

# Data Elements and the Development of New Productivity in Biomedical Innovation

Caifen Xiang

Guangzhou College of Commerce, Guangzhou, Guangdong, 511363, China

## ABSTRACT

New productivity has emerged as a core driver of high-quality economic development, and its formation is increasingly shaped by data elements and technological innovation. Building on the theoretical framework of new productivity, this study conducts a systematic empirical investigation of how data elements promote new productivity through technological innovation pathways. Using a combination of Chinese provincial-level panel data and firm-level microdata from 2011 to 2023, the paper examines the driving effects, underlying mechanisms, heterogeneity patterns, and spatial spillovers associated with data elements. The results of this paper proposes policy implications including improving data openness and data trading systems, strengthening targeted support for breakthrough innovation, and optimizing interregional coordination in factor allocation. The study provides empirical evidence and practical insights for fostering new productivity in a context-sensitive and innovation-driven manner.

## KEYWORDS

Data elements; New productivity; Breakthrough innovation; Spatial spillover; Heterogeneity analysis

## 1 Introduction

New productivity represents an advanced form of productive capacity driven by disruptive technological innovation and characterized by qualitative improvements in efficiency and value creation. In knowledge- and data-intensive fields such as biomedicine, the development of new productivity is closely associated with advances in scientific research, technological innovation, and the effective integration of data resources. In this context, biomedical innovation has become an important domain in which the transformative role of data elements can be observed and evaluated.

With the rapid development of the digital economy, data have emerged as a critical production factor in biomedical innovation. Large volumes of biomedical data—including clinical records, research data, genomic information, and innovation-related enterprise data—provide essential inputs for research and development, technology diffusion, and innovation commercialization. Beyond their direct informational value, data elements contribute to new productivity by improving resource allocation efficiency, reducing information asymmetry, and facilitating the emergence of disruptive biomedical technologies. As a result, data-driven innovation is increasingly viewed as a key pathway for enhancing productivity in the biomedical sector.

To address these gaps, this study integrates provincial-level panel data with firm-level data from biomedical-related listed companies to construct a comprehensive evaluation framework for new productivity in biomedical innovation. The paper focuses on three core research questions. First, do data elements significantly promote the development of new productivity in biomedical innovation, and to what extent? Second, does this effect operate primarily through breakthrough innovation rather than incremental innovation? Third, how does the impact of data elements vary across regions and firm types within the biomedical sector, and are there spatial spillover effects? By providing systematic empirical evidence on these issues, this study aims to clarify the mechanisms through which data elements drive new productivity in biomedical innovation and to offer evidence-based insights for data-driven innovation policies.

## 2 Literature Review and Theoretical Framework

### 2.1 Conceptualization and Measurement of New Productivity in Biomedical Innovation

New productivity can be understood as a qualitative upgrading of productive capacity resulting from the coordinated transformation of core production elements, including labor, production resources, and production objects. In the context of biomedical innovation, this transformation is reflected not only in higher efficiency but also in the emergence of new research paradigms, advanced technologies, and innovation-driven industrial structures. Rather than representing a simple quantitative expansion, new productivity emphasizes structural optimization and qualitative change in innovation capability.

Existing studies generally conceptualize new productivity from three interrelated dimensions: the upgrading of human capital, the digitalization and intelligence of production resources, and the high-value and green transformation of production objects. These dimensions are particularly relevant for biomedical innovation, which is characterized by high

knowledge intensity, strong dependence on research infrastructure, and increasing requirements for sustainability and efficiency. Accordingly, an indicator-based evaluation framework can provide a systematic approach to capturing the multidimensional nature of new productivity in biomedical innovation.

From the labor perspective, new productivity is closely linked to the skill structure and innovation capacity of the workforce. In biomedical sectors, highly educated personnel and R&D professionals play a decisive role in knowledge creation and technological breakthroughs. Indicators such as average years of education and the proportion of R&D personnel are therefore used to reflect the upgrading of human capital and labor productivity in biomedical innovation activities.

From the perspective of production resources, digital and traditional infrastructures jointly shape the innovation environment. Digital infrastructure, including internet penetration and data connectivity, supports data-intensive biomedical research and cross-institutional collaboration. Traditional infrastructure, such as transportation networks, remains important for supporting industrial development and the diffusion of biomedical products. In addition, innovation inputs and outputs—measured by R&D expenditure intensity and patent authorization—capture the technological foundation and innovation performance that underpin productivity upgrading.

From the perspective of production objects, new productivity is associated with changes in industrial structure and development orientation. In biomedical innovation, this is reflected in the expansion of strategic emerging industries, the upgrading of industrial value chains, and increasing attention to green development. Indicators such as the output share of strategic emerging industries and energy consumption per unit of GDP are used to capture the degree of industrial upgrading and environmental efficiency.

## 2.2 Research Hypotheses

Based on the above analysis, this study proposes the following research hypotheses:

H1: Data elements significantly promote the development of new productivity in biomedical innovation.

H2: Data elements drive the development of new productivity in biomedical innovation primarily through breakthrough innovation rather than incremental innovation.

H3: The productivity-enhancing effect of data elements exhibits significant heterogeneity across regions, firm sizes, and industry segments within the biomedical sector.

H4: The impact of data elements on new productivity in biomedical innovation generates positive spatial spillover effects.

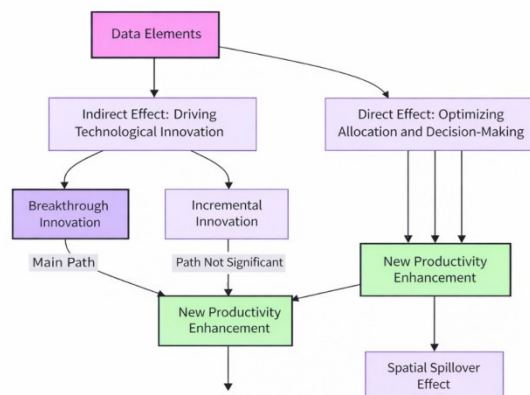


Figure 1 Theoretical Framework of Data-Driven New Productivity Development

## 3 Research Design

### 3.1 Data Sources and Variable Definitions

This study employs a balanced panel dataset covering 30 provincial-level regions in China (excluding Tibet and Hong Kong, Macao, and Taiwan) from 2011 to 2023, and complements the macro analysis with firm-level evidence based on China's A-share listed companies to validate mechanisms and heterogeneity patterns. Provincial data are collected primarily from the *China Statistical Yearbook* and the *China Statistical Yearbook on Science and Technology*. Firm-level and supplementary indicators are obtained from the Peking University Digital Finance Research Center and the CNRDS and CSMAR databases.

### 3.1.1 Dependent variable: New Productivity in Biomedical Innovation (NQP)

Following the literature on new productivity, we construct a composite index from the perspective of (i) the quality of labor, (ii) the upgrading of production resources, and (iii) the transformation of production objects. To align the measurement with the biomedical innovation context, the indicator system emphasizes human capital for R&D, innovation inputs and outputs, digital infrastructure supporting data-intensive research, and industrial upgrading toward technology-intensive sectors. The provincial annual index is calculated using the entropy-weighted method (Table 1). At the firm level, F\_NQP is constructed based on listed-company information following an established approach in related studies, and is used for micro-level robustness and mechanism testing (details reported in the Appendix/Methodology section).

### 3.1.2 Key explanatory variable: Data Elements (Data)

We measure the development level of data elements using an entropy-weighted composite index along three dimensions:

(1) Data infrastructure and connectivity (e. g., domain resources, IPv4 addresses, broadband access and related indicators);

(2) Data transformation capability (e. g., software and information service revenue, ICT employment and related capacity measures);

(3) Data-enabled application and utilization relevant to biomedical innovation ecosystems (e. g., digital platform development and digital inclusion proxies, and indicators reflecting data-intensive service adoption). This composite approach is intended to capture both the “availability” and the “utilization” of data elements in regional innovation systems.

Mediating variables: Technological Innovation. To differentiate innovation types, we use patent-based measures. Breakthrough innovation (B\_Inno) is proxied by the number of authorized invention patents, while incremental innovation (I\_Inno) is proxied by the sum of authorized utility model and design patents. All patent variables are log-transformed after adding one to address skewness.

### 3.1.3 Control variables

At the provincial level, we control for economic development (log GDP per capita), industrial structure (share of tertiary industry value-added), government intervention (fiscal expenditure/GDP), human capital (share of higher-education students), and openness (total imports and exports/GDP). At the firm level, we include standard controls such as firm size, firm age, leverage, profitability, and ownership structure.

## 3.2 Model Specification

Baseline model (two-way fixed effects). We estimate the baseline relationship between data elements and new productivity using a province-year two-way fixed-effects specification:

$$NQP_{it} = \alpha_0 + \alpha_1 Data_{it} + \sum_j \beta_j Controls_{it}^j + \mu_i + \lambda_t + \varepsilon_{it}$$

where  $i$  and  $t$  denote province and year, respectively;  $\mu_i$  and  $\lambda_t$  capture province and year fixed effects.

Mediation analysis. To test whether technological innovation mediates the effect of data elements on new productivity, we adopt a stepwise approach:

$$Tech_{it} = \gamma_0 + \gamma_1 Data_{it} + \sum_j \beta_j Controls_{it}^j + \mu_i + \lambda_t + \varepsilon_{it} \quad NQP_{it} = \delta_0 + \delta_1 Data_{it} + \delta_2 Tech_{it} + \sum_j \beta_j Controls_{it}^j + \mu_i + \lambda_t + \varepsilon_{it}$$

We implement the mediation tests separately for breakthrough innovation (B\_Inno) and incremental innovation (I\_Inno), consistent with H2.

Spatial econometric model. To examine spatial spillover effects (H4), we estimate a Spatial Durbin Model (SDM):

$$NQP_{it} = \rho WNQP_{it} + \beta_1 Data_{it} + \theta_1 WData_{it} + \sum_j \beta_j Controls_{it}^j + \mu_i + \lambda_t + \varepsilon_{it}$$

where  $W$  denotes a spatial weight matrix constructed based on geographical adjacency and/or economic distance;  $\rho$  is the spatial autoregressive coefficient; and  $\theta_1$  captures the spatially lagged effect of data elements.

## 4 Empirical Results and Analysis

### 4.1 Descriptive Statistics and Baseline Regression Results

Descriptive statistics for the main variables show substantial cross-provincial variation in both the level of new productivity in biomedical innovation and the development of data elements, indicating pronounced regional imbalance in data-driven productivity transformation.

Table 1 reports the baseline regression results. The estimated coefficient of Data is positive and statistically significant at the 1% level across all specifications, regardless of whether control variables are included. These findings provide initial empirical support for H1, suggesting that data elements constitute an important driver of new productivity enhancement in biomedical innovation.

Table 1 Baseline Regression Results

Variable	(1) NQP <sup>↔</sup>	(2) NQP <sup>↔</sup>	(3) NQP <sup>↔</sup>
Data <sup>↔</sup>	0.6385*** <sup>↔</sup> (0.1178) <sup>↔</sup>	0.7363*** <sup>↔</sup> (0.1043) <sup>↔</sup>	0.3176*** <sup>↔</sup> (0.0598) <sup>↔</sup>
Control variables	None	Partial	Full
Prov./Time Fixed Effects	Yes	Yes	Yes
Observations	360	360	360
Adjusted R <sup>2</sup>	0.077	0.126	0.534

**Note:** Robust standard errors are reported in parentheses. \*\*\*, \*\*, and \* denote statistical significance at the 1%, 5%, and 10% levels, respectively. All regressions include province and year fixed effects. Control variables are gradually added across specifications.

### 4.2 Robustness Checks

To ensure the reliability of the empirical findings, a series of robustness checks are conducted. First, the core explanatory variable is replaced by an alternative proxy. Specifically, a dummy variable indicating whether a region is designated as a national-level comprehensive big data pilot zone is used to substitute for the data elements index. The estimated results remain statistically significant.

Second, potential endogeneity concerns are addressed using an instrumental-variable approach. The interaction term between the number of post offices in each city in 1995 and the lagged volume of express delivery services is employed as an instrument for data elements, and a two-stage least squares (2SLS) estimation is conducted. The Kleibergen–Paap rk LM test rejects the null hypothesis of under-identification, and the Cragg–Donald Wald F statistic exceeds the critical value, indicating that the instrumental variable is valid. The second-stage coefficient of Data remains significantly positive.

Third, to mitigate the influence of extreme observations, all continuous variables are winsorized at the 1% level on both tails. Fourth, the sample period is adjusted by excluding the years 2020–2021, during which the pandemic experienced concentrated outbreaks.

All robustness checks consistently support the baseline conclusions.

### 4.3 Mechanism Analysis: The Key Mediating Role of Breakthrough Innovation

Table 2 reports the results of the mediation effect analysis. Column (1) shows that data elements significantly promote breakthrough innovation (B\_Inno). In column (2), after incorporating B\_Inno into the baseline regression, its coefficient is significantly positive, while the coefficient of Data remains significant but decreases in magnitude. This indicates that breakthrough innovation plays a partial mediating role, which is further confirmed by a significant Sobel test result.

Column (3) shows that data elements do not have a statistically significant effect on incremental innovation (I\_Inno). In column (4), after adding I\_Inno to the regression, its coefficient remains insignificant.

These findings provide clear empirical support for H2, indicating that data elements primarily drive the qualitative upgrading of new productivity by stimulating breakthrough and disruptive technological innovation, rather than through incremental improvements.

**Table 2** Mechanism Analysis: Mediating Effects of Breakthrough and Incremental Innovation in Biomedical Innovation

Variables	(1) B_Inno	(2) NQP	(3) I_Inno	(4) NQP
Data	0.2094*** (0.0499)	0.2541*** (0.0572)	0.0170 (0.0135)	0.3168*** (0.0599)
B_Inno		0.1012** (0.0401)		
I_Inno				0.0047 (0.0035)
Control Variables	YES	YES	YES	YES
Fixed Effects	YES	YES	YES	YES
Observations	360	360	360	360
R <sup>2</sup>	0.659	0.548	0.334	0.534

#### 4.4 Heterogeneity Analysis

(1) Regional heterogeneity. To examine regional differences, the sample is divided into eastern, central, and western regions. The results indicate that the positive effect of data elements on new productivity in biomedical innovation is strongest and most statistically significant in the eastern region, followed by the central region, while the effect in the western region is weaker and occasionally insignificant. This pattern is consistent with regional disparities in biomedical data infrastructure, R&D resource concentration, and the maturity of innovation ecosystems, which condition the effectiveness of data-driven productivity enhancement.

(2) Firm size heterogeneity. Firm-level evidence based on listed companies shows that data elements exert a significantly larger positive effect on new productivity for small and medium-sized enterprises (SMEs) than for large firms. SMEs in biomedical-related sectors tend to be more flexible in organizational structure and innovation strategy, allowing them to respond more rapidly to data-driven opportunities. In contrast, large firms often face higher coordination and adjustment costs in integrating data elements into existing innovation processes.

(3) Industry heterogeneity. The productivity-enhancing effects of data elements are more pronounced in data-intensive and technology-intensive industries, including biomedical-related manufacturing, pharmaceuticals, medical devices, and high-technology sectors. These industries rely heavily on R&D activities, digital platforms, and data-enabled experimentation, which amplifies the marginal returns from data element integration. Overall, the heterogeneity results provide consistent empirical support for H3.

#### 4.5 Spatial Spillover Effect Analysis

The results of the spatial econometric analysis are presented in Table 4. The spatial autoregressive coefficient ( $\rho$ ) is positive and statistically significant, indicating spatial dependence in the development of new productivity across regions. Moreover, the coefficient of the spatially lagged data elements term ( $W \times \text{Data}$ ) is also significantly positive, suggesting that improvements in data elements within one region not only enhance local biomedical innovation productivity but also generate positive spillover effects on neighboring regions.

These spatial spillovers may arise through multiple channels, including cross-regional knowledge diffusion, collaboration along biomedical innovation and supply chains, and demonstration effects associated with data-driven research and industrial practices. The findings confirm H4 and highlight the importance of interregional coordination in the development and utilization of biomedical data resources.

### 5 Conclusions and Policy Implications

This study investigates how data elements contribute to the development of new productivity in biomedical innovation by integrating provincial-level data with firm-level evidence. By combining baseline regressions, mechanism analysis, heterogeneity tests, and spatial econometric models, the paper provides a systematic assessment of the productivity effects of data elements and the conditions under which these effects materialize.

The empirical results consistently show that data elements play a significant role in enhancing new productivity in biomedical innovation. This effect remains robust after addressing potential endogeneity concerns and applying multiple robustness checks, suggesting that the observed relationship is not driven by model specification or sample selection. Data elements therefore represent more than a supportive background condition; they function as an active input shaping innovation efficiency and productivity upgrading in data-intensive biomedical activities.

Further analysis reveals that the productivity-enhancing effect of data elements operates primarily through breakthrough innovation rather than incremental innovation. This finding highlights a qualitative dimension of data-

driven development: data elements appear to facilitate disruptive technological change by expanding research possibilities, reducing information constraints, and enabling new combinations of knowledge. Incremental innovation, by contrast, does not exhibit a comparable mediating role, indicating that the contribution of data elements lies in fostering structural transformation rather than marginal improvement.

The analysis also uncovers substantial heterogeneity in the impact of data elements. The effects are stronger in regions with more developed data infrastructure and innovation ecosystems, as well as among small and medium-sized enterprises and technology-intensive industries closely related to biomedical innovation. These differences suggest that the effectiveness of data elements depends on complementary conditions, including organizational flexibility, human capital, and the maturity of local innovation systems. In addition, the presence of positive spatial spillover effects indicates that data-driven productivity gains can extend beyond regional boundaries through knowledge diffusion, interregional collaboration, and innovation network linkages.

These findings carry important implications for policy and governance in the context of data-driven biomedical innovation. Rather than focusing solely on expanding data volume, institutional efforts should prioritize improving the conditions under which data can be effectively utilized. Clear and stable data governance frameworks—covering access rights, secure sharing, and value realization—can reduce uncertainty and encourage broader participation in data-driven innovation activities. The orderly opening of public-sector data relevant to biomedical research may further lower entry barriers and enhance research efficiency.

The results also suggest that innovation policy should place greater emphasis on supporting exploratory and high-risk research activities that are more likely to generate breakthrough outcomes. Evaluation and funding mechanisms that tolerate uncertainty and experimentation can help align data utilization with transformative innovation goals. At the same time, addressing disparities in data access and digital capabilities remains essential. Targeted support for less-developed regions and smaller firms, combined with mechanisms that facilitate cross-regional collaboration and data sharing, can strengthen the diffusion of productivity gains.

Overall, this study contributes to a more nuanced understanding of how data elements shape productivity transformation in biomedical innovation. By emphasizing mechanisms, heterogeneity, and spatial interactions, the analysis underscores that data-driven productivity development is not automatic but contingent on institutional design and innovation capacity. These insights provide a basis for more adaptive and evidence-based approaches to fostering sustainable, data-enabled biomedical innovation.

## Funding

This research was supported by the 2025 University-Level Research Projects of Guangzhou College of Commerce (NO: 2025XJYB017 and 2025KYQD20)

## About the Author

Caifen Xiang, Ph.D., Associate Professor, Research interests: digital trade and cross-border e-commerce.

## References

- [1] Leonelli S. (2024). Is data science transforming biomedical research? Evidence, expertise, and experimental design. *Philosophy of Science*, 91(1), 1–22.
- [2] Liu Y., Chen Y., & Han L. (2023). Bioinformatics: Advancing biomedical discovery and innovation in the era of big data and artificial intelligence. *The Innovation Medicine*, 1(1), 100012.
- [3] da Silva R. G. L., Figueiredo A. M., & Buss P. M. (2024). The advancement of artificial intelligence in biomedical research and health innovation: Challenges and opportunities in emerging economies. *Globalization and Health*, 20, Article 44.
- [4] Sriram, V., Greene C. S., & Barak B. (2025). Addressing biomedical data challenges and opportunities to improve data quality, sharing, and analysis. *Scientific Reports*, 15, Article 90453.
- [5] Topol E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.