doctor profile

Feature (Abbreviation)	Type	Definition
ID	Numerical	The desensitized ID of the doctor.
Gender	Categorical	The gender of the doctor.
Title (DT)	Categorical	The professional title of the doctor.
Hospital	Categorical	The hospital that the doctor works in.
Department	Categorical	The department of the doctor in the hospital.
Hospital_Level (HL)	Categorical	The grade of the hospital under the Chinese system.
Province	Categorical	The province where the hospital is located.
City	Categorical	The city where the hospital is located.
Specialty	Text	Diseases that doctors excel in diagnosing and treating.
Popularity	Numerical	The recommendation popularity calculated by platform.
Satisfaction	Numerical	Satisfaction with online services on the doctor.
Consultation_Num (HC)	Numerical	The number of historical online consultation of the doctor.
View	Numerical	The number of visits to the doctor's homepage.
Article_Num (AN)	Numerical	The number of articles related to diseases published by doctors on the platform.
After_Consultation (PC)	Numerical	The number of patients that re-consults after consultation.
Vote (PV), Thanks (TL) Gift (GF)	Numerical	The number of votes, thanks letter, and gift operated by patients to the doctor.
Waiting	Numerical	The waiting time to receive the consultation of the doctor.
Online, Registration	Time	The last online time and the registration time of the doctor.
Stopped, Outpatient	Time	The stopped outpatient information and the outpatient schedule of the doctor.

Table F.4 Feature list of matching pair instance

Feature	Type	Definition
Patient	Numerical	The desensitized ID of the patient.
Disease	Categorical	The disease of the patient diagnosed by the doctor.
Pat_Province	Categorical	The province of the patient living in.
Severity	Categorical	The severity level of the patient instance.
Pay_Time	Time	The time to pay the order.
Type	Categorical	The product type of the order.
Doc	Numerical	The desensitized ID of the doctor.

Supplementary Appendix G: Discrimination criteria of doctor features to conduct RCT

In the estimation of ATE, difference-in-means method requires the randomized control on data instances, which sets the treat group and control group as $[0, 1]$ of features. By observing the potential values within each feature, in line with the domain knowledge on medicine, we utilize the discrimination condition to transform important features for conducting RCT as shown in Table G.5. That is, if the feature of one doctor satisfies the discrimination criteria of treatment group, it is $y_\bullet = 1$ where $\bullet = \{DT, HL, AN\}$.

Table G.5 The discrimination criteria in RCT

Feature	Group	Discrimination condition
Title (DT)	Treatment group	Chief nurse, Chief physician, Chief rehabilitation specialist, Chief pharmacist, Chief technician, Chief inspection physician, Psychotherapist
	Control group	Other doctor titles
Hospital_Level (HL)	Treatment group	Grade A tertiary hospital, Tertiary hospital
	Control group	Other hospital levels
Article_Num (AN)	Treatment group	More than the average of article number
	Control group	Less than the average of article number

Supplementary Appendix H: Complexity analysis of CARE

This section discusses the time complexity of our proposed CARE method to enhance its feasibility in practice. In CARE method, the causality analysis is separated from the real time two-sided matching process so that its implementation time is not included into total time. We just need to analyze the time consumption of pre-training part and weak reciprocal matching.

Pre-training part: For graph construction, the embedding of doctor specialty text via word2vec framework is $\mathcal{O}(kv)$ with sliding time window size k and vocabulary size v. And the usage of KNN nearly consumes the time $\mathcal{O}(|\mathcal{D}|^2)$ in similarity calculation and ranking. Given L layers to repeat GraphTransformer message passing, the complexity is $\mathcal{O}(L(|\mathcal{D}| + M \cdot |\mathcal{E}|) + |\mathcal{E}| \cdot h_c \cdot M)$ with the number of negative sampling M and competency embedding dimension h_c.

Weak reciprocal matching: This process is defined as a binary classification task through two layes MLP, respectively resulting in the complexity of $\mathcal{O}(T \cdot |\mathcal{M}_{obs}| \cdot h_{hidden}(h_{f1} + 1)$ and $\mathcal{O}(T \cdot |\mathcal{M}_{obs}| \cdot h_{hidden}(h_{f2} + 1)$ for P2D and D2P direction, where h_{hidden} is the hidden dimension of MLP. As reciprocity considered in two directions, $h_{f1} = h_0 + h_c + h_s$ and $h_{f2} = h_0 + h_c$ correspond to factors used in prediction.

Hence, the overall time complexity is summarized as follows:

$$\mathcal{O}(kv + |\mathcal{D}|^2 + L(|\mathcal{D}| + M \cdot |\mathcal{E}|) + |\mathcal{E}| \cdot h_c \cdot M) + T \cdot |\mathcal{M}_{obs}| \cdot h_{hidden}(h_{f1} + h_{f2} + 2)). \tag{6}$$

References

Cheng D, Li J, Liu L, Liu J, Le TD (2024) Data-driven causal effect estimation based on graphical causal modelling: A survey. *ACM Computing Surveys* 56(5):1–37.

Liang C, Hong Y, Gu B (2024) Monitoring and the cold start problem in digital platforms: Theory and evidence from online labor markets. *Information Systems Research* DOI: 10.1287/isre.2021.0146.

Lonati S, Wulff JN (2024) Hic sunt dracones: On the risks of comparing the itcv with control variable correlations. *Journal of Management* DOI: 10.1177/01492063241293126.

Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66:688–701.

VanderWeele TJ, Ding P (2017) Sensitivity analysis in observational research: Introducing the e-value. *Annals of Internal Medicine* 167(4):268–274.