Evolutionary Game Analysis of Health Big Data Open Utilization Considering Trust Gains and Losses

Authors: Wang Dandan, Xie Shicheng, Cheng Yingying, Dandan Wang

Date: 2024-08-30T00:00:00+00:00

Abstract

【Purpose/Significance】Trust among relevant stakeholders has become a key factor determining the effectiveness of health big data open utilization. Based on the process of health big data open utilization and considering trust profit and loss factors, this study investigates the influence mechanism of trust on the behaviors of various stakeholders, providing forward-looking references for improving the effectiveness of health big data open utilization in China. 【Method/Process】Focusing on the decision-making behaviors of data platforms, data users, and patients under the consideration of trust profit and loss factors, and based on analysis of the open utilization process, this study establishes a tripartite evolutionary game model, conducts simulation analysis through MATLAB, and finally proposes countermeasures and suggestions based on equilibrium point analysis and simulation results. 【Results/Conclusions】Obtaining patient trust is the fundamental motivation for data users to utilize data transparently; enhancing the security of the research environment can fundamentally solve trust issues; strengthening supervision and increasing the trust profit and loss of data users helps achieve a benign state of open utilization; while improving patients' trust level in data platforms, it is more important to strengthen constraints on data user behavior. China may consider building a state-led health big data open utilization platform, strengthening the construction of trusted research environments and actively promoting their application, implementing a data usage registration system and establishing a standard system, and effectively guiding public participation in the collaborative governance process of health big data.

Full Text

Preamble

Evolutionary Game Analysis of Open Utilization of Healthcare Big Data Considering Trust Gains and Losses

Wang Dandan, Xie Shicheng, Cheng Yingying
Business School, Henan University of Science and Technology, Luoyang 471023

Abstract

[**Purpose/Significance**] Trust among relevant stakeholders has become a key determinant of the effectiveness of healthcare big data open utilization. Based on the process of healthcare big data open utilization and considering trust gain and loss factors, this study investigates the influence mechanism of trust on the behavior of various parties, providing forward-looking references for improving the effectiveness of healthcare big data open utilization in China. [**Methods/Process**] Focusing on the decision-making behaviors of data platforms, users, and patients under considerations of trust gains and losses, this study establishes a tripartite evolutionary game model based on analysis of the open utilization process. Simulation analysis is conducted using MATLAB, and countermeasures and recommendations are proposed based on equilibrium point analysis and simulation results. [**Results/Conclusion**] Gaining patient trust is the fundamental motivation for users to transparently utilize data. Enhancing research environment security can fundamentally solve trust issues. Strengthening regulation and increasing the trust gains and losses for users contribute to achieving a virtuous state of open utilization. While improving patient trust levels in data platforms, it is even more important to strengthen constraints on user behavior. China should consider establishing a nationally-led healthcare big data open utilization platform, strengthening the construction of trusted research environments and actively promoting their application, implementing a data use registration system and establishing a standard system, and effectively guiding public participation in the collaborative governance of healthcare big data.

Keywords: Healthcare big data; Trust; Transparency; Open utilization; Evolutionary game

Healthcare big data refers to health and medical-related data generated during disease prevention, treatment, and health management processes [1]. The United States pioneered the open data movement in 2009, launching a government data open platform that extended to the health and medical data domain, establishing the HealthData.gov platform. The Centers for Disease Control and Prevention, Centers for Medicare & Medicaid Services, and Food and Drug Administration have opened structured data on medical insurance and subsidies, health expenditures, disease control, and public health. The United Kingdom built the National Health Service (NHS) system, opening demographic statistics and health management and quality monitoring data. Relying on the NHS system, the UK has been at the forefront of healthcare big data open utilization worldwide, launching the care.data project in 2013. However, due to improper data use that led to loss of public trust, the project was terminated in July 2016. In 2018, the UK established Health Data Research UK (HDR UK), dedicated to integrating health data across the UK for scientific research purposes, which continues to operate today [2].

Domestic scholars have conducted research on healthcare big data from perspectives including rights allocation [3-5], privacy protection [6], obstacles and governance in open utilization [7], quality evaluation [8], and value realization paths [9-10]. International scholars primarily discuss technological advancements, legal improvements, and ethical guidance, with technological improvements being the most common focus, such as access security [11-12], storage security [13], network attack prevention [14], sharing security [15], processing security [16], transmission security [17], transaction security [18], and building trusted research environments [19]. The importance of privacy, transparency, accountability, fairness, and consent for healthcare big data is fully recognized [20-22], and the construction of trust among stakeholders [23-25] has received significant attention. China attaches great importance to the open utilization of healthcare big data. The "Guiding Opinions on Promoting and Regulating the Development and Application of Healthcare Big Data" emphasizes promoting data sharing and establishing and improving open systems. The "National Healthcare Big Data Standards, Security, and Service Management Measures (Trial)" establishes a state-led approach to data management and development.

This work is supported by the National Social Science Fund of China project titled "Research on the Open Utilization and Collaborative Governance of Healthcare Big Data" (Grant No. 23BTQ068).

Author(s): Wang Dandan, Vice Dean of Business School, Henan University of Science and Technology, Professor, PhD, Doctoral Supervisor, Email: wangdan430806@163.com; Xie Shicheng, Doctoral Candidate; Cheng Yingying, Lecturer, PhD.

In light of this, this paper analyzes the UK's experience in healthcare big data open utilization, constructs a process model for healthcare big data open utilization, and considers the mutual influences among data users' transparency in data use, regulatory intensity, and patient participation behavior under established privacy protection technologies, institutions, and infrastructure levels. A tripartite evolutionary game model is established to study the game behaviors of various stakeholders and propose countermeasures and recommendations.

1.1 Review of the UK's Healthcare Big Data Open Utilization Journey

The NHS aggregates life-course data for the entire UK population from birth to death, providing a rich data foundation for health research in the UK. The

UK government has actively promoted the open utilization of healthcare big data. Under the leadership and guidance of the NHS, the Health and Social Care Information Centre (HSCIC) launched the flagship healthcare big data platform care.data in 2013, collecting the most detailed electronic health data from public medical institutions and general practitioner clinics nationwide. The project aimed to better understand patients, develop drugs and treatments, predict public health and disease trends, monitor drug and treatment safety, and compare medical quality across regions [26]. The project allowed the NHS and eligible private companies to use data for these public purposes.

In the care.data project, the HSCIC, guided by the NHS, served as the data collector and data use regulator. Data was made available to research institutions and some eligible private companies for a fee, with usage restricted to public interest-related research projects. However, after the NHS sold anonymized data to health organizations, research universities, hospital trusts, pharmaceutical companies, and other private companies, although anonymization was performed and usage was agreed to be for public interest, numerous incidents of personal information leakage occurred. This information was used by insurance companies, marketing firms, and others for commercial purposes, seriously infringing on patients' privacy rights and autonomy. Although the NHS subsequently strengthened regulation, due to the negative externalities arising from the inherent replicability of data, privacy breaches and data misuse did not stop. A large number of the public chose to opt out of the project, leading to the termination of the care.data project in July 2016 [27].

The care.data practice demonstrates several key lessons. First, encryption technology cannot guarantee absolute data security. Anonymized data is not absolutely secure and remains potentially re-identifiable, requiring continued security protection. Second, when absolute data security cannot be guaranteed, transparency in the data circulation process and patient consent are fundamental prerequisites for ensuring public trust. For example: "Should patient consent be obtained when family doctors provide data to HSCIC and when HSCIC opens data to external parties?" and "Is data used for high-quality scientific research and other public purposes rather than commercial purposes?" Only by addressing these issues can the trust of the public and healthcare professionals be won and good data supply be ensured [28]. In 2018, the UK established HDR UK, dedicated to using UK health and medical data for ethical purposes [2]. HDR UK primarily builds public confidence and ensures public trust through Trusted Research Environments (TRE), Data Use Registers (DUR), and Patient and Public Involvement and Engagement (PPIE) systems.

1.2 Evolutionary Game Theory and Trust Gains and Losses

The open utilization of healthcare big data is a dynamic process of strategic choices among multiple stakeholders, involving the public and patients, data custodians (medical institutions, trusts, storage centers, etc.), data users (medical research institutions, research institutes, pharmaceutical companies, etc.), data regulators (government agencies, data platforms, etc.), and technical service providers. Under conditions of bounded rationality, these stakeholders continuously trial-and-error, evolve, and make choices. Some scholars have used evolutionary game methods to study the game process among medical institutions, technology support enterprises, and the government [29], constructing two-party evolutionary game models from the perspective of privacy protection between patients and medical service institutions, as well as three-party evolutionary game models among patients, medical service institutions, and the government to explore the sharing willingness of various stakeholders [30-31].

Trust is a relationship in which one party has reasonable beliefs in another party's competence and goodwill, or has explicit commitments to perform certain actions [32-34]. Regarding healthcare big data, public trust manifests in several ways: (1) Trust in the state and government. The public may have lower willingness to share due to distrust in their country's medical research level [35] or political system [36-37]. (2) Trust in researchers and doctors. In healthcare systems, researchers' recommendations have been proven to increase the likelihood of public participation in research [38-40]. (3) Trust in healthcare systems and user institutions. The public's willingness to share sensitive information with an entity varies according to their level of trust and familiarity [41-43]. (4) Trust in regulatory agencies. The public pays attention to the reputation of individuals or organizations that regulate data [44-46]. This paper conceptualizes trust loss (or gain) as the expected actual benefit loss (or gain) that one party generates for another when it perceives a decrease (or increase) in the other party's competence and goodwill and takes action accordingly. For example, the trust gain obtained by a data platform through strong regulation—that is, when the platform's strong regulatory behavior causes patients to perceive an increase in the platform's competence and goodwill—leads patients to actively participate in data sharing, thereby generating expected benefits for the data platform.

2.1 Problem Description

The process of healthcare big data open utilization is illustrated in Figure 1 [Figure 1: see original paper]. In the data collection phase, patients share data with medical institutions (hospitals, community clinics, individual doctors, etc.) due to treatment needs. Medical institutions have control over patient data, while patients retain privacy rights [3,47]. To promote open utilization, an opt-out patient informed consent mechanism is adopted. The data platform collects health and medical data from medical institutions and aggregates it into usable healthcare big datasets. In the open utilization phase, the data platform anonymizes the data and provides it to data users serving the public interest, either for a fee or free of charge. Both parties sign data use agreements specifying the scope of data use, and users utilize the data according to the agreement.

For users, non-transparent data use may yield greater benefits. Non-transparent use refers to users reprocessing data for projects beyond the agreed scope or sharing it with other users, in addition to using data within the agreed scope [48]. Although agreements clearly limit the scope of use during data use applications, due to low data replication costs and the difficulty of detecting data reprocessing or flow when confidentiality is maintained, users may have a fluke mentality and use data non-transparently when facing temptation. This significantly increases the risk of patient data re-identification and leakage. Additionally, due to the complexity of the collection process, data is stored in multiple locations. If leakage occurs, the diversity of sources makes it difficult to quickly and accurately identify the origin.

For the data platform, providing data to users for a fee or free of charge yields data sharing benefits or social benefits from users' data utilization. Due to data leakage risks, the platform has a responsibility to regulate the use process. With privacy protection technology levels fixed, urging users to use data compliantly is the focus of platform regulation, including transparency construction in data use and tracing data leakage sources. Transparency construction in data use is a series of tasks requiring users to invest certain costs, such as urging the implementation of PPIE systems to allow public representatives to participate throughout the data use process; urging the implementation of DUR systems to maintain transparency in the data use process; and strengthening internal data access permission management and employee training. The platform has minimum requirements for users' privacy protection technology levels and transparency construction, investigates data leakage incidents, and imposes penalties.

For patients, patients have the rights to disagree, withdraw consent, and delete data [49-50]. Throughout the entire data open utilization process, patients can choose whether to continue participating at any time. Patient participation (or non-withdrawal) may be motivated by benefits such as improved healthcare quality [51], personal economic interests [52-53], or perceived social public benefits [54-55]. Users' non-transparent data use increases patients' perceived privacy risks, causing concerns about data misuse and privacy leakage. The stronger the privacy sensitivity of patient data, the greater the perceived loss from data leakage.

2.2 Model Assumptions

The three game players—data platform, user, and patient group—are illustrated in Figure 2 [Figure 2: see original paper]. The data platform makes decisions on regulatory intensity. Strong regulation means higher regulatory costs but yields public trust gains. In this case, the probability of detecting users' non-transparent data use increases. The platform's strategic choices are (Strong Regulation, Weak Regulation). Users decide whether to use data transparently. Transparent data use means strictly following the agreement, while non-transparent use means facing the temptation of non-compliant use for additional benefits while also bearing the risk of data leakage or detection by regulators. Users' strategic choices are (Transparent Use, Non-transparent Use). Patients decide whether to withdraw from data sharing. Withdrawal means non-participation, while non-withdrawal means continued participation. Patients' strategic choices are (Participate, Not Participate). The parameters and their meanings are shown in Table 1 .

Assumption 1: The data platform, users, and patients are all bounded rational agents. The data platform focuses on long-term sustainable development and considers trust gains and losses, which can be viewed as considerations of future benefits or losses. Users weigh choices between short-term and long-term benefits; non-transparent data use in the short term can bring immediate benefits but carries the risk of trust loss, while long-term considerations of trust gains and losses can be viewed as considerations of future benefits or losses. Patients' trust in the data platform influences their own decisions, which can be viewed as perceived trust gains and losses [56].

Assumption 2: Let x represent the proportion of users choosing transparency, then $1-x$ represents the proportion choosing non-transparency. Let y represent the proportion of the data platform choosing strong regulation, then $1-y$ represents the proportion choosing weak regulation. Let z represent the proportion of patients choosing participation, then $1-z$ represents the proportion choosing non-participation. $x, y, z \in [0, 1]$.

Assumption 3: The data platform's benefits and costs. When patients participate, they bring direct economic benefits and social benefits to the data platform, denoted as R. When the data platform chooses strong regulation, it bears regulatory costs C_1. Strong regulation increases public confidence in the security of data open utilization, yielding trust gains T_1 [56]. Under strong regulation, users' non-transparent data use may not always be detected; let the detection probability be p_1. When detected, the platform imposes a penalty F_1 on non-transparent users and forces them to cease non-compliant behavior. Under strong regulation, if a data leakage event occurs, the platform conducts a strict investigation and charges a penalty F_2 to the data leaker. Strong and weak regulation are relative concepts; to simplify the model, assume the regulatory cost under weak regulation is negligible. Under weak regulation, when a data leakage event occurs, the data platform suffers trust loss L_1 and charges a penalty F_2 to users after detection.

Assumption 4: Users' benefits and costs. Users must have established privacy protection technology levels and a certain degree of transparency construction to qualify for data use, with fixed investment C_2. Transparent data use yields benefits R_1. Non-transparent data use yields additional benefits R_2. The trust gain from transparent data use is T_2, which is also the fluke benefit when non-transparent use occurs under strong regulation but is neither detected nor leads to data leakage. Under strong regulation, when non-transparent use occurs and data leakage happens, users suffer trust loss L_2. Under weak regulation, when data leakage occurs, trust loss L_3 is incurred. Since the source of leakage cannot be confirmed under weak regulation, but trust loss occurs.

Assumption 5: Patients' benefits and costs. The sum of direct benefits or perceived social benefits when patients participate is denoted as R_3. The loss when privacy leakage occurs is L_4. The privacy leakage risk under transparent use is r_1, and under non-transparent use is r_2. Since transparent use reduces privacy risk, $r_1 < r_2$. The trust gain for patients when the data platform implements strong regulation is T_3, representing improved patient-perceived security. The net benefit for non-participating patients is 0, with no risk incurred.

2.3 Payoff Matrix

The payoff matrix for the tripartite evolutionary game among the data platform, users, and patients is constructed as shown in Table 2 .

3.1 Replicator Dynamic Equations

Based on Table 2, the expected payoffs for the data platform choosing "Strong Regulation" and "Weak Regulation" strategies are:

$$U_{11} = yz(R+T_1-C_1+p_{1F}1)+y(1-z)(R+T_1-C_1+p_{1F}1)+(1-y)z(R-L_1)+(1-y)(1-z)(R-L_1)$$

$$U_{12} = yz(R-C_1)+y(1-z)(R-C_1)+(1-y)zR+(1-y)(1-z)R$$

According to the dynamic equation construction method, the replicator dynamic equation for the data platform is:

$$F(y) = y(U_{11}-\bar{U}_1) = y(1-y)[(T_1+L_1-C_1)+p_{1F}1-(T_1+L_1)z]$$

Similarly, the replicator dynamic equations for users and patients are:

$$F(x) = x(U_{21}-\bar{U}_2) = x(1-x)[(T_2+L_3-R_2)-(T_2+L_2+F_2)y+(L_3-L_2)yz]$$

$$F(z) = z(U_{31}-\bar{U}_3) = z(1-z)[R_3-r_{2L}4+(r_2-r_1)L_{4x}+T_{3y}-(r_2-r_1)L_{4xy}]$$

3.2 Stability Analysis of Equilibrium Points

Further, setting $F(x)=0$, $F(y)=0$, and $F(z)=0$ yields 14 local equilibrium points. Since the asymptotically stable solution of a multi-population evolutionary game replication dynamic system must be a strict Nash equilibrium [57], only 8 pure strategy equilibrium points need to be considered: $E_1(0,0,0)$, $E_2(0,0,1)$, $E_3(0,1,0)$, $E_4(0,1,1)$, $E_5(1,0,0)$, $E_6(1,0,1)$, $E_7(1,1,0)$, $E_8(1,1,1)$. According to Friedman's method [58], partial derivatives are taken to obtain the Jacobian matrix J of the replication dynamic system:

$$J = \begin{pmatrix} (1-2x)[(T_2+L_3-R_2)-(T_2+L_2+F_2)y+(L_3-L_2)yz] & \cdots \\ \vdots & \ddots \end{pmatrix}$$

Substituting E_1 to E_8 into the Jacobian matrix and calculating the eigenvalues yields the results shown in Table 3 .

According to Lyapunov's first method [59], an equilibrium point is an Evolutionarily Stable Strategy (ESS) only when all eigenvalues of the Jacobian matrix have negative real parts. When one or more eigenvalues have positive real parts, the equilibrium point is unstable. When eigenvalues have both zero real parts and negative real parts, the equilibrium point is in a critical state, and its stability cannot be determined from the eigenvalue signs. Based on this, the following conclusions are drawn:

Conclusion 1: For $E_1(0, 0, 0)$ and $E_5(1, 0, 0)$, these are unstable points. T_2 represents users' consideration of trust gains. If users do not consider trust gains or cause trust losses due to short-term interests, i.e., if T_2 is small, $E_5(1, 0, 0)$ becomes a local equilibrium point, resulting in a situation where regardless of regulatory intensity, users choose non-transparent use while patients largely withdraw. This unstable point demonstrates that it is precisely users' consideration of trust gains that prevents the system from falling into the unfavorable local equilibrium state of (Regulation/Non-regulation, Non-transparent, Non-participation). Conversely, if data abuse becomes rampant and social trust is lacking, it is not surprising that people "do not want to and dare not" share data.

Conclusion 2: For $E_3(0, 1, 0)$ and $E_7(1, 1, 0)$, when $R_3 - r_{1L}4 < 0$, $E_3(0, 1, 0)$ is an ESS. However, in practical situations, when patients choose the non-participation strategy, the strategic choices of users and the data platform essentially lose meaning. For patients, the risk of continued participation is too great, with potential losses exceeding personal benefits. Even if users choose transparent use strategies, patients will still choose non-participation. This reflects that under established privacy protection technology levels and legal safeguards, when patients perceive excessive risks to privacy protection and basic personal rights, users' transparency strategies and regulators' regulatory strategies have almost no impact on patient behavior. A basic fact is that in reality, patients with extremely sensitive data have very low willingness to share data, such as information on HIV status, sexually transmitted infections, abortion records, IVF records, marital status, complaint records, and crime and abuse records, which should not be collected as noted in the care.data case [26].

Conclusion 3: For $E_2(0, 0, 1)$ and $E_6(1, 0, 1)$, when $R_3 - r_{2L}4 + T_3 > 0$, $E_2(0, 0, 1)$ is an ESS. This scenario also represents an extreme case where, under established technology levels and legal safeguards, when patients face minimal risks to privacy protection and basic personal rights, users' transparency strategies and regulators' regulatory strategies also have almost no impact on patient behavior. One can envision an ideal scenario: if an absolutely trusted research environment is in place, privacy protection technology can ensure privacy security, and legal regulations are well-developed, then no privacy leakage or other infringement issues would occur during data open utilization. In this case, due to patients' strong tendency to continue participating, the trust gain T_1 for data platforms from strong regulation is low, and the trust gain T_2 for users from transparent use is also low, making regulation and transparent use unnecessary. Data platforms would choose weak regulation strategies, and users would choose non-transparent use strategies.

Conclusion 4: For $E_4(0,1,1)$, when $C_1 > T_1 + L_1 + p_{1F}1$, $E_4(0,1,1)$ is an ESS. In this case, patients' risk losses from participation are less than their benefits, leading patients to 倾向于 choose participation strategies. Due to patient participation, the trust gain from strong regulation for data platforms becomes small, while costs exceed the sum of trust gains from strong regulation and trust losses from weak regulation, leading platforms to 倾向于 choose weak regulation strategies. Simultaneously, users' additional benefits from non-transparent use are less than the sum of trust gains from transparent use and trust losses from non-transparent use, leading users to choose transparent use strategies.

Conclusion 5: For $E_6(1,0,1)$, when $C_1 < T_1 + L_1 + p_{1F}1$, $E_6(1,0,1)$ is an ESS. In this scenario, the cost of strong regulation for data platforms is less than the sum of trust gains from strong regulation, trust losses from weak regulation, and penalty revenue from strong regulation, leading platforms to 倾向于 choose strong regulation strategies. Due to strong regulation, patients' trust gains increase, and patients' losses are less than the sum of direct benefits and trust gains, leading patients to 倾向于 participate. For users, due to the temptation of substantial additional benefits from non-transparent use and speculative psychology about evading regulation, users 倾向于 choose non-transparent strategies. This corresponds to real-world situations where potential additional benefits from non-transparent use are substantial.

Conclusion 6: For $E_8(1,1,1)$, when $C_1 < T_1 + L_1$, $E_8(1,1,1)$ is an ESS. In this case, the cost of strong regulation for data platforms is less than the sum of trust gains from strong regulation and trust losses from weak regulation, leading platforms to 倾向于 choose strong regulation strategies. Due to strong regulation, patients' trust gains increase, and patients' losses are less than the sum of direct benefits and trust gains, leading patients to 倾向于 continue participating. For users, the sum of trust gains from transparent use strategies and trust losses from non-transparent use strategies exceeds the difference between additional benefits and penalties from non-transparent use strategies, leading users to 倾向于 choose transparent use strategies. This corresponds to the best regulatory effectiveness scenario in practice.

4 Simulation Analysis

Considering the actual future scenario of healthcare big data open utilization in China, the initial state data is simulated based on the following facts: First, China attaches great importance to healthcare big data open utilization, so we assume the data platform's trust gain $T_1 > 0$. Second, as big datasets accumulate, the multiplier and 倍增 effects of data use become increasingly significant. We assume that all parties' R_2, R_3, and L_4 will increase rapidly, meaning users face growing temptation from non-transparent data use benefits.

Based on expert recommendations, the baseline initial values of various influencing factors are set as shown in Table 4 . Taking $x = y = z = 0.5$ and simulating in MATLAB, the evolution results are shown in Figure 3 [Figure 3: see original paper]a, indicating the system's initial equilibrium state is (Strong Regulation, Transparent Use, Participation). We then examine two anticipated state changes: (1) The continuous increase of R_2; (2) The effect of changes in major parameters on eliminating unstable fluctuations after R_2 increases to a certain extent.

Based on the set baseline initial values, changing the value of R_2 to 100, 200, and 400 sequentially yields evolution results shown in Figure 3 [Figure 3: see original paper]b. As R_2 increases, the data platform's strategic choice remains unchanged, still rapidly selecting the strong regulation strategy. However, when R_2 increases beyond a certain threshold (between 100 and 200), fluctuations appear in the strategic choices of data users and patients, and the larger R_2 becomes, the greater the fluctuation amplitude. The evolution process shows that when R_2 is large, the probability of users choosing non-transparent strategies increases, raising patient privacy risks and leading patients to 倾向于 choose non-participation. As patient participation decreases, users gradually return to transparent use, and patients subsequently return to participation, creating a cyclical pattern. The evolution results of users' and patients' strategic choices form two unstable wavy lines as shown in Figure 3 [Figure 3: see original paper]b.

The above analysis shows that when the benefits obtainable from users' non-transparent data use increase to a certain extent (e.g., $R_2 = 200$), significant fluctuations occur in the strategic choices of users and patients. Further research examines how changes in other parameter values can eliminate this unstable fluctuation. Setting $R_2 = 200$ in Table 4 while keeping other parameters unchanged creates a new set of initial values. Based on this, each simulation analysis changes only one parameter value.

Figure 3 [Figure 3: see original paper]c shows the evolution curves when $p_1 = 0.7$, $p_1 = 0.8$, and $p_1 = 0.9$. p_1 represents the data platform's regulatory intensity. Compared to $p_1 = 0.7$, the fluctuation amplitude of users' and patients' evolution curves is smaller when $p_1 = 0.8$. When $p_1 = 0.9$, users rapidly choose transparent use strategies, and patients subsequently choose participation strategies, restoring the system to the (Strong Regulation, Transparent Use, Participation) equilibrium state. This demonstrates that when p_1 increases beyond a certain threshold (between 0.8 and 0.9), unstable fluctuations in the system can be eliminated. Regulatory intensity has a positive impact on both data use transparency and patient participation. Similarly, Figures 3 [Figure 3: see original paper]d and 3e respectively show that the penalty amount for detecting non-transparent use under strong regulation and the penalty amount for data leakage events under strong regulation both have positive effects on data use transparency and patient participation.

Figure 3 [Figure 3: see original paper]f shows evolution results when $T_2 = 10$, $T_2 = 20$, and $T_2 = 30$. T_2 represents users' trust gains from transparent use, which can also be viewed as consideration of future benefits. When $T_2 = 10$, the proportion of transparent use by users stabilizes at 0.9. When $T_2 = 20$, users rapidly choose transparent use strategies, and patients subsequently choose participation strategies, restoring the system to the (Strong Regulation, Transparent Use, Participation) equilibrium state. This shows that users' trust gains from transparent use have a positive effect on data use transparency and patient participation. Similarly, Figure 3 [Figure 3: see original paper]g shows that users' trust losses from non-transparent use also positively affect data use transparency and patient participation.

T_3 is the trust gain patients receive when the data platform implements strong regulation, representing patients' trust level in the data platform. Figure 3 [Figure 3: see original paper]h shows evolution results when $T_3 = 0$, $T_3 = 4$, and $T_3 = 15$. When $T_3 = 0$, patients 倾向于 choose non-participation, while users 倾向于 choose transparent use. When $T_3 = 4$, due to patients' trust in the data platform, they 倾向于 choose participation, but this 反而 encourages users' speculative behavior, making users 更倾向于 choose non-transparent use. This demonstrates that to promote the long-term sustainable open utilization of healthcare big data, while improving patient trust levels, it is even more important to strengthen constraints on user behavior.

5.1 Main Conclusions

Practice in healthcare big data open utilization shows that trust among relevant stakeholders has become a key factor determining its effectiveness. Focusing on the decision-making behaviors of data platforms, users, and patients under considerations of trust gains and losses, this study established a tripartite evolutionary game model based on analysis of the open utilization process and conducted simulation analysis using MATLAB to reveal the influence mechanism of trust on stakeholder behavior. The findings show:

(1) Gaining patient trust is the fundamental motivation for users to transparently utilize data. Users' consideration of trust gains from transparent data use prevents the system from falling into the unfavorable local equilibrium state of (Regulation/Non-regulation, Non-transparent, Non-participation). If users lose patient trust, a large number of patients will choose to withdraw, and the amount of data available to users will drop sharply. Gaining patient trust is a necessary condition and fundamental motivation for users to choose transparent use strategies.

(2) Enhancing research environment security can fundamentally solve trust issues. When the open utilization environment has low security and patients perceive extremely high risks, they choose non-participation. When the open utilization environment has high security and patients perceive minimal risks, they choose participation. In these two extreme scenarios,

patients' participation decisions are independent of the decisions of the data platform and users, as well as their considerations of trust gains and losses.

(3) Strengthening regulation and increasing users' trust gains and losses contribute to achieving a virtuous state of open utilization. When the short-term benefits obtainable from users' non-transparent use are sufficiently large, they will neglect trust gains and losses and choose non-transparent strategies, leading to unstable fluctuations in all parties' strategic choices. After the data platform strengthens regulation, penalty amounts increase, and users' trust gains and losses are raised beyond certain thresholds, these fluctuations can be eliminated, promoting the system's return to the (Strong Regulation, Transparent Use, Participation) equilibrium state.

(4) While improving patient trust levels in data platforms, it is even more important to strengthen constraints on user behavior. When patients fully trust the data platform, they actively share data, but this 反而 encourages users' speculative behavior of non-transparent data use. When patients have high distrust in the data platform, they 倾向于 choose non-participation, while data users 反而倾向于 choose transparent use to regain patient trust. Governing user group behavior is fundamental to enhancing patient trust and promoting healthcare big data open utilization.

5.2.1 Nationally-Led Promotion of Healthcare Big Data Open Utilization Platform Construction

Nationally-led promotion of healthcare big data open utilization platform construction and implementation of strong regulation on user groups, fully leveraging the natural authority advantages of central and local governments, will help standardize user behavior and gain the trust of patient groups and data supply institutions. Chinese citizens have consistently shown very high trust in the government [60], and public trust in government directly affects whether government-issued information is convincing and whether formulated plans can be effectively implemented [61-62]. Currently, China has established nationally-led healthcare big data open utilization platforms such as the National Population Health Science Data Center, National Genomics Data Center, and National Bioinformatics Center, but these face issues such as poor data quality, insufficient data sharing, and low data utilization rates [63], requiring the development of more authoritative (e.g., CoreTrustSeal certified) and more open healthcare big data platforms. It is recommended that the National Health Commission take the lead, with local health commissions specifically promoting the establishment of a national-level healthcare big data research alliance, jointly aggregating and improving healthcare big datasets with China's healthcare big data custodians and research institutions. Meanwhile, central and local departments such as the Data Bureau, Ministry of Industry and Information Technology, and Public Security Bureau should collaborate with the Health Commission to strengthen the standardization of user behavior, enhance review of users' data use scope, strictly trace data leakage sources, and increase penalties for non-standard use behaviors, thereby improving public trust while vigorously standardizing user behavior to better achieve the aggregation, connectivity, and utilization of healthcare big data.

5.2.2 Strengthening Trusted Research Environment Construction and Actively Promoting Its Application

Enhancing the security of healthcare big data access through technical means is a "hard measure" for reducing patient privacy leakage risks, conducive to gaining the trust of patient groups and data supply institutions. Trusted Research Environments (TRE) are secure environments designed for researchers to access sensitive data based on network technology and management measures. They eliminate the need for physical data sharing among different users and truly achieve "data is usable but not visible" through five major aspects of security technology and management guarantees: secure people, secure projects, secure settings, secure data, and secure outputs. This ensures that researchers can quickly access and analyze required data while avoiding security responsibilities and risks such as data leakage caused by data downloads, safeguarding data security, establishing public trust, and achieving the highest standards of information governance, transparency, and security [64]. From a project implementation perspective, TRE ensures data security in privacy protection and other aspects while maximizing the use of high-performance computing and reducing data transmission and storage costs and security risks, thereby effectively promoting the conclusion of data utilization-related cooperation projects. It is recommended that healthcare big data open utilization platforms strengthen TRE construction, develop or introduce relevant technologies, enhance related infrastructure, and analyze highly sensitive data within TREs in conjunction with data classification and grading systems, reducing the flow of highly sensitive data and balancing utilization costs and benefits while ensuring data security.

5.2.3 Implementing Data Use Registration Systems and Establishing Standard Systems

Implementing data use registration systems can effectively reduce non-transparent use behaviors by data users. Clarifying information such as data users and use purposes promotes transparency in subjects, purposes, and processes of healthcare big data use, placing data users "in the open" to accept public review and supervision, thereby promoting transparent data use and enhancing public trust. Promoting data users to publish data use registers is a concrete way to implement data use registration systems. Data use registers, also known as data release registers or approved research lists, are public records of data that institutions provide secure access to for approved individuals and organizations for research, innovation, and service evaluation purposes. They typically contain information about the types of data accessed, purposes, approval dates, and the names of organizations using (or receiving) the data [65]. Data use registers, as the cornerstone of improving data use transparency, help build public confidence and enhance public trust. They help demonstrate the value and benefits of data use to society, facilitate researchers and research funding agencies in learning about more available datasets, and help relevant authorities understand data usage, generating broad social benefits. It is recommended that China reference the data use registration system and standardization practices of the UK Health Data Center [65] and advance the orderly development of China's data use registration work in accordance with national conditions.

5.2.4 Effectively Guiding Public Participation in the Collaborative Governance of Healthcare Big Data

Healthcare big data is essentially a macro aggregation of individual health data, involving the broadest public interests. Allowing the public to participate in decision-making regarding healthcare big data open utilization and exercise supervision rights helps standardize user behavior and enhance trust among patients and data supply institutions. The government should guide all sectors of society to value public worth, create an atmosphere for public participation, and integrate the concept of public participation in collaborative governance into the formulation of healthcare big data-related policies for top-down conceptual penetration. It should grant public participation rights and provide participation opportunities, offering legal and compliant participation channels through organizational structures and public activity releases. It is recommended that local health commissions establish public advisory committees composed of the public, with members determined through transparent elections, to fully leverage public advice and supervision functions. The government should also improve institutional normative systems to guide orderly public participation, such as developing 准入 policies for collaborative participation to ensure participating public members are representative, developing institutional norms for specific participation processes to clarify public participation goals and work content and standardize participation behavior, and developing feedback systems for public participation results. Finally, resource support should be provided to resolve barriers to public participation, such as offering knowledge support, pre-job training, and material incentives.

5.3 Research Limitations and Future Directions

Based on analysis of the UK's healthcare big data open utilization practices, this study focused on users' transparency construction and trust gains and losses of all parties, establishing a tripartite evolutionary model that provides beneficial exploration of the mutual influences among decision-making behaviors of healthcare big data open utilization stakeholders. However, to facilitate calculation with other benefits or costs, this study treated trust gains and losses as considerations of future benefits for all parties, conducting simple letter assignments and quantitative valuations for trust gains and losses, and examined the impact of individual variable changes on evolution results, which has certain limitations. Future research could more scientifically explore methods for quantifying trust gains and losses and simulation value settings, and further discuss the impact of simultaneous changes in multiple variables on evolution results. Additionally, this article only studied UK practices and their trust mechanisms; subsequent research could conduct more in-depth studies and comparative analyses of trust mechanism differences among different countries in healthcare big data open utilization, providing more references for China's participation in global health data governance.

References

[1] National Health Commission of the People's Republic of China. Notice on printing and distributing the "National healthcare big data standards, security, and service management measures (trial)"[EB/OL]. [2024-07-22]. http://www.nhc.gov.cn/mohwsbwstjxxzx/s8553/201809/f346909ef17e41499ab766890a34bff7.shtml.

[2] HDR UK. Our strategy[EB/OL]. [2024-07-25]. https://www.hdruk.ac.uk/about-us/our-strategy/.

[3] GAO F P. An entitlement of clinical data: the legal framework of clinical data sharing[J]. Modern law science, 2020, 42(04): 52-68.

[4] LI Y Q. Construction of the rights and interests system for the use of health data from the perspective of digital governance[J]. Journal of Hubei University (philosophy and social science), 2024, 51(02): 163-173.

[5] REN Y. Private law protection of the right to use medical data[J]. Journal of political science and law, 2023(06): 99-110.

[6] TANG J H. Risks and legal responses to healthcare big data sharing[J]. Study and practice, 2023(08): 54-60.

[7] LI J X, WANG L, SONG S J, et al. Research on open governance model for healthcare data driven by major public health emergencies[J]. Library and information service, 2022, 66(04): 23-32.

[8] ZHAI Y K, GUO R F, WANG Y, et al. Research on quality evaluation of medical and health big data from the perspective of data life cycle[J]. Journal of modern information, 2024, 44(01): 116-129.

[9] GUO Y, ZHANG C Y, LIU M T, et al. Research on influencing factors and paths for the reduction of users' information anxiety drives the realization of the value of medical and health big data[J]. Library and information service, 2023, 67(15): 25-34.

[10] ZHAI Y K, SONG X, WANG Y. Analysis of the value realization path of medical and health big data assets: based on information ecosystem theory[J]. Journal of technology economics, 2023, 42(11): 178-190.

[11] SINGH K, LI S, JAHNKE I, et al. Improving big data governance in healthcare institutions: user experience research for honest broker based application to access healthcare big data[J]. Behaviour & information technology, 2023, 43(6): 1067-1095.

[12] JIANG R, CHEN X, YU Y, et al. Risk and UCON-based access control model for healthcare big data[J]. Journal of big data, 2023, 10(01): 1-28.

[13] CHEN H, SONG Z, YANG F. Storage method for medical and health big data based on distributed sensor network[J]. Journal of sensors, 2023, 2(3): 1-10.

[14] SAROSH P, PARAH S A, MALIK B A, et al. Real-time medical data security solution for smart healthcare[J]. IEEE transactions on industrial informatics, 2022, 19(07): 8137-8147.

[15] ZHU L, ZHOU Y, FAN Q, et al. Enabling efficient and secure health data sharing for healthcare IoT systems[J]. Future generation computer systems, 2023, 149: 304-316.

[16] MAHAJAN B, JUNNARKAR A. Smart healthcare system using integrated and lightweight ECC with private blockchain for multimedia medical data processing[J]. Multimedia tools and applications, 2023, 82(28): 44335-44358.

[17] ALMALAWI A, KHAN A I, ALSOLAMI F, et al. Managing security of healthcare data for a modern healthcare system[J]. Sensors, 2023, 23(07): 3612.

[18] SARANYA K, PREMALATHA K. Multi attribute case based privacy-preserving for healthcare transactional data using cryptography[J]. Intelligent automation & soft computing, 2023, 35(02): 2029-2042.

[19] KERASIDOU X, MALONE M, DALY A, et al. Machine learning models, trusted research environments and UK health data: ensuring a safe and beneficial future for AI development in healthcare[J]. Journal of medical ethics, 2023, 49: 838-843.

[20] BRAUNACK-MAYER A, CAROLAN L, STREET J, et al. Ethical issues in big data: a qualitative study comparing responses in the health and higher education sectors[J]. Plos one, 2023, 18(04): e0282285.

[21] RASHI T, YOM-TOV E. Ethics of medical archival internet research data[J]. Journal of medical internet research, 2023, 25: e43754.

[22] GIUFFRÈ M, SHUNG D L. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy[J]. NPJ digital medicine, 2023, 6(01): 186.

[23] RENNIE S, ATUIRE C, MTANDE T, et al. Public health research using cell phone derived mobility data in Sub-Saharan Africa: ethical issues[J]. South African journal of science, 2023, 119(5-6): 1-7.

[24] TOMMEL J, KENIS D, LAMBRECHTS N, et al. Personal genomes in practice: exploring citizen and healthcare professionals' perspectives on person-

alized genomic medicine and personal health data spaces using a mixed-methods design[J]. Genes, 2023, 14(04): 786.

[25] SCHICKHARDT C, WINKLER E, SAX U, et al. Data infrastructures for health research: ethical framework, legal implementation[J]. Bundesgesundheitsblatt, gesundheitsforschung, gesundheitsschutz, 2023, 66(02): 160-167.

[26] VEZYRIDIS P, TIMMONS S. Understanding the care.data conundrum: new information flows for economic growth[J]. Big data & society, 2017, 4(01): 1-12.

[27] PRESSER L, HRUSKOVA M, ROWBOTTOM H, et al. Care.data and access to UK health records: patient privacy and public trust[J]. Technology science, 2015, 2015081103: 1-35.

[28] CARTER P, LAURIE G T, DIXON-WOODS M. The social licence for research: why care.data ran into trouble[J]. Journal of medical ethics, 2015, 41(05): 404-409.

[29] SU Q, JI L. The coordination mechanism of medical data sharing based on stochastic evolutionary game[J]. Information science, 2023, 41(09): 37-47.

[30] HAN P, GU L, ZHANG J M. Research on willingness to share medical data from perspective of privacy protection——based on tripartite evolutionary game analysis[J]. Journal of modern information, 2021, 41(03): 148-158.

[31] HAN P, ZHANG J M. Research on the willingness of medical data sharing considering privacy protection——from the perspective of evolutionary games[J]. Journal of modern information, 2021, 41(01): 111-120.

[32] BAIER A. Trust and antitrust[J]. Ethics, 1986, 96(02): 231-60.

[33] HAWLEY K. Trust, distrust and commitment[J]. Noûs, 2014, 48(01): 1-20.

[34] O'NEILL O. Trust, trustworthiness, and accountability[J]. Oxford university press, 2014.

[35] LHOUSNI S, DAOUDI F, BELMOKHTAR I, et al. Patients' knowledge and attitude toward biobanks in Eastern Morocco[J]. Biopreservation and biobanking, 2020, 18(03): 189-195.

[36] GASKELL G, GOTTWEIS H. Biobanks need publicity[J]. Nature, 2011, 471(7337): 159-160.

[37] GASKELL G, ALLANSDOTTIR A, ALLUM N, et al. The 2010 Eurobarometer on the life sciences[J]. Nature biotechnology, 2011, 29(02): 113-114.

[38] PAWLIKOWSKI J, WIECHETEK M, MAJCHROWSKA A. Associations between the willingness to donate samples to biobanks and selected psychological variables[J]. International journal of environmental research and public health, 2022, 19(05): 2552.

[39] SANDERSON S C, BROTHERS K B, MERCALDO N D, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US[J]. The American journal of human genetics, 2017, 100(03): 414-427.

[40] MILLS E J, SEELY D, RACHLIS B, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors[J]. The lancet oncology, 2006, 7(02): 141-148.

[41] MIDDLETON A, MILNE R, ALMARRI M A, et al. Global public perceptions of genomic data sharing: what shapes the willingness to donate DNA and health data?[J]. The American journal of human genetics, 2020, 107(04): 743-752.

[42] DIVE L, CRITCHLEY C, OTLOWSKI M, et al. Public trust and global biobank networks[J]. BMC medical ethics, 2020, 21(01): 1-9.

[43] MILNE R, MORLEY K I, ALMARRI M A, et al. Demonstrating trustworthiness when collecting and sharing genomic data: public views across 22 countries[J]. Genome medicine, 2021, 13(01): 92.

[44] AITKEN M, JORRE J, PAGLIARI C, et al. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies[J]. BMC medical ethics, 2016, 17(01): 1-24.

[45] LEMKE A A, WOLF W A, HEBERT-BEIRNE J, et al. Public and biobank participant attitudes toward genetic research participation and data sharing[J]. Public health genomics, 2010, 13(06): 368-377.

[46] SHABANI M, BEZUIDENHOUT L, BORRY P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review[J]. Expert review of molecular diagnostics, 2014, 14(08): 1053-1065.

[47] BALLANTYNE A. How should we think about clinical data ownership?[J]. Journal of medical ethics, 2020, 46(05): 289-294.

[48] CHEN Y. Research on healthcare data sharing and personal information protection[J]. Journal of intelligence, 2023, 42(05): 192-199.

[49] VOIGT P, BUSSCHE A. The EU general data protection regulation (GDPR), a practical guide[M]. Cham: Springer international publishing, 2017.

[50] GAO F P, LI Q T. The new civil law positioning of "personal information rights and interests"[J]. Dongyue tribune, 2023, 44(06): 166-176.

[51] WEITZMAN E R, KELEMEN S, KACI L, et al. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users[J]. BMC medical informatics and decision making, 2012, 12(01): 1-10.

[52] MEDFORD-DAVIS N, CHANG L, RHODES V. Health information exchange: what do patients want?[J]. Health informatics journal, 2017, 23(04): 268-278.

[53] ATHEY S, CATALINI C, TUCKER C. The digital privacy paradox: small money, small costs, small talk[R]. National Bureau of economic research, 2017.

[54] MÉJEAN C, EDELENYI F, TOUVIER M, et al. Motives for participating in a web-based nutrition cohort according to sociodemographic, lifestyle, and health characteristics: the NutriNet-Sante cohort study[J]. Journal of medical internet research, 2014, 16(08): e189.

[55] HOWE N, GILES E, NEWBURY-BIRCH D, et al. Systematic review of participants' attitudes towards data sharing: a thematic synthesis[J]. Journal of health services research & policy, 2018, 23(02): 123-133.

[56] QIU J P, XU Z Y, CHEN X H. Research on user privacy disclosure behavior in online health communities based on tripartite evolutionary game[J]. Information studies: theory & application, 2023, 46(01): 24-36.

[57] RITZBERGER K, WEIBULL J W. Evolutionary selection in normal-form games[J]. Econometrica, 1995, 63(6): 1371-1399.

[58] FRIEDMAN D. Evolutionary games in economics[J]. Econometrica, 1991, 59(03): 637-666.

[59] ZHU L L, RONG J M, ZHANG S Y. Three-party evolutionary game and simulation analysis of drug quality supervision under the government reward and punishment mechanism[J]. Chinese journal of management science, 2021, 29(11): 55-67.

[60] Edelman Public Relations Worldwide, the National Image Communication and Research Center at Tsinghua University. 2023 Edelman Trust Barometer China report[EB/OL]. [2024-08-06]. https://file.digitaling.com/eImg/uimages/20230403/1680510992119917.pd

[61] FANCOURT D, STEPTOE A, WRIGHT L. The Cummings effect: politics, trust, and trust behaviours during the COVID-19 pandemic[J]. The lancet, 2020, 396(10249): 464-465.

[62] KREPS S E, KRINER D L. Model uncertainty, political contestation, and public trust in science: evidence from the COVID-19 pandemic[J]. Science advances, 2020, 6(43): eabd4563.

[63] WU S Z, WANG A R, XIU X L, et al. Construction and implications of biomedical scientific data centers in Europe and the United States[J]. Digital library forum, 2022(04): 2-10.

[64] RUTH M. A pioneering programme to establish a national infrastructure for health data science[EB/OL]. [2024-08-02]. https://www.hdruk.ac.uk/news/a-pioneering-programme-to-establish-a-national-infrastructure-for-health-data-science/.

[65] KARRAR N, KHAN S K, MANOHAR S, et al. Improving transparency in the use of health data for research: recommendations for a data use register standard[EB/OL]. [2024-08-08]. https://zenodo.org/record/5902743#.Y19-A2DjzuE.

Author Contributions:
Wang Dandan: Determined the research topic and overall framework, wrote and revised the paper;
Xie Shicheng: Wrote and revised the paper;
Cheng Yingying: Revised the paper.

Note: Figure translations are in progress. See original paper for figures.