PRIVACY ANALYTICS

WHITE PAPER

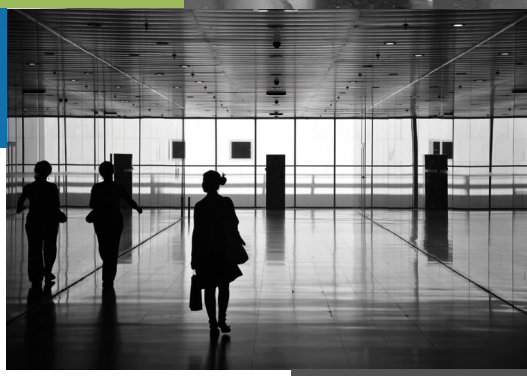
Specialty Pharmaceuticals and Patient Privacy

Khaled El Emam, President and Director of Real World Evidence Solutions, Privacy Analytics

Elizabeth Jonker, Privacy Analytics

Eileen Moyer, Senior Principal, Specialty Pharma Data Integration Center of Excellence (SPDI), IQVIA

Tiffany Williams, Engagement Manager, SPDI, IQVIA



PRIVACY ANALYTICS

a QuintilesIMS company

Contents

1. Introduction

The specialty biopharmaceutical market is growing at an astounding rate. These specialized, highly targeted products accounted for 30% of all pharmaceutical sales in 2015 and, according to recent projections, spending in this segment is expected to quadruple by 2020[1]. Although the patient population for many of these specialty drugs is small, each prescription is very valuable. The annual cost of treatment for a single patient can be in the hundreds of thousands.

A comprehensive view of the market is the foundation upon which specialty biopharmaceuticals base decisions vital to the future of their businesses. However, their products make their way into patients' hands through so many different and evolving channels that obtaining the information needed to form a complete picture can be extremely challenging. For example, when specialty biopharmaceuticals first appeared, they were mainly administered by physicians in-office and billed as a treatment under the medical benefit of the individual's health insurance[2]. Currently, there is a push toward self-administered specialty drugs that can be administered by patients in their homes and billed under the pharmacy benefit of their insurance[3]. This change may be beneficial to companies looking to track usage as claims under the medical benefit are opaquer and the bundling of specialty drugs with other services is not uncommon[4]. There are various channels of distribution that may be utilized, such as wholesalers, boutique pharmacy networks, individual specialty pharmacies, infusion centers, medical clinics, and others[5] for either route of administration. Patients can also sign up for patient support provided by a Patient Support Hub for many specialty biopharmaceutical products. The Hub both collects data on patients and requires access to patient data to provide support services. Therefore, in order to obtain a complete view of the patient experience, companies developing these drugs must acquire data from numerous disparate sources, many of which are not covered by syndicated data sources.

2. Data Privacy

Privacy can pose an additional challenge for companies looking to obtain and consolidate patient-related data from a number of different sources. Data protection regulations may prevent some data sources from providing patient data or limit the amount and type of data that is shared with pharmaceutical companies. Also, the extent to which companies are permitted to link data from disparate sources and the purposes for which they may process this data may also be limited under these regulations. Not to mention limitations on the sharing of information needed to fuel new research discoveries. The risk associated with data sharing is elevated for the small, specific populations that these companies target; however, the small size of these populations also means that a greater number of studies is needed to ensure the robustness and generalizability of the findings[6].

---

[1] J Krasowski, 'Lessons in Specialty Data Integration: Best Practices for Small and Emerging Biopharma Companies' (IMSHealth 2015) <http://bit.ly/2z9N9EX>.

[2] Tony Pinsonault, 'The World of Specialty Pharma' [2005] *Pharmaceutical Executive* <http://bit.ly/2zbA2ks> accessed 30 August 2017.

[3] I Spatz and N McGee, 'Health Policy Briefs - Specialty Pharmaceuticals' [2013] Health Affairs <http://bit.ly/2c1JDii> accessed 30 August 2017; B Bolgar and B Tonkovic, 'The Pharmacy Benefit vs the Medical Benefit' [2011] PharmacyTimes <http://bit.ly/2h7JwID> accessed 11 September 2017.

[4] Spatz and McGee (n 3).

[5] Krasowski (n 1); Spatz and McGee (n 3).

[6] Pamela Buffone, 'Specialty Pharma and Patient Privacy Protection' <http://bit.ly/2ipGD2x> accessed 31 October 2017.

One solution to meet the requirements and data protection regulations in such data sharing and aggregation scenarios is to de-identify the data prior to collection and/or use. According to the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which governs the collection, use, retention, and disclosure of protected health information, "health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information."[7] Therefore, de-identified patient data is not subject to regulations which restrict the purposes for which personal information can be used and disclosed.

Being able to aggregate the patient information required from all the various sources in a de-identified form for which there is a single anonymous ID/key assigned to each patient allows companies to obtain a complete picture of the patient journey that is essential to inform their business decisions while protecting the privacy of individual patients. Having a single anonymous key for each patient allows a biopharmaceutical company to link records from various points of contact (prescriber, Patient Support Hub, specialty pharmacy, etc.), as well as integrating these with syndicated data sources (laboratory data, retail data, etc.) to improve reporting and analytics capabilities.

The integration of these data sources makes it possible for companies to assess the complete patient experience and ensure patients are receiving the support they need from the Patient Support Hub and/or other patient programs; they are able to keep track of patients who switch from one specialty pharmacy to another; compare the performance of patients going through a support hub against that of patients purchasing directly from a specialty pharmacy; assess the impact of free goods or starter programs on patient conversion; determine payer issues by measuring patient time to fill, conversion rates, and percentages of claims that are rejected or reversed; measure the effectiveness of copay and savings card programs on persistence and compliance; and measure trends in all of these over time.

This de-identified patient data also helps companies assess the value of a given drug and the effectiveness of the various methods of distribution. For example, the data from the Patient Support Hub and other sources of support allows companies to determine the return on investment of the patient programs in which they invest by measuring their impact on persistence, compliance, and retention and to, consequently, determine if they should provide more of an incentive to patients to encourage the use of support programs. Having access to this data allows companies to compare the patient-level performance of specialty pharmacies across a number of metrics, such as time to fill, conversion rate, compliance, and others, in order to inform decisions regarding distribution. Analysis of de-identified patient data can illuminate opportunities to improve prescriber performance by identifying high reject and reversal rates on insurance claims and providing assistance with prior authorization support and education. Changes in trends over time can be measured with the use of integrated, de-identified data to inform decisions regarding the future viability of products. De-identified patient data can also be shared more openly for research purposes, facilitating new research discoveries and potential breakthroughs.

Some of the data flows and data uses outlined above may not even be possible without de-identification. If they were possible, through some legal authority, the obligations placed on companies to allow for the use of identifiable patient data would likely be onerous and/or expensive. On the other hand, when patient data is de-identified these vital activities can proceed unencumbered by such obligations while still ensuring that all legal requirements are met and patient privacy is protected.

[7] US Congress, 'The Health Insurance Portability and Accountability Act of 1996; 45 Code of Federal Regulations 164.154(b)1 Expert Determination' <http://bit.ly/2xqRMGw>.

Many biopharmaceutical companies outsource the aggregation, integration and even reporting/analytics functions for their specialty products. De-identification can help to ensure that any patient data companies share with service providers is properly safeguarded with respect to privacy. The vendors that they partner with should demonstrate a strong understanding of patient privacy requirements and proactively manage the acquisition, integration and dissemination of this protected data asset.

3. De-identification Standards

Most systematic methods used to de-identify individual-level data can be grouped into two camps: (1) Heuristic-based methods and (2) Risk-based methods. Risk-based de-identification methods reflect the emerging perspective from regulators, and have been encoded in various standards and guidelines[8].

The HIPAA Privacy Rule, in conjunction with subsequent guidance published by the Office for Civil Rights (OCR) at the US Department of Health and Human Services (HHS)[9], provide the most prescriptive guidelines for the de-identification of personal health information.

Section 164.514(b) of the Privacy Rule contains the implementation specifications that a covered entity, or affiliated business associate, must follow to meet the de-identification standard. In particular, the Privacy Rule outlines two routes by which health data can be designated as de-identified. These are illustrated in Figure 1.

[8] For example, Office of the Information and Privacy Commissioner of Ontario, 'Deidentification Guidelines for Structured Data' (Office of the Information and Privacy Commissioner of Ontario 2016) <http://bit.ly/2ypH3B1> accessed 30 August 2017; Health Information Trust Alliance, 'HITRUST De-Identification Framework' (HITRUST Alliance 2015) <http://bit.ly/1F8hgrE>; Office for Civil Rights, 'Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule' (Department of Health and Human Services 2012); The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, 'Accessing Health And Health-Related Data in Canada' (Council of Canadian Academies 2015).

[9] Office for Civil Rights (n 8).

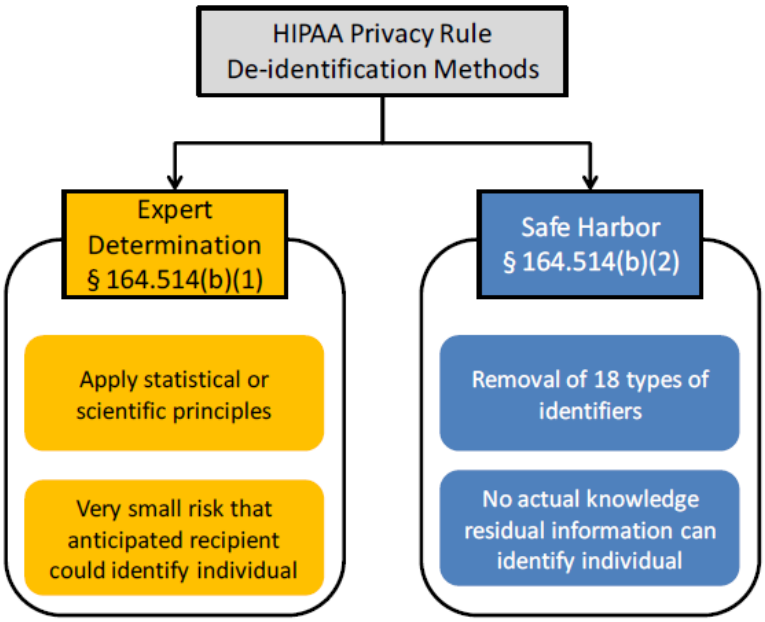


Figure 1: The two de-identification standards in the HIPAA Privacy Rule.[10]

The first route is the "Safe Harbor" method. Safe Harbor is a heuristic which requires the manipulation of 18 fields in a data set. The Privacy Rule requires that a number of these data elements be "removed". This method is not considered to be risk-based.

While the application of Safe Harbor is straightforward, there are instances in which dates and sometimes more fine-grained geographic information are necessary. In practice, the Safe Harbor standard would remove critical geospatial and temporal information from the data, potentially reducing the utility of the data. Many meaningful analyses of health data sets require the dates and event order to be clear. For example, in a Safe Harbor data set it would not be possible to include the dates on which prescriptions were dispensed.

In recognition of the limitations of de-identification via Safe Harbor, the HIPAA Privacy Rule provides for an alternative in the form of the Expert Determination method, which is a risk-based approach to de-identification. This method has three general requirements:

The de-identification must be based on *generally accepted statistical and scientific principles* and *methods for rendering information not individually identifiable*. This means that the covered entity, or affiliated business associate, needs to ensure that there is a body of work that justifies and evaluates the methods that are used for the de-identification, and that these methods must be generally known.

The risk that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to re-identify an individual who is a subject of the information needs to be "very small". However, the mechanism for measuring re-identification risk is not defined in the HIPAA Privacy Rule, and what would be considered very small risk is also not defined. Therefore, the de-identification methodology must include some manner of measuring re-identification risk in a defensible way, and have a repeatable process to follow that allows for the definition of very small risk.

[10] This diagram was copied from the OCR guidance document, ibid..

Finally, the methods and results of the analysis that justify such determination must be documented. While the basic principles of de-identification are expected to be consistent across all clinical trials, the details will be different for each study and these details also need to be documented.

These requirements establish a baseline for acceptable risk-based de-identification methods.

4. Basic Principles of Risk-based De-identification

There are three fundamental concepts underpinning risk-based de-identification methods.

The first concept differentiates among the different types of information in data. Data may have direct identifiers, which are things like a patient's NHS number or social security number. Direct identifiers are assumed to directly identify a patient with a high probability of success. These are typically pseudonymized or removed to create pseudonymous data. This type of data is still considered personal information [11]. Another type of information would be quasi- (or indirect) identifiers. These are things like demographics (e.g., age, gender, race), socio-economic information (e.g., income, and years of education), and clinical events (e.g., rare diagnoses, and hospital visits). There is evidence that this kind of information can be used to identify individuals [12]. Dealing with this kind of information can produce anonymous data. Figure 2 below presents further examples of direct and quasi-identifiers.

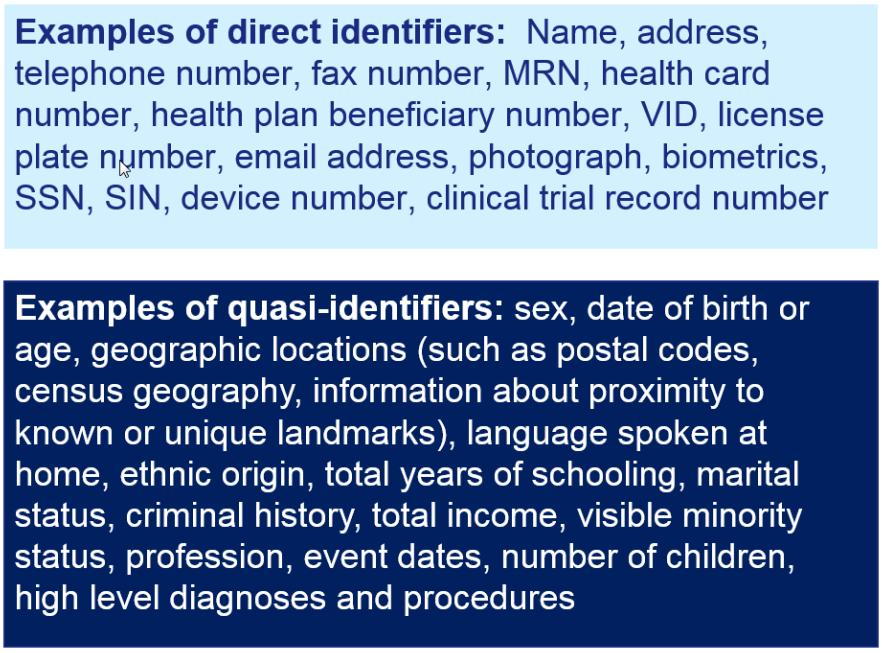


Figure 2: Direct and Quasi-identifiers.

[11] K El Emam, 'Pseudonymous Data Is Not Anonymous Data' (*The BMJ*, 20 November 2014) <http://bit.ly/1uf8wX9>.

[12] Khaled El Emam and others, 'A Systematic Review of Re-Identification Attacks on Health Data' (2011) 6 PLoS ONE <http://bit.ly/2hYogS0>.

Therefore, the act of de-identification is focused on the quasi-identifiers only. The assumption is that pseudonymization has already been applied to address re-identification risks from direct identifiers.

The second underlying concept is that risk-based methods are quantitative. The quantity that is being measured is the risk of re-identification of an individual in the data. The initial step is to set an acceptable threshold for this risk. This means that the acceptable risk of re-identification is going to be some value larger than zero. Setting the threshold at zero risk means that no useful data will be shared at all.

The actual risk of re-identification is then measured on the data. This measured value is compared to the threshold. If the measured risk is above the threshold then the data is not considered anonymous. If the measured risk is below the threshold then the data is considered anonymous. If the data is not anonymous then various transformations can be applied to bring the measured risk below the threshold. These transformations may include generalizing certain values in the data (for example, generalizing a date of birth to a year of birth), or suppressing certain values in the data that make individuals stand out.

The general measurement process is illustrated in Figure 3.

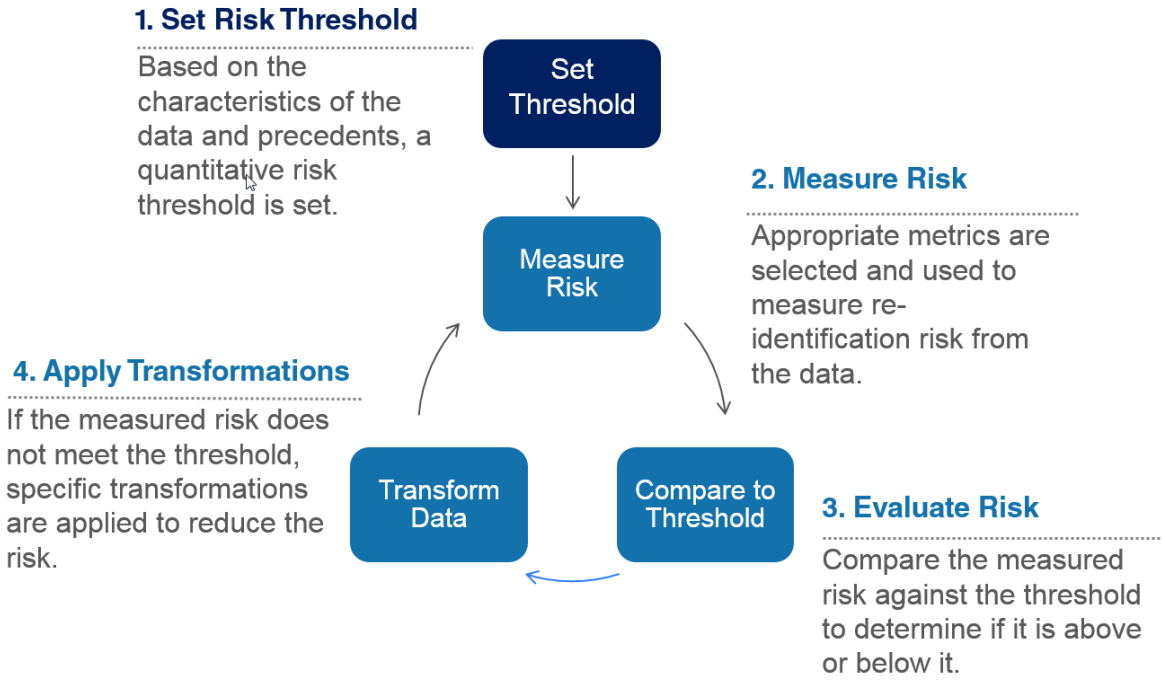


Figure 3: Overview of the risk measurement process.

The third concept pertains to the context of the data. The actual risk of re-identification is a function of both the data and the context. The context represents the security, privacy, and contractual controls that are in place. For example, one context can be a public data release (e.g., an open data initiative). Another context would be a researcher who analyzes the data in a very secure enclave. These are two very different contexts and the risk of re-identification is different in each of these, even for the same data.

The exact definition of the context is illustrated in Figure 4. It consists of characteristics of the data recipient / holder, the contract or data use agreement, and the data itself [13].

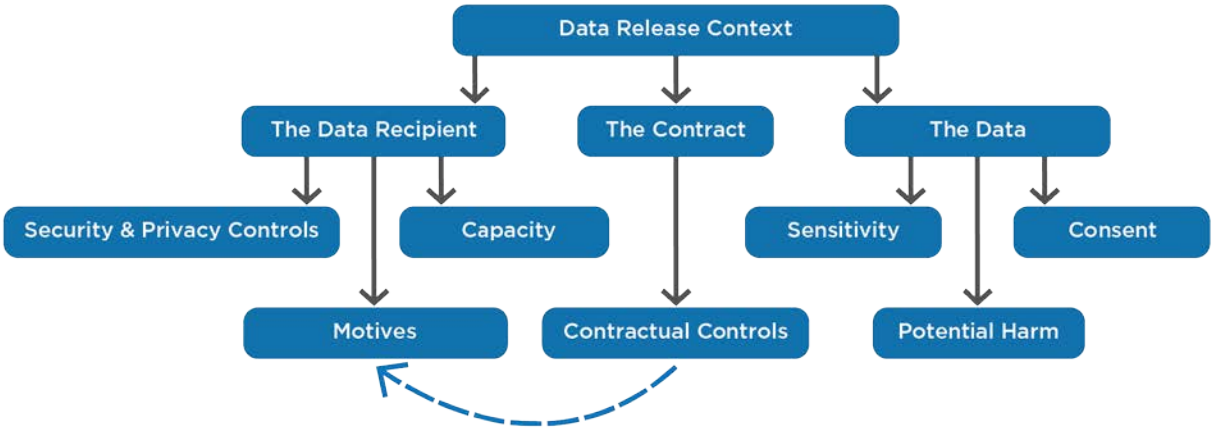


Figure 4: Definition of the data context.

This is illustrated in the diagram in Figure 5. Here we see that the overall risk is a function of both, the data risk and the context risk. When expressed as probabilities, the overall risk of re-identification is the multiplication of these two numbers.

This means the same data can have different levels of risk if it is processed in different contexts. It also means the same data can have different risk levels as it moves from one organization to another in the same data flow (i.e., over time). For example, if the data moves from an organization performing analytics to a say, a researcher, the risk may be low in the first instance but increase in the second instance after the transfer.

[13] Khaled El Emam, *Guide to the De-Identification of Personal Health Information* (CRC Press (Auerbach) 2013).

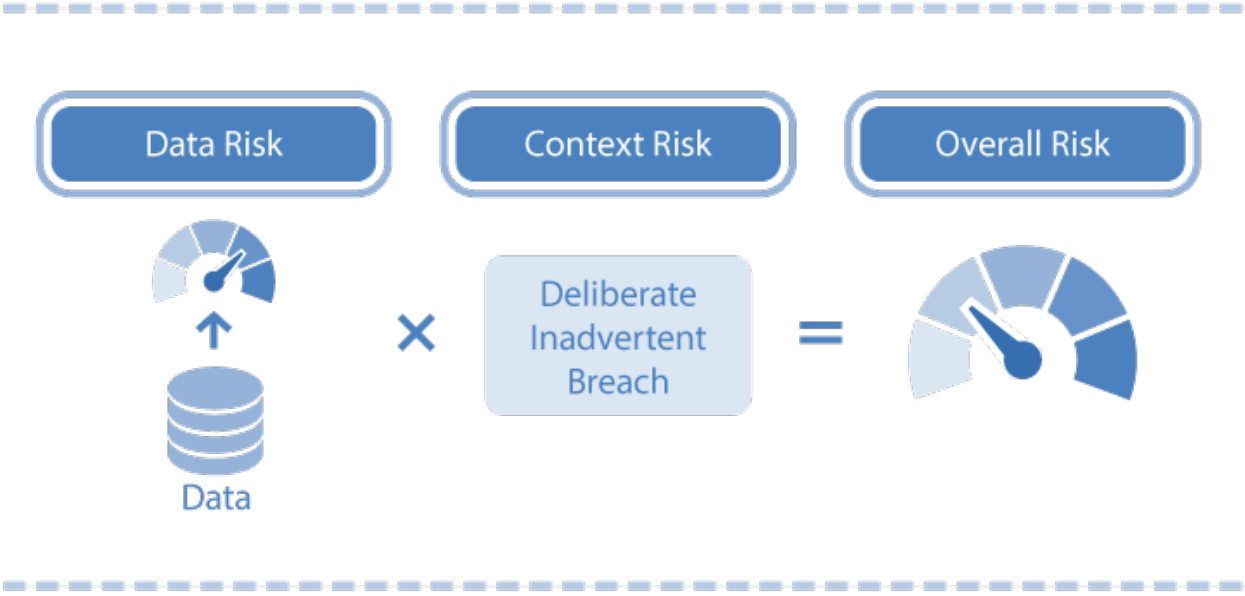


Figure 5: Re-identification risk is a function of data risk and context risk.

5. The De-identification Process

The steps of the combined pseudonymization and the full de-identification process are summarized at a high level in Figure 6. Once the variables in the data have been classified into direct and quasi-identifiers, the direct identifiers can be pseudonymized, and then the quasi-identifiers can be dealt with to de-identify the data.

Documentation of the process and results is important to ensure questions about the assumptions can be answered later if need be, and to allow replicability of the analysis.

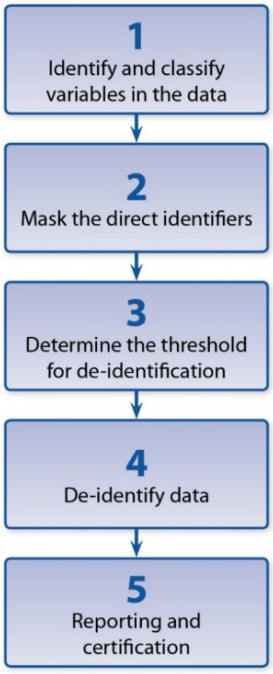


Figure 6: The de-identification process.

6. Conclusion

By de-identifying data either at the data source or after retrieval (i.e. during the data aggregation and integration process) specialty biopharmaceutical companies can ensure they are able to tap into the broad base of evidence required to support and grow their business in a responsible manner which maintains individual privacy.

Safeguarding patient privacy is critical to gaining and maintaining patient trust. People are more willing to accept risk (i.e. re-identification risk) if they are confident in the organizations and methods being used to manage that risk[14]. However, when they are not confident the risk is being properly managed or their trust in an organization has been damaged due to a breach, patients' willingness to share their information is negatively affected [15]. Privacy breaches have also been shown to lead to a loss of business/customers for companies experiencing a breach[16].

[14] Paul Slovic, 'Perceived Risk, Trust, and Democracy' (1993) 13 Risk Analysis 675; Nancy E Kass and others, 'The Use of Medical Records in Research: What Do Patients Want?' (2003) 31 Journal of Law, Medicine & Ethics 429.
[15] Sharyl J Nass, Laura A Levit and Lawrence O Gostin, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research* (The National Academic Press 2009); Laura J Damschroder and others, 'Patients, Privacy and Trust: Patients' Willingness to Allow Researchers to Access Their Medical Records' (2007) 64 Social Science & Medicine 223; Thomas S Mayer, 'Privacy and Confidentiality Research and the U.S. Census Bureau Recommendations Based on a Review of the Literature' (Statistical Research Division, US Bureau of the Census 2002) Survey Methodology 2002-01; Eleanor Singer, John Van Hoewyk and Randall J Neugebauer, 'Attitudes and

Biopharmaceutical companies must ensure the patient data purchased from specialty pharmacies and/or used to consolidate with other patient data (such as claims, hubs) is properly safeguarded with respect to privacy. As previously mentioned, many companies outsource the aggregation, integration and even reporting/analytics for their specialty products. It is essential that the vendors they partner with demonstrate a strong understanding of patient privacy requirements and proactively manage the acquisition, integration and dissemination of their patient data. Risk based de-identification provides a strong method of ensuring patient privacy is protected while also producing high quality data that can illuminate the kinds of insights that will shape the future of this sector.

Behavior: The Impact of Privacy and Confidentiality Concerns on Participation in the 2000 Census' (2003) 67 The Public Opinion Quarterly 368.

[16] 'Ponemon Institute's 2017 Cost of Data Breach Study: Global Overview' (IBM Security and Ponemon Institute 2017) <https://ibm.co/2v2AUIi> accessed 1 November 2017.

Contact Us

Mail: 251 Laurier Avenue W, Suite 200
Ottawa, Ontario, Canada
K1P 5J6

Phone: +1 613 369 4313

Toll Free: +1 855 686 4781

Fax: +1 613 369 4312

Email: sales@privacy-analytics.com

Online: www.privacy-analytics.com