Combining data tokenization and real-world patient insights to bridge the gap for a more diverse and complete dataset

Parexel demonstrates data tokenization and linkage in a seamless direct-to-patient study

David Barratt
Senior Vice President,
Enterprise Account Leader

Clinical research using real-world data (RWD) typically faces shortcomings in linking that data with a patient's personal characteristics and experience, clouding the researchers' view of the individual patient. For example, characteristics such as race and ethnicity and social determinants of health (SDOH) are often missing in external datasets.

Parexel initiated a study protocol that aimed to evaluate technologies for linking participant engagement with RWD to try to close this gap. With this initiative, we incorporated innovative data science techniques using streamlined technology integration and tokenization. Our goal was to demonstrate that we could develop a more complete and holistic view of an individual patient through this methodology. We wanted to determine the viability of reaching out directly to a specific group of patients to gain their informed consent to complete a simple questionnaire about their health. Then, for patients willing to provide additional consent, we would tokenize the data obtained via the questionnaire and link it with that same individual's data in RWD sources.

In designing the patient questionnaire, the Parexel team had the foresight to include several questions that might not routinely be considered for a study. Because Parexel strongly supports diversity in clinical research, we included questions about race and ethnicity and SDOH. These included questions about income, educational attainment, and the source of patients' healthcare insurance.

The initiative enabled us to reach patients experiencing a specific disease (in this case, chronic heart disease being treated with anticoagulants) and connect RWD with confidential information about themselves, their circumstances, and their healthcare experiences. We succeeded in demonstrating that linking data through the use of tokens allows us to gain valuable insights by completing and complementing RWD.

What is tokenization/data linking?

A token – an encrypted string of characters – can be generated from a patient's personal information that securely and uniquely identifies that patient in a dataset. Tokens can be used to match data patient-by-patient across multiple data sources. This data linking makes a more holistic picture possible to help us put patients first.

How we did it

To achieve this proof of concept, Parexel managed and integrated the contributions of several of our partners to deliver our vision and create a seamless experience for the patient:

- Partner A provided an Institutional Review Board (IRB) review of Parexel's novel study documents and brought knowledge and experience of tokenization and privacy to help us ensure a fit-for-purpose protocol.
- Partner B used outreach techniques to pre-screen patients and seamlessly transfer them into Partner C's environment if they were suitable according to their responses.
- In the Partner C platform, patients were informed further about the study and, if willing, provided consent and answered the questions posed in our brief questionnaire. Further informed consent was sought to add tokens for their data.
- Partner D's tokens were generated from the information supplied by the patient within Partner C's platform.
- Partner E compensated the patients' time and effort for the first 100 participants to complete the study with a gift voucher triggered within the platform.
- Finally, Partner F provided the data platform into which we transferred the tokens and questionnaire data in preparation for analysis at Parexel.

What we found

A total of 676 patients participated in our study by signing the informed consent form (ICF) during the 40 days the system was open to recruitment. Of these patients, 95% consented to tokenizing their data and potentially linking it with other data for additional research. Of these, 52% went on to provide personal identification information in the survey, and 305 participants had their tokens generated.

Through the use of Partner D's tools, for those 305 patients, we found matching tokens in 19 of 20 commercially available datasets (none of which are insurers or providers). Two of the datasets checked achieved matches with more than 85% of tokenized patients.

Did our data agree with the linked data?

Our tokenization process used name, gender, and date of birth. Patients participated based on self-reported cardiovascular conditions, corresponding with the participants' diagnoses and treatments according to the linked external data. This suggests that our funnel to select a specific group of patients was effective and that patients report honestly – regardless of the compensation offered – when interacting with the study systems.

Did our population parameters match and enrich and complete the linked data?

In the data we linked, we found that almost 40% of patients were missing data such as race and ethnicity. Collecting this information directly from patients adds richness to the combined dataset. The same applies much more broadly to other socioeconomic factors, such as education and income, which are usually absent from external datasets.

160 of the patients responded to our survey to provide details of their healthcare insurance provision. Of these patients, 71% were covered by Medicare. This compares to 35% in the linked dataset. By using the survey approach, we were able to represent a greater proportion of Medicare patients in our analysis.

Does our patient population match populations in trials?

By adjusting our targeted potential identification and screening approaches, we can potentially gain insights from Asian, Hispanic, Native American, or other groups whose representation we seek. The 6.4% of linked participants reporting as Black/African-American is a much higher representation of these patients in clinical trials for 24 cardiovascular drugs granted FDA approval between 2006 and 2020.

What did we learn about patients after the active participation?

Three patients are reported in the linked data to have died several months after participation in the survey phase of our study. Looking in detail at these cases, including patient comments in the survey, provides valuable insights into the patient journey and validates the use case for tokenization in pursuing very long-term patient mortality outcomes. Insights include:

- Signs of distress.
- Social and familial pressures on patients that affect treatment adherence.
- Necessary improvements in patient communication.
- Care coordination.
- Risks associated with comorbidities and healthcare vulnerabilities (noting that our study overlaps with the COVID-19 pandemic).

Benefits of our approach

Where does Parexel's agile and configurable approach, as demonstrated in this study, help us find solutions to research challenges? It's possible to imagine that the approach applied during the study could achieve benefits in many situations, including:

- Pre-screen patients for planned studies and "pre-tokenize" patients expressing an interest in a study well before any clinic visits.
- Reducing effort from site personnel for tokenization helps them focus on delivering patient care.
- Tokenization and its potential for data-linking within the direct-to-patient construct can facilitate extended long-term follow-up and help reduce or even remove the costs of conventional follow-up.

We see many potential applications of the approach used in our study. Examples include using the platform as:

- A pre-screener for participation in a specific trial.
- A virtual waiting room for patients expressing an interest in trial participation.
- A means of long-term follow-up.
- As the basis for a natural history or burden of disease study.

What capabilities we developed

A direct-to-patient study incorporating innovative data science techniques presents several challenges that Parexel has been able to solve, including:

- Seamlessly passing the patient from one environment to another within a single user-friendly interface.
- Explaining tokenization and data-linking in accessible language for volunteers.
- Facilitating the withdrawal of consent should a participant wish to do so.
- Making sure that, whether a participant consented or decided not to proceed, the system provided appropriate appreciation and recognition in the direct-to-patient environment.
- Ensuring that participants were comfortable with privacy and reassuring that sensitive data used to generate tokens was secure and stored only as long as necessary to generate the token.

Conclusion

This study demonstrated that linking data through tokens allows us to gain valuable insights by complementing RWD and acts as proof of concept for the use cases outlined in this article. With data that relates to race, ethnicity, and SDOH, we can obtain a holistic view of patients who otherwise may be underrepresented. Customizing the Internet-based targeting and the series of questions in our direct-to-patient approach could help identify specific populations for research using RWD based on characteristics incomplete or absent from the RWD. In addition, from the study, we were able to deliver insights such as those related to patients' digital literacy and how to manage latency in external datasets effectively.

We believe that there are many other strong use cases for tokenization. Embedding it in a seamless direct-to-patient setting minimizes patient burden while facilitating further research directly with the patient by combining data we collect with external data sources. Using this approach, important questions about the health of enrolled participants can be answered. We can understand their treatment pathways and preferences, facilitate remote and long-term follow-up for endpoints and outcomes, as well as conduct post-authorization safety surveillance.

We are very excited that from this study, we were able to demonstrate the value of combining questionnaire responses with the longitudinal data that may extend both before and after the study for the participants we enrolled.

The know-how we have gained from this study and others of this kind prepares us well for innovative study implementation. With our success in achieving our vision through a blend of technology, effective vendor management, and patient-centricity, we look forward to extending these capabilities to our clients' innovative study challenges.

Diversity in the study patient population

Of the 334 participants, 265 also consented to having their data tokenized and, linked and matched in the claims dataset. In this group, 134 were female and 131 were male. 228 were white (86%), 17 Black/African-American (6.4%), 12 mixed race (4.5%), 4 Latino (1.5%), 3 Asian (1.1%), and 1 Native American (0.4%). Participants were distributed across more than 20 U.S. states.

For our linked participants, median and mean age were both 66 years, ranging from 42-85, a group typical of this disease state. Interestingly, throughout this age group, patients were comfortable using the web-based screening, consent, and questionnaire interface.

Social determinants of health (SDOH)

Income and educational achievement:

- Income levels in the 265 linked participants ranged from $2,500 to $162,500.
- Median income was $27,000.
- Not all participants completed their highest achieved education levels, but for 34%, it was high school, 10% college, and 7% postgraduate.

Health insurance:

- 68% of linked patients used Medicare, Medicaid (primarily government-funded healthcare), or a combination of the two.
- 19% used commercial healthcare only.
- 7% combined commercial insurance with Medicare.
- 4% had no insurance.
- 0.6% (1 patient) was covered by military/veterans' benefits.

With Heart™

We're always available for a conversation

Parexel International Corporation
2520 Meridian Pkwy, Durham, NC 27713, USA
+1 919 544-3170

Offices across Europe, Asia, and the Americas

www.parexel.com

parexel®