INSTITUTE FOR DEFENSE ANALYSES

Methodology to Assess Risk from Strategic Competitor Acquisition of U.S. Biological Data and Application to an Agricultural Bioprocessing Case Study

Robert Cubeta
Kristen Bishop
Janet Marroquin Pineda
Ashley Farris
Clay Hamill
Jay Shah

August 2023
Approved for public release;
distribution is unlimited.
IDA Paper P-33619
Log: H 23-000308

INSTITUTE FOR DEFENSE ANALYSES
730 E.Glebe Rd
Alexandria, VA 22305

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses under contract HQ0034-19-D-0001, project AI-6-5283, "Biology Data Risk Assessment Methodology" for the Director, Science & Technology Exploitation and Analytics, Maintaining Technology Advantage (MTA), Office of the Under Secretary of Defense, Research & Engineering. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The authors wish to thank, Jeff Grotte, Carly Cox, and Laura Odell for their thoughtful reviews, and the publications team Florestine Purnell and Amberlee Mabe-Stanberry.

For More Information:
Mr. Robert L. Cubeta Project Leader
rcubeta@ida.org, 703-575-4681
Ms. Jessica L. Stewart, Director, SFRD
jstewart@ida.org, 703-575-4530

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-33619

Methodology to Assess Risk from Strategic Competitor Acquisition of U.S. Biological Data and Application to an Agricultural Bioprocessing Case Study

Robert Cubeta
Kristen Bishop
Janet Marroquin Pineda
Ashley Farris
Clay Hamill
Jay Shah

This page is intentionally blank.

Executive Summary

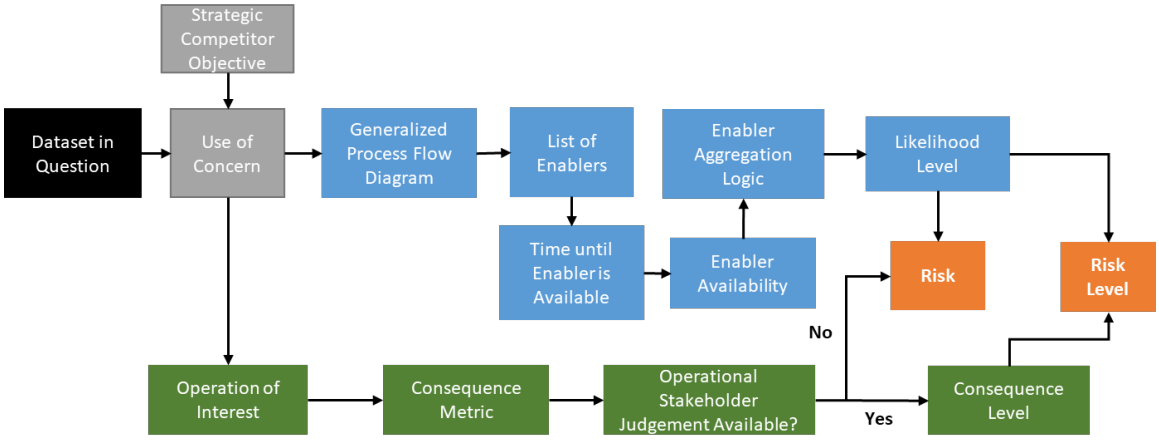
Strategic competitors use a variety of licit and illicit methods to acquire U.S. biological data. Extraction methods include: research partnerships, investments, mergers, acquisitions, cyber intrusions, and combinations of these tactics. Additionally, U.S. biological data is willingly provided to firms owned by strategic competitors through fee-for-service arrangements such as those for genetic sequencing and analytics. Such data sharing initially may be conducted for legitimate and permissible uses, but the biotechnologies using these datasets are inherently dual-use and also can enable nefarious applications, posing national security risks.[1]

The Director, Science & Technology Exploitation and Analytics, Maintaining Technology Advantage (MTA) of the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) asked the Institute for Defense Analyses (IDA) to:

1. Develop a repeatable methodology to assess the national security risk posed by strategic competitor acquisition of U.S. biological datasets either alone or when combined with other data, and
2. Apply the methodology to representative case studies that illustrate both the threat and risk of strategic competitor acquisition of U.S. biological data to facilitate messaging across the Department of Defense (DOD) and broader National Security audiences.

IDA's methodology assesses risk as the product of 1) the likelihood of a strategic competitor successfully achieving a user-specified application of a given dataset, and 2) the resulting consequence to a user-specified operation of interest. This operation of interest need not be a specific military operation. Consequence to other national security activities such as intelligence activities or economic competitiveness can be considered. IDA's risk assessment methodology is presented in the following figure.

[1] Edward H. You, "Safeguarding the Bioeconomy: U.S. Opportunities and Challenges," Testimony for the U.S.-China Economy and Security Review Commission (Washington, DC: March 16, 2017), https://www.ehidc.org/sites/default/files/resources/files/Ed_You_Testimony_USCC.pdf.



Overview of IDA's Risk Assessment Methodology

The risk assessment methodology is executed twice for a given dataset to determine the change in risk associated with the strategic competitor's access to the data. In some cases, a strategic competitor may already be capable of achieving the use of concern, and therefore acquiring the dataset does not increase risk to the operation of interest. In other cases, the dataset in question may represent a critical bottleneck, and therefore the data acquisition increases risk to the operation of interest. Datasets that fall into the latter category should be considered for safeguarding.

The topics of the presented case studies were selected based on the guidance of the sponsor and other national security stakeholders. The goal was to consider a diverse range of biological datasets and potential applications to include military operations, intelligence operations, and economic competitiveness. The case studies spanned four categories:

1. Human genetic data
2. Human microbiome data
3. Data relating to industrial biotechnology
4. Geospatially tagged biological data[2]

In this abridged document, we describe IDA's risk assessment methodology and summarize the analysis for one illustrative example pertaining to agricultural bioprocessing data. Specifically, the example examines China's use of privately held bioprocessing data to develop genetically engineered (GE) disease-resistant pigs. Using IDA's risk assessment methodology, we find that it is **Very Unlikely**, both with and without the dataset in question, that China could reduce 18.3 percent of the value of global U.S. pork exports via

[2] For the case studies, geospatially tagged biological data refers to any type of biological data that is associated with a specific geographic location.

the production of porcine respiratory reproductive syndrome (PRRS)-resistant pigs for domestic consumption in the next five years.

In its current form, IDA's risk assessment methodology possesses several limitations that should be taken into consideration by the user, as these limitations may impact the methodology's utility:

- Given the time and resources available, IDA did not validate the accuracy or reproducibility of the methodology.
- The methodology requires a substantial level of expertise and time to execute, which may pose a challenge to other users. We did not assess how the use of the methodology by someone with constrained time or expertise would impact the results.
- The risk posed by a single user-specified use of concern toward a single user-specified operation of interest may not be the greatest or even representative risk associated with a given dataset. Users may lack the time to execute the methodology on multiple uses and operations for a given dataset.

A follow-on assessment of the methodology's usability by those outside the IDA team will inform methodological developments that ideally would overcome these limitations and improve the methodology's utility.

This page is intentionally blank.

Contents

This page is intentionally blank.

1. Introduction

A. Background and Objectives

Strategic competitors are using a variety of licit and illicit methods to acquire U.S. biological data. Extraction methods include: research partnerships, investments, mergers, acquisitions, cyber intrusions, and combinations of these tactics. Additionally, U.S. biological data is willingly provided to firms owned by strategic competitors through fee-for-service arrangements such as those for genetic sequencing and analytics. Such data sharing may initially be conducted for legitimate and permissible uses, but the biotechnologies using these datasets are inherently dual-use and can also enable nefarious applications, posing national security risks.[3]

An "Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy" was issued by the President on September 12, 2022.[4] The order identifies the need to:

> safeguard the United States bioeconomy, as foreign adversaries and strategic competitors alike use legal and illegal means to acquire United States technologies and data, including biological data, and proprietary or precompetitive information, which threatens United States economic competitiveness and national security.[5]

Given the threat posed by strategic competitors and the directions contained within the executive order, the Director, Science & Technology Exploitation and Analytics, Maintaining Technology Advantage (MTA) of the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) asked the Institute for Defense Analyses (IDA) to:

---

[3] You, "Safeguarding the Bioeconomy: U.S. Opportunities and Challenges."

[4] The White House, "Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for a Sustainable, Safe, and Secure American Bioeconomy," Presidential Action (Washington, DC: The Whitehouse, September 12, 2022), https://www.whitehouse.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/.

[5] Ibid, 2.

1. Develop a repeatable methodology to assess the national security risk posed by strategic competitor acquisition of U.S. biological datasets either alone or when combined with other data, and
2. Apply the methodology to representative case studies that illustrate the assessment of risk posed by a strategic competitor's acquisition of U.S. biological data to facilitate messaging across the DOD and broader National Security audiences.

B. Scope

The sponsor requested that the risk assessment methodology handle a wide variety of applications. This includes the ability for the methodology to assess the risk posed by a diverse range of biological data. IDA adopted the definition of biological data presented in the 2022 EO on advancing biotechnology: "The term 'biological data' means the information, including associated descriptors, derived from the structure, function, or process of a biological system(s) that is measured, collected, or aggregated for analysis."[6]

Given this broad definition, we developed a methodology that can assess the risk associated with any type of biological data. Examples of potential biological datasets include data relating to humans, plants, animals, microorganisms, and industrial biotechnology processes. The focus of this project was specifically on biological datasets, but the methodology is agnostic to the type of data. The framework could also be modified to assess the risk of acquisition of non-biological datasets and other enabling technologies.

In addition to considering a wide scope of types of biological data, the sponsors also requested that the methodology consider a wide range of national security implications, not just those relating to the DOD. Example national security implications could be those impacting military operations, intelligence operations, or U.S. economic competitiveness.

C. Paper Organization

Chapter 2 introduces the risk assessment methodology and details the procedure for executing it. Chapter 3 contains an illustrative case study applying the risk assessment methodology. This case study analyzes the potential risk to U.S. economic competitiveness due to the acquisition of a dataset related to genetically modified agricultural products. Seven additional case studies were analyzed and used to inform development of the risk assessment methodology. These other case studies are described in detail in a classified version of this paper.

---

[6] Ibid, 16.

2. Methodology

A. Overview

Much of the theoretical foundation of IDA's risk assessment methodology comes from the Joint Risk Analysis Methodology (JRAM).[7] One of the JRAM's purposes is to facilitate risk communication across the DOD through the use of common terminology. Therefore, we adopted many of the foundational terms and constructs established in the JRAM. Principal among these definitions is that risk is the "probability and consequence of an event causing harm to something valued."[8] Therefore, a critical component in the methodology is the identification of both the event causing harm and the valued object experiencing the harm.

For the purposes of IDA's risk assessment methodology, the event causing harm is not a strategic competitor's acquisition of some U.S. biological dataset, hereafter referred to as the *Dataset in Question*, but rather it is some use of that data.[9] In other words, we postulate that simply possessing the dataset does not directly harm U.S. national security interests and by extension does not pose a direct risk. Therefore, the first step of the risk assessment methodology is to identify a *Use of Concern*, defined as the specific application of the *Dataset in Question*, whose risk is to be assessed.

Given the wide scope of national security implications that the methodology should be able to consider, the object experiencing the harm caused by the *Use of Concern* is also user-specified. A consideration of the holistic risk posed to all facets of U.S. national security is infeasible given the diverse range of applications of a given biological dataset. Instead, users identify a specific *Operation of Interest* that serves as the object at risk. We define *Operation of Interest* as a broad term describing one of the diverse range of activities whose success is of concern to national security. Additional discussion of selecting an *Operation of Interest* is included in Section 2.D.1.

Taken together, the user-specified *Use of Concern* and *Operation of Interest* form the basis for our definition of *Risk* as: the *Likelihood* of the strategic competitor successfully

---

7 Department of Defense, *Joint Risk Analysis Methodology*, Chairman of the Joint Chiefs of Staff, CJSM 3105.01A, (Washington, DC: CJCS, October 12, 2021).

8 Ibid, B-1.

9 Terms presented in capitalized italics are components of IDA's risk assessment methodology with specific definitions. Appendix A contains a summary list of these terms and their associated definitions.

achieving the *Use of Concern* and the resulting *Consequence* to the *Operation of Interest*. We define *Likelihood* as the chance that the strategic competitor successfully achieves the *Use of Concern* within a user-specified *Timeframe of Interest*. For reasons discussed above, we used a *Timeframe of Interest* of five years for the development of the case studies. *Consequence* is defined as the impact or resulting harm to the *Operation of Interest* if the strategic competitor successfully achieves the *Use of Concern*. The risk assessment methodology assumes the *Consequence* remains the same over the *Timeframe of Interest* and does not account for *Risk* mitigating measures that may be taken in response.

To better understand the national security implications of a strategic competitor's acquisition of a given dataset, it is informative to consider not just the absolute *Risk* associated with a given *Use of Concern*, but also how the acquisition of the *Dataset in Question* changes that *Risk*. In some cases, a strategic competitor may already possess sufficient data to successfully achieve the *Use of Concern*, in which case, accessing the additional data may have diminishing risk implications. There may be cases in which the strategic competitor's access to the *Dataset in Question* is the final missing piece that enables successful achievement of the *Use of Concern*. In such a case, the data acquisition represents a significant *Driver of Risk*.

To account for this, *Likelihood* is calculated twice for a given *Use of Concern*: first assuming the strategic competitor possesses the *Dataset in Question*, and again assuming they do not. These two assessed *Likelihoods* produce two associated *Risks*. Considering the difference, or lack thereof, between the two assessed *Risks* provides insight into how essential the *Dataset in Question* is given the strategic competitor's other capabilities. However, if the *Likelihood* (or other component of the methodology, such as *Consequence*) indicates zero or very low risk, the user should cease the risk assessment or consider an alternative *Use of Concern*. Section 2.E details this component of the methodology.

Figure 1 diagrams the major components of the risk assessment methodology. The process starts with the primary input, the *Dataset in Question* (black box). Next, users identify the *Use of Concern* through consideration of a *Strategic Competitor Objective* (gray boxes).[10] The blue boxes along the top of the diagram outline the Likelihood assessment process, culminating in a categorical *Likelihood Level* (e.g., Very Likely, Likely, etc.).[11] The green boxes along the bottom of the diagram outline the *Consequence* assessment process.

As discussed in detail in Section 2.D, the *Consequence* assessment could generate two outputs. The first is the *Consequence Metric*: an objectively assessed, ideally

---

[10] We define *Strategic Competitor Objective* as a goal the strategic competitor wishes to achieve.

[11] We define *Likelihood Level* as a categorical measure of likelihood. The categories are: Very Likely, Likely, Unlikely, and Very Unlikely. See Section 2.C.3 for more details.

quantitative measure of harm to the *Operation of Interest*. The second is the *Consequence Level*: a categorical measure of *Consequence* (e.g., Minor, Modest, etc.) informed by subjective judgement of the *Consequence Metric* by operational stakeholders and subject matter experts. These two characterizations of *Consequence* generate two characterizations of *Risk* (orange boxes), the process of which is detailed in Section 2.E. If users cannot consult with operational stakeholders or experts, then the *Consequence Metric* is presented together with the assessed *Likelihood Level* as two constituent components of *Risk*.[12] If users can consult with operational stakeholders or experts, then the *Consequence Level* is combined with the *Likelihood Level* to generate a categorical *Risk Level* (e.g., High, Medium, Low).

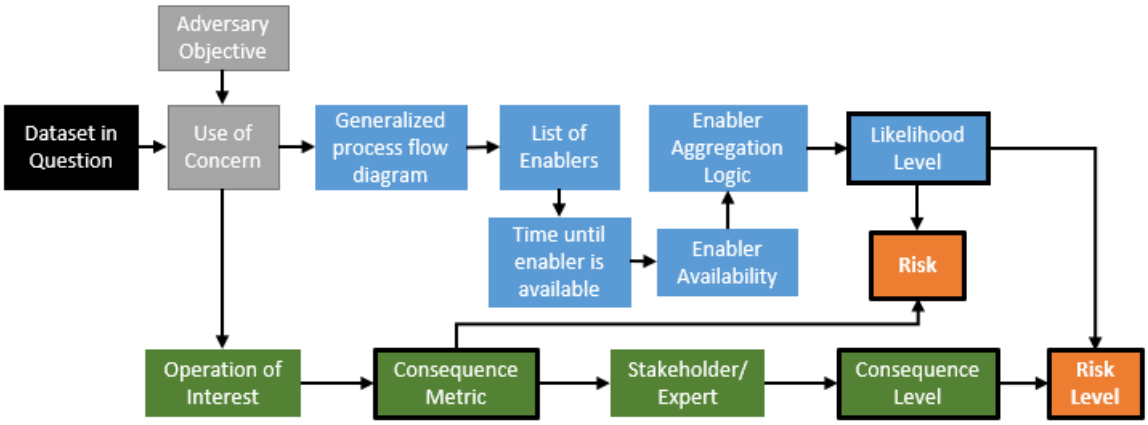


Figure 1. Overview of IDA's Risk Assessment Methodology

Note: Black: *Dataset in Question*, Gray: *Objective* and *Use of Concern*, Blue: *Likelihood* pathway, Green: *Consequence* pathway, Orange: *Risk* assessment. Methodology outputs are bolded in black.

As previously mentioned, the *Likelihood* assessment (blue boxes) is conducted twice for a given *Use of Concern*: once assuming the strategic competitor possesses the *Dataset in Question* and then assuming they do not. The two resulting *Likelihood Levels* result in two assessments of *Risk Level* (given operational stakeholder input) and *Risk* (if operational stakeholder input is unavailable). The risk assessment methodology only considers the effect of the *Availability* of the *Dataset in Question* on whether or not the strategic competitor can successfully achieve the *Use of Concern.* Recall that *Consequence* is the resulting harm if the strategic competitor successfully achieves the *Use of Concern*. Therefore, changes in the *Likelihood* associated with the *Availability* of the *Dataset in*

[12] When *Risk* is presented as the combination of the *Likelihood Level* and *Consequence Metric*, it can be summarized with a statement like: "It is Very Likely that the strategic competitor can successfully achieve the *Use of Concern* in the next five years, which would result in *x*% damage to the *Operation of Interest*."

Question do not change the *Consequence*. In other words, the *Consequence* assessment is conducted only once for a given *Use of Concern*.

Below is a list of the procedural steps constituting the risk assessment methodology procedure. Each step includes a reference to the section of the chapter where it is discussed in detail. Each of these steps is demonstrated by the corresponding numbered subsection in the Detailed Analysis of Chapter 3.

1. Identify the *Strategic Competitor Objective* (Section 2.B)
2. Identify and characterize the *Dataset in Question* (Section 2.B)
3. Identify the *Use of Concern* for the *Dataset in Question* by considering the *Strategic Competitor Objective* (Section 2.B)
4. Analyze the *Enablers*[13] required for successfully achieving the *Use of Concern* (Section 2.C)
    a. Select *Enablers* (Section 2.C.1)
        i. Create a *Generalized Process Flow Diagram*[14] depicting the major process steps associated with successfully achieving the *Use of Concern* (Section 2.C.1.a)
        ii. Identify *Enablers* associated with each step of the Process Flow Diagram by referencing the provided list of *Enabler* categories (Section 2.C.1.a)
    b. Assess *Availability*[15] of each *Enabler* under the assumption that the strategic competitor possesses the *Dataset in Question* (Section 2.C.2)
        i. Estimate a *Best Guess*, *Lower Bound*, and *Upper Bound* for the *Time until Available*[16] for each *Enabler* (Section 2.C.2.a)

---

[13] *Enablers* are capabilities and information that are required to successfully achieve the *Use of Concern*. An assessment of whether or not the strategic competitor possesses the *Enablers* serves as the foundation of the *Likelihood* assessment.

[14] A *Generalized Process Flow Diagram* is a schematic depiction of the major process steps associated with successfully achieving the *Use of Concern*. It guides users in the identification of *Enablers*.

[15] *Availability* refers to the likelihood that the strategic competitor possesses the *Enabler* within the *Timeframe of Interest*.

[16] The *Time until Available* is a set of estimated times it will take for the strategic competitor to possess a given *Enabler*. Its range of possible values is accounted for by the user making a *Best Guess* and providing *Lower* and *Upper Bounds* on the estimate. In some cases, the *Upper Bound* may be infinite and require special consideration (Section 2.C.2.a.2).

ii. Use the time estimates to generate a probability distribution describing the uncertainty in the assessed *Time until Available* for each *Enabler* (Section 2.C.2.a)

iii. Determine the *Availability* of each *Enabler* by integrating the associated distribution from zero to the *Timeframe of Interest* (Section 2.C.2.a)

5. Assess *Likelihood* (Section 2.C)
    a. Calculate the numeric *Likelihood Metric*[17] as the minimum *Availability* across all *Enablers* under the assumption that the strategic competitor does not possess the *Dataset in Question* (Section 2.C.3)
    b. Convert the *Likelihood Metric* into the corresponding categorical *Likelihood Level* (Section 2.C.3)
    c. Repeat steps 4.b, 5.a, and 5.b under the assumption that the strategic competitor does not possess the *Dataset in Question*
6. Assess *Availability* of each *Enabler* and *Likelihood* under the assumption that the strategic competitor possesses the *Dataset in Question* (Section 2.C.2)
    a. Repeat steps 4.b, 5.a, and 5.b under the assumption that the strategic competitor does not possess the *Dataset in Question*
7. Assess *Consequence* (Section 2.D)
    a. Identify an *Operation of Interest* (Section 2.D.1)
    b. Identify and evaluate a *Consequence Metric* (Section 2.D.2)
    c. Convert the *Consequence Metric* to a categorical *Consequence Level* through consultation with subject matter experts or stakeholders (Section 2.D.3)
8. Assess *Risk* (Section 2.E)
    a. Combine the *Consequence Level* with the *Likelihood Level* determined under the assumption that the strategic competitor possesses the *Dataset in Question* to determine the *Risk Level* (Section 2.E)

[17] The *Likelihood Metric* is a numeric measure of *Likelihood* that ranges from 0 (least likely) to 1 (most likely).

b. Repeat the previous step using the *Likelihood Level* determined under the assumption that the strategic competitor does not possess the *Dataset in Question* to determine the *Risk Level* (Section 2.E)

c. Characterize the change in *Risk Level* associated with the strategic competitor's access to the *Dataset in Question* by comparing the *Risk Levels* determined in the previous two steps (Section 2.E)

d. If a *Consequence Level* was not determined, present *Risk* in its two constituent components: *Likelihood Level* and *Consequence Metric* (2.E)

9. Identify *Drivers of Risk*[18] (Section 2.F)

B. Identify Use of Concern

As shown in Figure 1, the first step of the methodology is to identify the *Use of Concern*. The *Use of Concern* is identified by considering a *Strategic Competitor Objective*, defined as a goal the strategic competitor wishes to achieve. When identifying the *Strategic Competitor Objective*, it is important to consider both what the strategic competitor is stating publicly (e.g., policy documents or leadership speeches), and what is reported by intelligence. When possible, the strategic competitor's publicly stated objectives should be corroborated by intelligence to both refine the user's understanding of that goal and confirm the strategic competitor's commitment to achieving said goal.

Not all *Uses of Concern* can be directly tied to strategic competitor statements, be they public or internal. Intelligence assessments and other national security analytic products can inform the identification of uses that may be concerning but are not associated with strategic competitor statements. Users may not be aware of all applicable *Strategic Competitor Objectives*. This may be due to incomplete intelligence reporting or a lack of access or available time to review relevant intelligence products. Users with familiarity of a strategic competitor's goals or access to those who do will be able to generate more informative risk assessments.

Chapter 3 contains examples of *Strategic Competitor Objectives* and their associated sources. Users of the methodology are encouraged to reference these *Strategic Competitor Objectives* and associated citations; they may apply directly to other *Datasets in Question* or may serve as the starting point for identifying an applicable *Strategic Competitor Objective*.

It is worth noting that a given *Dataset in Question* likely has multiple and potentially diverse *Uses of Concern*, each with its own, potentially unique, associated *Risk*. Therefore,

---

[18] *Drivers of Risk* are capabilities or data that the *Likelihood* is highly sensitive to.

a complete characterization of the *Risks* associated with a given *Dataset in Question* involves characterizing the range of *Risks* associated with various potential *Uses of Concerns*. It is likely time-prohibitive—if not impossible—for a user to identify all potential *Uses of Concern* associated with a *Dataset in Question*. Similarly, each *Use of Concern* can harm multiple and potentially diverse *Operations of Interest*, which users may not be aware of or able to assess. Section 2.F discusses this limitation in more detail.

Ultimately, the inability to assess the myriad ways in which the strategic competitor could use a biological dataset is a limitation of the risk assessment methodology. The implication of this limitation relates to how quickly the methodology can be executed by a user. The faster the methodology can be executed, the more *Uses of Concern* can be considered—ultimately providing a more complete depiction of risks. We did not assess the speed at which users outside of the IDA team could execute the methodology; however, such an assessment would be central to a follow-on analysis of the methodology's usability.

C. Assess Likelihood

As stated in the previous section, our definition of *Likelihood* is: the chance that the strategic competitor successfully achieves the *Use of Concern* in the *Timeframe of Interest*, and the process for its assessment is shown as the blue boxes along the top of Figure 1. It is important to clarify that the definition of *Likelihood* excludes an assessment of the probability that the strategic competitor would choose to implement the specified *Use of Concern*. In some cases, there may be multiple ways for the strategic competitor to achieve a given objective. Such alternatives are not considered in the risk assessment methodology. Readers should not conclude that the *Uses of Concern* presented in the case studies represent the most likely course of action a strategic competitor would take.

Additionally, the methodology assumes the strategic competitor will pursue the *Use of Concern* for the duration of the *Timeframe of Interest*. It does not account for the probability that the strategic competitor may abandon the effort at some point during the *Timeframe of Interest*. The validity of this assumption likely breaks down for longer *Timeframes of Interest* over which the strategic competitor is more likely to abandon or redirect unsuccessful efforts.

The methodology's assessment of *Likelihood* centers on the consideration of an analytic construct called *Enablers*. *Enablers* are capabilities and information that are required to successfully achieve the *Use of Concern*. Different *Uses of Concern* for a given *Dataset in Question* may have different *Enablers*. Similar *Uses of Concern* may have similar *Enablers*, so users should consider referencing an existing set of *Enablers* as a starting point for alternative *Uses of Concern*. However, care should be taken to ensure the existing *Enablers* are applicable to the alternative *Use of Concern* and assess if additional *Enablers* are required. Furthermore, a single *Use of Concern* may have multiple sets of

Enablers if there are multiple ways of achieving the use. The fact that the methodology only considers one set of *Enablers* at a time is a limitation of the methodology.

The consideration of *Enablers* to derive a *Likelihood* estimate is a three-step process. The first step is to identify the set of *Enablers*. The second step is to assess the *Availability* of each *Enabler*. An *Enabler*'s *Availability* is defined as the chance that the strategic competitor will possess the *Enabler* within the *Timeframe of Interest*. The final step is to aggregate the *Availabilities* of all the *Enablers* to generate a final estimate for the *Likelihood* of the strategic competitor successfully achieving the *Use of Concern*. Each of these steps are detailed in the subsequent sections.

The process of identifying and assessing the *Enablers* is a substantial analytic effort. The IDA team observed that the majority of the time executing the methodology was spent on assessing the *Enablers*. The overall accuracy of the methodology depends on the user's ability to accurately identify and assess the *Enablers*.

1. Identify Enablers

The first step in considering *Enablers* is to identify them. The process of identifying the *Enablers* for a given *Use of Concern* requires a certain level of familiarity with the scientific and technical concepts at play. Users who lack the requisite expertise should consult subject matter experts to ensure the list of *Enablers* is both complete and does not include non-essential elements. The risk assessment methodology involves a structured approach to help guide users in identifying pertinent *Enablers*. This approach involves 1) dividing the *Use of Concern* into discrete processes and identifying the associated *Enablers* for each process, and 2) providing a predefined list of *Enabler* categories for the user to reference. Taken together, these two components provide the user with a starting point to guide their research.

When generating the list of *Enablers,* it is important that users consider that each *Enabler*, by definition, is necessary for successfully achieving the *Use of Concern*. That is, the *Use of Concern* cannot be achieved if any one or more *Enabler* is missing. Furthermore, the set of all *Enablers* for a given *Use of Concern* must be sufficient. That is, it must completely account for all necessary capabilities and information. Users may encounter capabilities or information that may contribute to successfully achieving the *Use of Concern*, but in and of themselves are not essential, in which case they would not be considered an *Enabler*. However, these non-essential capabilities or information are still worth noting as they may be informative to assessing the *Availability* of the *Enablers* to the strategic competitor.

a. Generalized Process Flow Diagram

As mentioned above, the process of identifying the *Enablers* involves breaking down the *Use of Concern* into constituent processes. This is accomplished through the creation of a *Generalized Process Flow Diagram*, which we define as a schematic depiction of the major process steps associated with successfully achieving the *Use of Concern*. Figure 2 shows an example for a notional *Use of Concern*. The *Generalized Process Flow Diagram* for an actual *Use of Concern* will have its own process steps related to that specific use. The example below is intended to generally represent the scope of the number and level of specificity of each step.

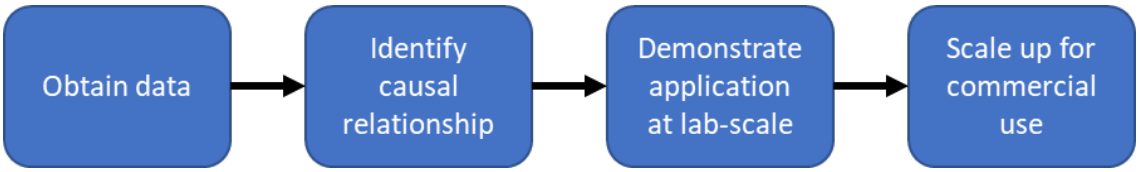


Figure 2. Notional *Generalized Process Flow Diagram*

Ideally, the *Generalized Process Flow Diagram* is sufficiently generalized so the diagram created for one *Use of Concern* can be directly used or readily adapted for another. However, the identified processes should also provide specific actionable guidance on what types of *Enablers* a user should consider. The *Generalized Process Flow Diagram* included in Chapter 3 and the other case studies presented in the classified version of this paper demonstrate an appropriate balance of generalizability and specificity. The diagrams in these case studies may potentially work as-is for other *Uses of Concern*, or may serve as a point of departure for generating new diagrams.

Users then identify the associated *Enablers* for each process step in the *Generalized Process Flow Diagram* by considering the categories in the list below. The set of *Enablers* must include all the data required to successfully achieve the *Use of Concern*. In some cases, the *Dataset in Question* is sufficient for the use, but in other cases, a strategic competitor may require additional data. In addition to data, the *Enabler* categories include the technological capabilities, scientific knowledge, and infrastructure required for the *Use of Concern*.

Data

- **Additional observations of the same type of data**: This type of *Enabler* includes other datasets containing data with the same characteristics as the *Dataset in Question*. Datasets in this category could be directly combined with the dataset of interest for a specific use. For example, if the *Dataset in Question* contains genetic single nucleotide polymorphism (SNP) data for a given set of individuals, then the same type of genetic data for other individuals not in this

dataset would be considered as this type of *Enabler*. This category of *Enabler* occurs in cases where the *Dataset in Question* alone is insufficient to successfully achieve the *Use of Concern*.

- **Related data of other types**: In some cases, the *Dataset in Question* alone is insufficient for achieving the *Use of Concern* because it does not capture all of the required types of information. In these cases, the strategic competitor will require related data of other types. Data in this category may not have the exact same characteristics as the *Dataset in Question*, but still help to achieve the *Use of Concern*. These may include datasets from different domains, datasets having different features, or datasets which may help complete the *Use of Concern* at a different step than the dataset of interest. For example, if the *Dataset in Question* consists of SNP data for a set of individuals, then short tandem repeat (STR) data for those same individuals could be considered as this type of *Enabler*.
- **Ability to generate data**: These *Enablers* highlight the capability of the strategic competitor to generate additional observations of the same type of data, or related data of other types. Generation of data may be primary generation such as experimentation and making observations, or secondary generation such as acquiring datasets from other entities. If the *Dataset in Question* is SNP data for a set of individuals, the ability to obtain and process SNP data on additional individuals would be an example of this type of *Enabler*.

Technological capabilities

- **Computational capability**: These *Enablers* capture the computational power required to achieve the *Use of Concern*. This includes both raw computing power (i.e., computations performed by processors) and other essential hardware capabilities, such as the ability to store generated data.
- **Machinery/tools capability:** These *Enablers* consist of equipment which may help achieve the *Use of Concern*. These include equipment both commercially available or developed by government entities.
- **Analytic capability:** These *Enablers* describe the analytic algorithms, computer programs, models, and simulations, for example, that are required to achieve the *Use of Concern*. While similar to **computational capability**, **analytic capability** focuses on software as compared to hardware.

Scientific knowledge

- **Technical skills and capabilities in use domain:** These *Enablers* highlight the technical skills and capabilities associated with the domain. This may include

applications similar to the specified use, and can act to identify existing workflows for development of the *Use of Concern*.

- **Personnel with requisite scientific knowledge:** These *Enablers* may be individuals, teams, or entire institutions who possess the requisite understanding of the scientific phenomenon associated with the *Use of Concern*.

Infrastructure

- **Academic, governmental, and industrial institutions**: These *Enablers* consist of the non-human capabilities that institutions provide. *Enablers* such as professional networks and access to funding and other incentives fall into this category.
- **Raw materials:** These *Enablers* consist of the physical materials required to achieve the *Use of Concern*. This may include access to reagents, pre-packaged kits for workflows, or specific minerals required for the use.
- **Regulatory environment:** These *Enablers* consist of legislation and initiatives that can catalyze the development of the *Use of Concern*.

The *Enabler* categories presented above may not be sufficiently exhaustive for all *Uses of Concern*. Additionally, some *Enablers* may not neatly fit into just one or any of the categories. These categories are intended to facilitate, not constrain, the process of generating the set of *Enablers*.

One final note on how the user should characterize the identified *Enablers*. For some *Enablers,* simply asking "is this capability or information essential for successfully achieving the *Use of Concern*?" is inadequate. Instead, the question should be "how much of this capability or information is essential for successfully achieving the *Use of Concern*?" This is particularly true for data *Enablers*. Just because the strategic competitor possesses pertinent data, does not necessarily mean they have a sufficient quantity of that data.

Therefore, when identifying these *Enablers*, users should quantify—to the extent possible—not just the type of data that is required, but also the quantity needed to successfully achieve the *Use of Concern*. This quantification of *Enablers* plays an important role in determining whether the *Dataset in Question* contains a sufficient quantity of data to successfully achieve the *Use of Concern*, or whether the strategic competitor requires additional data.

Even with the use of the *Generalized Process Flow Diagram* and the categories provided above, correctly identifying all *Enablers* for a given *Use of Concern* will likely depend on the user's available time and level of expertise. We have not assessed the sensitivity of the methodology regarding the completeness or accuracy of the identified

Enablers. Such an assessment is being considered as part of a follow-on assessment of the usability of the risk assessment methodology.

2. Assess Availability of each Enabler

Following the identification of the *Enablers,* the next step of the methodology is to assess the *Availability* of each *Enabler*. As defined above, an *Enabler's Availability* is the likelihood that the strategic competitor will possess it within the *Timeframe of Interest.* The *Availability* of an *Enabler* is quantified on a scale from 0 (least likely) to 1 (most likely).

Recall that *Likelihood* is assessed twice for a given *Use of Concern*: once assuming that the strategic competitor does not possess the *Dataset in Question* and then again assuming they do. To do this, the methodology considers how the strategic competitor's access to the *Dataset in Question* changes the *Availability* of each *Enabler*, and in turn the *Likelihood* of successfully achieving the *Use of Concern*.

a. Assess Time until Available

The process of assessing an *Enabler's Availability* centers on estimating its *Time until Available*, which we define as how long it will take the strategic competitor to possess a given *Enabler*. As implied by its definition, the *Availability* of an *Enabler* is a probabilistic event and accordingly so is the *Time until Available*. To reflect the uncertainty in how long it may take a strategic competitor to possess an *Enabler*, the risk assessment methodology uses three estimates for the *Time until Available*:

1. A *Best Guess*: a reasonable best estimate of when the strategic competitor would possess the *Enabler,*
2. A *Lower Bound*: a reasonable estimate for the soonest time at which the strategic competitor would possess the *Enabler,* and
3. An *Upper Bound*: a reasonable estimate for the latest time at which the strategic competitor would possess the *Enabler*.

These three estimates for the *Time until Available* are then combined to form a distribution to characterize the likelihood of the strategic competitor possessing the *Enabler* in the *Timeframe of Interest*. This process is described in the next section.

For some *Enablers*, it may not be possible to estimate a finite *Upper Bound* on the *Time until Available*. This is particularly true for *Enablers* that depend on yet-to-be demonstrated scientific phenomenon, or if the *Use of Concern* requires a dataset that cannot be generated. In such a case, the *Lower Bound* and *Best Guess* may be estimated based on how long the requisite scientific experiments will take; however, there is no guarantee that those experiments will generate positive results. Even if the scientific line of inquiry is pursued for years on end, if the hypothesis to be proven is physically impossible, then the

strategic competitor can never possess the *Enabler*. In these cases, the *Upper Bound* for the *Time Until Available* is considered infinite. The quantification of the *Availability* of these *Enablers* requires special consideration (see Section 2.C.2.a.2).

If the strategic competitor already possesses a given *Enabler*, all three estimates for the *Time until Available* are trivially set to zero years. For *Enablers* that the strategic competitor does not possess, the user must generate the required time estimates.

The process of estimating the *Time until Available* will likely vary from one *Enabler* to another given the diversity of *Enablers* that may be involved. We provide general guidance and suggestions for generating these estimates; however, the ultimate accuracy of this step of the methodology may depend on the user's level of expertise, their access to both open and classified literature, and their available time. We have not assessed how well a user with constrained expertise or time could assess the *Time until Available* for each *Enabler* in the model, nor have we assessed how sensitive the final risk assessment is to errors in this process. Both should be considered as part of a follow-on usability assessment.

Throughout this process, users should bear in mind that the ultimate objective of understanding is whether the *Enabler* will be available during the *Timeframe of Interest* (e.g., five years). Therefore, precise estimates down to a timeframe of less than a few months is not needed. Similarly, accuracy in estimating times greater than five years is not critical. Ultimately, we incorporated multiple point estimation (i.e., *Best Guess*, *Lower Bound*, and when applicable, *Upper Bound*) to capture the uncertainty in this estimation and reduce the sensitivity of the final assessed *Risk* to errors in this estimation process.

One potential approach to estimate *Time until Available* is the use of analogous situations. For example, if the U.S. already possesses the *Enabler*, then how long did that process take? Could a strategic competitor achieve the same progress in the same amount of time, sooner, or later? Alternatively, the user can consider analogous technologies or data applications. Perhaps the application of interest to the *Enabler* has been demonstrated in a different context. If so, how long did it take to develop that application? Would you expect the development of the analogous application to be faster, slower, or about the same as the *Enabler* in question?

For all of these cases, the applicability of an analogous situation depends on the level of similarity to the *Enabler*. Users can increase the range between the estimated *Lower* and *Upper Bounds* to reflect the uncertainty in their estimate. For an example of using analogous situations to inform the estimate of *Time until Available,* see the use of the regulatory timelines for genetically engineered cotton and papaya to inform the regulatory approval timeline of gene-edited pigs in Section 3.B.4.b.8.

Another technique users may consider to inform their estimation of the *Time until Available* is to break down a given *Enabler* into constituent steps and assess each step

independently. With this approach, the *Best Guess* and the *Lower* and *Upper Bounds* can be estimated for each constituent step and then aggregated to develop an overall estimate range for the *Enabler*. For an example of this approach, see the breakdown of developing gene-resistant pigs into the constituent parts of basic research, gene delivery, and pig reproduction and development in Section 3.B.4.b.2.

The open scientific literature will likely serve as an important source for assessing the *Time until Available*. Ideally, users will consult native language literature of strategic competitors, too. Accessing translations of native language sources is substantially more difficult than English language sources. A user with constrained time and resources may not be able to conduct a review of these sources. In addition to open sources, intelligence reporting on the strategic competitor is an important source for assessing the *Time until Available*.

Recall that the assessment of *Likelihood* is conducted twice: once assuming that the strategic competitor does not possess the *Dataset in Question* and once assuming that they do. To that end, users must assess each *Enabler's Time until Available* twice. First, the *Time until Available* is assessed assuming the strategic competitor does not have access to the *Dataset in Question*. This assessment is based on what information and capabilities the strategic competitor currently possesses or those that can be internally developed by the strategic competitor within the *Timeframe of Interest*.

Next, the assessed *Time until Available* is updated to reflect the strategic competitor gaining access to the *Dataset in Question.* For some *Enablers*, the strategic competitor's access to the *Dataset in Question* will directly impact its *Time until Available*. For example, if the *Dataset in Question* contains all the necessary data for the *Use of Concern*, then under the assumption that the strategic competitor possesses the *Dataset in Question*, the *Time until Available* for the data *Enabler* becomes zero (i.e., the strategic competitor is guaranteed to possess the *Enabler*).

However, for some *Enablers*, the strategic competitor's access to the *Dataset in Question* will indirectly change that *Enabler's Availability*. For example, possessing the *Dataset in Question* may increase the chances that the strategic competitor develops some capability during the *Timeframe of Interest*. In this case, the *Time until Available* for that *Enabler* would be decreased to reflect how acquiring the *Dataset in Question* makes it easier for the strategic competitor to possess the *Enabler*.

Finally, the *Availability* of some *Enablers* may not be affected by the strategic competitor's acquisition of the *Dataset in Question.* For these *Enablers,* their assessed *Time until Available* would be the same whether or not the strategic competitor acquires the *Dataset in Question*. Chapter 3 provides examples of how the *Times until Available* are adjusted to reflect the strategic competitor's acquisition of the *Dataset in Question*.

There is one final caveat on how we assessed the *Availability* of the *Enablers*. Our assessments of *Enabler Availability* did not consider the possibility that the strategic competitor could steal or otherwise acquire the *Dataset in Question* or similar data from an alternative protected source. For example, if the *Dataset in Question* is held by a U.S. company, we did not consider the possibility of the strategic competitor acquiring similar data from a different U.S. company. For example, users would need to know the cyber security vulnerabilities of the company holding the data and the related capabilities of the strategic competitor to determine how long it would take a strategic competitor to possess the data through nefarious means.

After the user assesses the *Time until Available* for each *Enabler* the next step is to convert the three time estimates (i.e., the *Best Guess, Lower Bound*, and *Upper Bound*) into a distribution that describes these time ranges for each *Enabler*. These distributions are then used to determine the *Availability* of each *Enabler*. The form of the probability distribution depends on if the *Upper Bound* for the *Time until Available* is finite or infinite.

1) Enablers with Finite Upper Bound on the Time until Available

As implied by its definition, the risk assessment methodology considers an *Enabler's Availability* as a probabilistic event. That is, there is uncertainty in estimating the time at which the strategic competitor will possess a given *Enabler*. A common and long-established method for estimating probabilistic activity times is the use of the program evaluation and review technique (PERT) distribution.[19] The PERT distribution has a general bell shape and is uniquely specified by a minimum, maximum, and most likely value. The PERT distribution is a transformation of the four-parameter Beta distribution with a probability density function (PDF) of

$$f_{PERT}(x) = \frac{(x-1)^{\alpha-1}(c-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)(c-a)^{\alpha+\beta-1}}$$

$$\alpha = 1 + 4\frac{b-a}{c-a}$$

$$\beta = 1 + 4\frac{c-b}{c-a}$$

where a, b, and c are the minimum, most likely, and maximum values of the distribution and B is the beta function. PERT distributions are readily available in many computational software, including Microsoft Excel.[20] Figure 3 illustrates representative PERT

---

[19] S Mohan et al., "A Lognormal Approximation of Activity Duration in PERT Using Two Time Estimates," *Journal of Operational Research Society* 58, no. 6 (December 21, 2017): 827-831, https://www.tandfonline.com/doi/full/10.1057/palgrave.jors.2602204.

[20] In Excel, the PDF for a PERT distribution is calculated using the equation $f_{PERT}(x) =$ BETA.DIST$(x, \alpha, \beta,$FALSE$, a, c)$, where $\alpha, \beta, a,$ and c are all defined as above.

distributions each with a minimum of 1, a maximum of 5, and a most likely value of 2 (blue), 3 (orange), or 4 (gray).

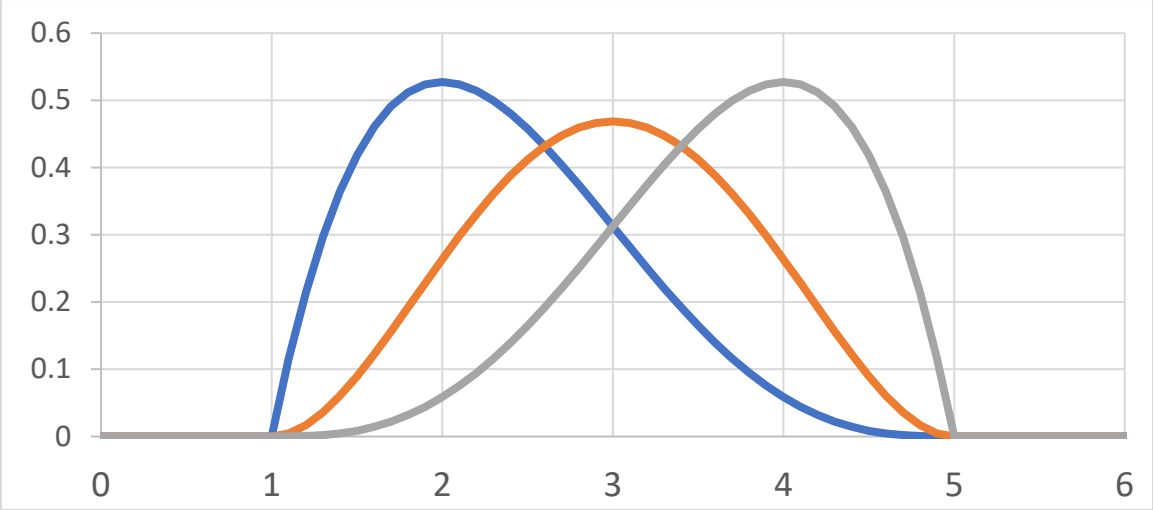


Figure 3. Representative PERT Distributions with a Minimum of 1, a Maximum of 5, and most likely Values of 2 (blue), 3 (orange), and 4 (gray).

For each *Enabler*, the *Best Guess*, *Lower Bound*, and *Upper Bound* estimates for the *Time until Available* are used as the mode, minimum, and maximum of a PERT distribution to define a PDF that describes the variation in the estimated *Time until Available*.

The three parameters of the PERT distribution must satisfy the following constraint: $a < b < c$. However, users may assess the *Lower Bound* is the same as the *Best Guess*. In this case, the *Best Guess* should be increased by the small value of 0.01 to ensure the constraints on the PERT distribution parameters are satisfied. Similarly, in the event that the *Best Guess* is assessed to be the same as the *Upper Bound*, the *Upper Bound* should be increased by the same value of 0.01.[21]

Finally, in the event that the user assesses all three estimates to be the same, then the distribution becomes a point estimate at the single assessed value. If the point estimate is less than the *Timeframe of Interest*, then the *Availability* is 1. Whereas, if the point estimate is greater than the *Timeframe of Interest*, then the *Availability* is 0.

The *Availability* of each *Enabler* is then determined by integrating the associated PDF from 0 to the *Timeframe of Interest*, or equivalently, evaluating the associated cumulative density function (CDF), $F_{PERT}(x)$, at the *Timeframe of Interest.*[22] In other words, the PERT distribution describes the range of possible times at which the strategic competitor would

[21] IDA did not assess the sensitivity of the methodology to the choice of the small value.

[22] In Excel, the CDF for a PERT distribution is calculated using the equation $f_{PERT}(x) =$ BETA.DIST$(x, \alpha, \beta,$TRUE$, a, c)$, where $\alpha, \beta, a,$ and c are all defined as above.

come to possess the *Enabler*, and the area under that distribution up to the *Timeframe of Interest* (e.g., five years) represents the probability of the strategic competitor possessing that *Enabler* in that time period. Figure 4 depicts an example of an *Enabler* with an assessed *Best Guess* for *Time until Available* of four years and a *Lower* and *Upper Bound* of one and six years, respectively. In this example, the *Enabler's Availability* is 0.90. The process for implementing these calculations in Microsoft Excel is included in Appendix B.

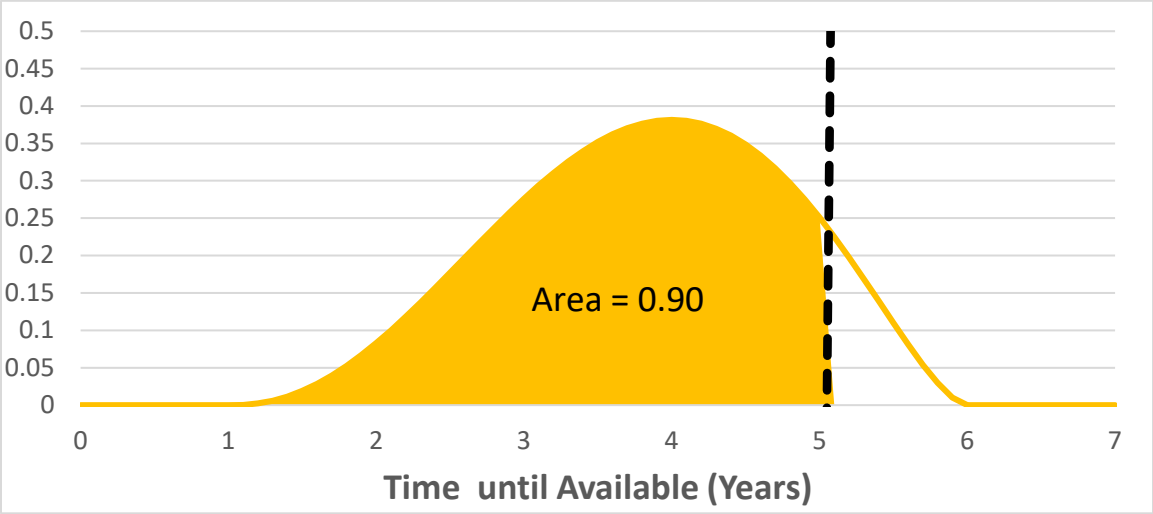


Figure 4. Notional Distribution for an *Enabler* with a *Time until Available* of 4 (1-6) years and a Corresponding *Availability* of 0.90.

2) Enablers with Infinite Upper Bound on the Time Until Available

Recall that some *Enablers* may have a nonfinite *Upper Bound* on their *Time until Available.* In these cases, the unbounded range on the *Time until Available* is incompatible with the PERT distribution, which requires a finite value for the maximum. Therefore, an alternate distribution is required. Mohan et al. advocate using lognormal distributions in lieu of PERT distributions to describe probabilistic activity times when only two estimates are known—the most likely and either the maximum or minimum.[23]

The use of lognormal distributions in this manner for characterizing the variation in the *Time until Available* is appealing for two reasons. First, the PDF of a lognormal distribution extends to positive infinity. Therefore, it can adequately handle the situation in which an *Enabler* lacks a finite *Upper Bound*. Second, the required parameters—the minimum and most likely—are already captured as the *Lower Bound* and *Best Guess* for the *Time until Available.* Therefore, we replace the use of a PERT distribution with a lognormal distribution for *Enablers* that lack a finite *Upper Bound*. Once the distribution

[23] Mohan et al., "A Lognormal Approximation of Activity Duration in PERT Using Two Time Estimates."

has been characterized, the process of calculating the *Availability* by evaluating the CDF at the *Timeframe of Interest* remains the same as described above.

The PDF for the lognormal distribution used in the risk assessment methodology is

$$f_{lognormal}(x) = \frac{1}{x\sigma^*\sqrt{2\pi}} e^{\frac{-(\ln(x)-\mu^*)^2}{2\sigma^{*2}}}$$

$$\sigma^* = \frac{1}{2}z - \left[\frac{1}{4}z^2 + \log\left(\frac{a}{b}\right)\right]^{1/2}$$

$$\mu^* = \ln(a) + z\sigma^*$$

where a and b continue to be the minimum and most likely values, and z is the number of standard deviations between the minimum and most likely value. Following the guidance of Mohan et al., a value of $z = 3$ is used in the risk assessment methodology.[24] The formulation of the lognormal distribution above places four constraints on acceptable parameter values.[25]

1. If the estimate for the *Lower Bound* is 0 (i.e., $a = 0$), then σ^* and μ^* are undefined due to $\ln(0)$. In this case, users should increase the estimate of the *Lower Bound* by the small value of 0.01.
2. If the estimate for the *Lower Bound* is the same as that of the *Best Guess*, then $\sigma^* = 0$, which results in a division by zero in the PDF of the lognormal distribution. In this case, users should increase the estimate of the *Best Guess* by 0.01. In the event that the *Lower Bound* and the *Best Guess* are both assessed to be 0 (i.e., $a = b = 0$), then the *Lower Bound* should be set at 0.01 and the *Best Guess* should be set to 0.02 to ensure constraint 1) is satisfied.
3. The formulation of σ^* requires $\frac{1}{4}z^2 + \log\left(\frac{a}{b}\right) > 0$ to avoid taking the square root of a negative number. This occurs when the assessed value for the *Lower Bound* and the *Best Guess* satisfy the following inequality $z < \sqrt{-4\ln(a/b)}$. If this occurs, users should set $z = \sqrt{-4\ln(a/b)}$.[26]

As is the case for the PERT distribution, the PDF and CDF of a lognormal distribution can readily be calculated in numerous computational software to include Microsoft Excel. Figure 5 shows an example of a lognormal distribution with a minimum of 1 and a most

---

[24] IDA did not assess the sensitivity of the methodology to the choice of value for z.

[25] IDA did not assess the sensitivity of the methodology to these choices of how to ensure valid parameter values.

[26] A small value of 0.001 is added to z to ensure floating point numerical imprecision does not result in $z < \sqrt{-4\ln(a/b)}$.

likely value of 3 (blue) in addition to a PERT distribution with the same minimum and most likely value but a maximum of 7 (orange).

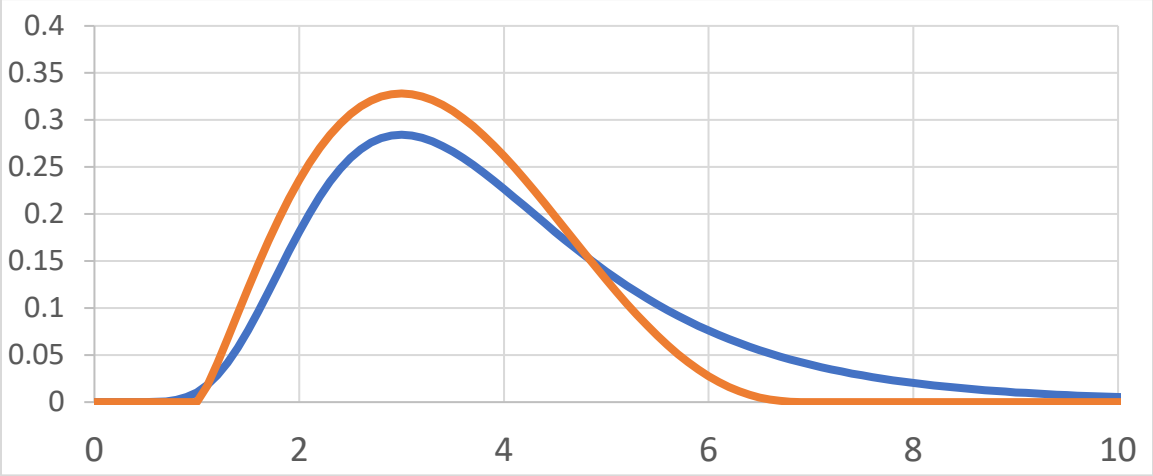


Figure 5. Representative Lognormal Distribution (blue) and PERT Distribution (orange).

A limitation with using lognormal distributions to describe the *Time until Available* is that the *Availability* of the *Enabler* will asymptotically approach 100 percent with increasing *Timeframe of Interest*. However, the strategic competitor may never possess these types of *Enablers,* given they rely on yet-to-be demonstrated scientific phenomenon. Therefore, the *Availability* of these *Enablers* in the long term should approach some value less than 100 percent. This limitation is mitigated through the consideration of only short *Timeframes of Interest* (e.g., five years). Users should not consider longer *Timeframes of Interest* as the methodology may overestimate the *Availability* of *Enablers* with infinite *Upper Bounds*.

3) Discussion on Assessing Enabler Availability

The process of assessing the *Availability* of each *Enabler* is one of the most labor-intensive portions of the methodology. We believe that generating multiple estimates for the *Time until Available* for a given *Enabler*, while research intensive, is a rigorous way to account for the uncertainty associated with forecasting strategic competitor capability development. Users could consider adopting an alternative approach in which they directly assess the *Availability* of each *Enabler* within the *Timeframe of Interest* without consideration of the *Time until Available*.

That is, a user would answer the question "does the strategic competitor already possess this capability or can they acquire it in the *Timeframe of Interest*?" instead of generating estimates for a *Best Guess*, *Lower Bound*, and *Upper Bound*. It may be the case that simplifying this step provides a meaningful improvement in the methodology's

usability at a minimal loss of accuracy. However, we neither employed this simplification nor assessed its impact on the methodology's results.

At this point in the methodology the user has a set of *Enablers* describing the information and capabilities that the strategic competitor must have to achieve the *Use of Concern*. Each of these *Enablers* has two assessed *Availabilities.* The first describes the *Likelihood* of the strategic competitor possessing that *Enabler* assuming the strategic competitor does not possess the *Dataset in Question*. The second assessed *Availability* describes the *Likelihood* of the strategic competitor possessing that *Enabler* assuming they do possess the *Dataset in Question*.

Reviewing the assessed *Enabler Availabilities* provides insight into potential *Enabler(s)* or other capabilities that the *Likelihood* of the strategic competitor achieving the *Use of Concern* is highly sensitive to. For example, *Enablers* with low assessed *Availabilities* (i.e., close to 0) represent bottlenecks that may prevent a strategic competitor from achieving the *Use of Concern*. Such capabilities or datasets are candidates for efforts to prevent access by the strategic competitor.

Conversely, *Enablers* with high assessed *Availabilities* (i.e. close to 1) are unlikely to prevent the strategic competitor from achieving the *Use of Concern*. Efforts to prevent the strategic competitor's access to these capabilities or dataset may not be impactful enough to prevent the specified *Use of Concern*. Consideration of the *Availabilities* of a set of *Enablers* in this fashion provides insight into the *Drivers of Risk* for a given *Use of Concern*. See Section 2.F for additional discussion of *Drivers of Risk.*

It is critical to remember that both the *Enablers* and their assessed *Availabilities* are assessed in the context of a specific *Use of Concern*. A different use of the *Dataset in Question* may have different *Enablers* with different *Availabilities*. Therefore, an *Enabler* that has no impact on *Likelihood* for one *Use of Concern* may be a bottleneck for another.

3. Aggregate the Availability of all Enablers to Determine Likelihood

The final step in assessing the *Likelihood* of the strategic competitor achieving the *Use of Concern* is to aggregate the *Availability* of the *Enablers.* A *Likelihood Metric* is calculated as the minimum *Availability* across all of the *Enablers.* The resulting *Likelihood Metric* is then converted into a categorical *Likelihood Level* as shown in Table 1. The *Likelihood Levels* and associated *Likelihood Metric* ranges are adapted from those presented in the JRAM.[27]

[27] Department of Defense, *Joint Risk Analysis Methodology*.

Table 1. *Likelihood Levels*

Likelihood Level	***Likelihood Level***
[0.0-0.2)	Very Unlikely
[0.2-0.5)	Unlikely
[0.5-0.8)	Likely
[0.8-1.0]	Very Likely

The process of aggregating *Enabler Availability* is done twice for each *Use of Concern*: once using the *Enabler Availabilities* assessed and assuming that the strategic competitor does not possess the *Dataset in Question*, and then a second time assuming they do. Accordingly, two assessed *Likelihood Levels* are obtained. The difference between these two *Likelihood Levels* represents how the strategic competitor's acquisition of the *Dataset in Question* changes the *Likelihood* of them achieving the *Use of Concern*. In some cases, acquiring the *Dataset in Question* may have a large impact on the *Likelihood*. This would be expected in cases where the strategic competitor already possesses all other capabilities needed for the specified use, and the *Dataset in Question* is the one missing piece. In such cases, safeguarding the *Dataset in Question* is essential in preventing that *Use of Concern*.

There may be cases in which the strategic competitor's acquisition to the *Dataset in Question* does not change the *Likelihood* for a given *Use of Concern*. This may be because the strategic competitor can already achieve the *Use of Concern* with their current data and capabilities. Or, this may be due to a lack of *Availability* of some other *Enabler* that prevents the *Use of Concern* even when the strategic competitor acquires the *Dataset in Question.*

The methodology as documented above generates a change in *Likelihood* that reflects the impact of the strategic competitor acquiring the *Dataset in Question*. However, this process of generating a change in *Likelihood* can also be extended to any *Enabler* associated with *Use of Concern*. A user can execute the methodology assuming that any given *Enabler* either is or is not available to the strategic competitor and examine the resulting change in *Likelihood* and resulting *Risk*. By doing this, a user would have a better understanding of *Drivers of Risk* for the *Use of Concern*. That being said, the case studies presented here only consider the change to *Likelihood* based on acquiring the *Dataset in Question.*

As discussed previously, the process of predicting whether or not a strategic competitor could achieve a given *Use of Concern* is a challenging and imprecise process. We developed the methodology outlined above to account for the inherent uncertainty in this process through the consideration of multiple estimates for the *Time until Available* for each *Enabler*. The ultimate quantitative result of this process, the *Likelihood Metric*, is

a point estimate. This single metric mathematically accounts for the uncertainty inherent in the assessment, but presented alone does not convey that uncertainty to a reader. Instead, the single numeric point estimate of *Likelihood* may suggest an unintended level of certainty in the result. Therefore, the quantified *Likelihood Metric* is not intended to be reported as the assessed *Likelihood*. Instead, it is solely to be used to identify the associated *Likelihood Level.* This categorial result is the ultimate product of the *Likelihood* assessment and is incorporated with *Consequence* to determine *Risk*.

D. Assess Consequence

Once the user assesses the *Likelihood* of the strategic competitor achieving the *Use of Concern*, the next methodological step is to assess the *Consequence* of that use. As previously defined, *Consequence* is the impact or resulting harm to the *Operation of Interest* if the strategic competitor achieves the *Use of Concern*. Unlike *Likelihood*, which has a narrow definition, the *Consequence* of a *Use of Concern* can vary depending on how the strategic competitor employs it and the context in which the *Consequence* is viewed.

For example, a *Use of Concern* may have a substantial tactical-level consequence in the context of a specific setting, but in and of itself may fail to result in a meaningful strategic-level impact. Furthermore, a *Use of Concern* may fail to present a substantial *Consequence* to one national security interest (e.g., military operational success) and simultaneously present substantial harm to another interest (e.g., economic competitiveness).

Just as it is impractically time-consuming—if not impossible—for a user to consider all potential *Uses of Concern* for a given *Dataset of Interest*, it is likewise infeasible for a user to consider all potential *Consequences* associated with a given *Use of Concern*. Therefore, the risk assessment methodology focuses on assessing the *Consequence* to a single user-specified *Operation of Interest*. For the purpose of the risk assessment methodology, an *Operation of Interest* is defined as one of the diverse range of activities whose success is of concern to national security. These activities may range from being tied to broad strategic-level goals down to more narrowly focused tactical activities.

Additionally, the *Operation of Interest* is not limited to just the military components of national security. Activities relating to intelligence operations and industrial and economic sectors are also included in this definition. Additional discussion on selecting the *Operation of Interest* is presented in the next section.

It is important to note that within the risk assessment methodology, *Consequence*, unlike *Likelihood*, does not change depending on whether or not the strategic competitor has access to the *Dataset in Question*. This follows from the fact that the *Consequence* is assessed assuming the strategic competitor achieves the *Use of Concern*—regardless of how likely or unlikely that is.

In other words, *Consequence* and *Likelihood* are assessed independently. If a situation arises in which the *Dataset in Question* provides the strategic competitor with some improved capability that they would not otherwise possess, this improved capability should be considered as a different *Use of Concern*, with its own associated risk assessment.

To summarize, execution of the risk assessment methodology will result in two assessed *Likelihood Levels* (with and without access to the *Dataset in Question*), but only one assessed *Consequence*. Future development of the methodology could consider how *Consequence* may change given the strategic competitor's access to the *Dataset in Question*.

The methodology for assessing the *Consequence* to the user-specified *Operation of Interest* is shown in the green boxes across the bottom Figure 1 and consists of:

1. Identifying the *Operation of Interest,*
2. Identifying and evaluating a *Consequence Metric*, and
3. Consulting with stakeholders or subject matter experts to assess the associated *Consequence Level*.

Discussion of each of these steps and definitions of newly introduced terms are contained in the following sections.

1. Identify the Operation of Interest

The first step of assessing *Consequence* is to identify the *Operation of Interest*. The *Operation of Interest* can be selected based on the user's particular interests or mission space. Alternatively, a user can rely on subject matter expertise to select an *Operation of Interest* that may be particularly vulnerable to the *Use of Concern*. The JRAM presents examples of national security activities that can inform a user's selection of an *Operation of Interest*. These include the following strategic-level activities:

- Safeguarding the security of the United States, its population, civil society, Allies, and Partners;
- Safeguarding the security of the U.S. economy and the global economic system;
- Preserving and extending universal values; and
- Advancing and maintaining a U.S.-led international order.[28]

Considering the *Consequence* to such high-level national security activities may seem appealing, as it would present a holistic understanding of national security risk. However, in most cases it is unlikely a user will be able to directly assess the impact of the *Use of*

[28] Department of Defense, *Joint Risk Analysis Methodology*.

Concern on these activities. Therefore, the choice of a more narrowly focused *Operation of Interest* will be required in most cases.

In addition to broad strategic-level national security activities, the JRAM also presents narrowly focused operational activities. Examples from the JRAM include:

- Achieving planned mission objectives;
- Obtaining DOTMLPF-P[29] capabilities relative to threat capabilities;
- Obtaining modernization goals;
- Maintaining critical maintenance; and
- Maintaining readiness.

Users can reference the above examples and the additional activities presented in the JRAM as a starting point for identifying an *Operation of Interest*. It is worth highlighting that as a DOD product, the JRAM is focused on military risk. However, other national security activities, especially those relating to the economy, economic competitiveness, and industry, should also be considered as candidate *Operations of Interest*.

Finally, the case study in Chapter 3 illustrates an example *Operation of Interest* that may be informative to future applications of the methodology. As discussed above, the potential for a user to select an inconsequential *Operation of Interest* is a limitation of the methodology that is ideally ameliorated through subject matter expertise or consideration of multiple operations.

2. Identify and Evaluate a Consequence Metric

The next step is for the user to assess the level of harm to the *Operation of Interest* by identifying and evaluating a *Consequence Metric*. A *Consequence Metric* is a unit of measure, ideally quantitative, for assessing the extent of the impact of the *Use of Concern* on the *Operation of Interest*.

It is difficult to provide general guidance on how users should select a relevant *Consequence Metric* given the diversity of uses of biological data. An ideal metric is quantifiable, able to be accurately evaluated, and directly relevant to the *Operation of Interest*. Quantitative, as compared to qualitative, metrics are preferred as they provide a more objective and unambiguous characterization of the level of harm caused to the *Operation of Interest*. It may not always be possible to identify a relevant quantifiable metric. A relevant qualitative metric is preferable to an irrelevant quantitative metric.

Ideally, the metric can be accurately evaluated; it is not useful without data or methodology for evaluating it. Precision, while desirable, may not be possible. Rough

[29] Doctrine, organization, training, materiel, leadership and education, facilities, and policy.

order-of-magnitude estimates that are grounded in data and derived with sound analytic methods are still preferable over subjective descriptions of *Consequence*. Finally, and most importantly, a *Consequence Metric* should be relevant to the *Operation of Interest.* A metric that can be accurately assessed but is of little relevance provides limited utility in assessing *Consequence*.

The case studies in Chapter 3 and the classified version of the paper provide examples of *Consequence Metrics* that span various national security operations including military, intelligence, and economic. Users should consult these examples to see if any of these metrics are applicable or adaptable to their *Use of Concern* and *Operation of Interest*.

We recognize the challenges in identifying and evaluating a useful *Consequence Metric*, especially for users lacking familiarity with the *Operation of Interest* or related subject matter expertise. To the extent possible, stakeholders and other individuals with subject matter expertise relating to the *Operation of Interest* should be consulted. This is particularly important to ensure the choice of a relevant *Consequence Metric.*

Users may need to explore multiple *Consequence Metrics* for a given *Use of Concern* to identify a satisfactory choice. There is no guarantee that the selected *Consequence Metric* captures the most consequential effect of the *Use of Concern* on the *Operation of Interest*. This is a limitation of the methodology that is ideally ameliorated through stakeholder input.

3. Assess Consequence Level

The final step in assessing *Consequence* is translating the *Consequence Metric* to a *Consequence Level*. For the purposes of the risk assessment methodology, *Consequence Level* is a categorical measure of the level of harm to the *Operation of Interest*. The *Consequence Levels* used in the risk assessment methodology are adapted from those presented in the JRAM and shown in Table 2.[30] As shown in the table, each *Consequence Level* has multiple qualitative descriptors to account for the variety of *Operations of Interest* that a user may wish to assess.

[30] Department of Defense, *Joint Risk Analysis Methodology*.

Table 2. *Consequence Levels*

Consequence Level	Description
Minor	• Confined damage to *Operation of Interest* • *Operation of Interest* objectives achievable (mission success)
Modest	• Considerable damage to *Operation of Interest* • *Operation of Interest* objectives mostly achieved
Major	• Catastrophic damage to *Operation of Interest* • *Operation of Interest* objectives minimally achieved
Extreme	• Existential/permanent damage to *Operation of Interest* • *Operation of Interest* failure

The process of translating the *Consequence Metric* into a *Consequence Level* requires a subjective judgement. The expertise and operational familiarity required to make an informed judgement is likely very different from the scientific and technical expertise required to execute the other components of the methodology (e.g., identifying and assessing *Enablers*). It is likely that the user of the risk assessment methodology will lack the qualifications to make an informed judgement of *Consequence Level*. Therefore, users will require input from stakeholders involved with the *Operation of Interest*. The users of the methodology should present the stakeholders with the evaluated *Consequence Metric* and ask them to assess the associated *Consequence Level*.

Ideally, stakeholders will have already been consulted to ensure the selected *Consequence Metric* is relevant. If not, the user should first confirm the selected *Consequence Metric* is relevant and if not, work with the stakeholder to identify one that is. In the event that the user cannot obtain stakeholder judgement on the *Consequence Level*, *Risk* can still be assessed using the *Consequence Metric* and assessed *Likelihood Level*.[31] This process is detailed in the next section.

The IDA team did not consult operational stakeholders for the development of the case studies. The case studies include our assessment of the *Consequence Metric*, but do not include an expertly informed assessment of the associated *Consequence Level*. However, the case studies do provide a notional *Consequence Level* to illustrate the risk assessment methodology.

E. Assess Risk

The final step in the methodology combines the assessed *Likelihood* and *Consequence* to determine *Risk*. As shown in the orange boxes in Figure 1, the methodology has two

[31] Obtaining stakeholder judgement was outside of the scope of the present analysis; therefore, *Consequence Levels* were not assessed for the case studies.

ultimate outputs: *Risk* and *Risk Level*. If the user cannot obtain stakeholder judgement on the *Consequence Level*, then the output of the methodology is the *Likelihood Level* and associated *Consequence Metric.* Even though these two components cannot be combined to form a single measure of risk, they do provide an informative description of how likely it is that the strategic competitor can achieve the *Use of Concern* (i.e., *Likelihood Level*) and if the use is achieved, what is the resulting harm to the *Operation of Interest* (i.e., *Consequence Level*).

If a *Consequence Level* is assessed, then it can be combined with the *Likelihood Level* to determine an overall *Risk Level.* For the purpose of the methodology, *Risk Level* is a categorical measure of *Risk* posed by the *Use of Concern* to the *Operation of Interest* and is determined by the *Risk Matrix* shown in Figure 6. The *Risk Matrix* is a two-dimensional mapping of each potential combination of *Consequence Level* and *Likelihood Level*.

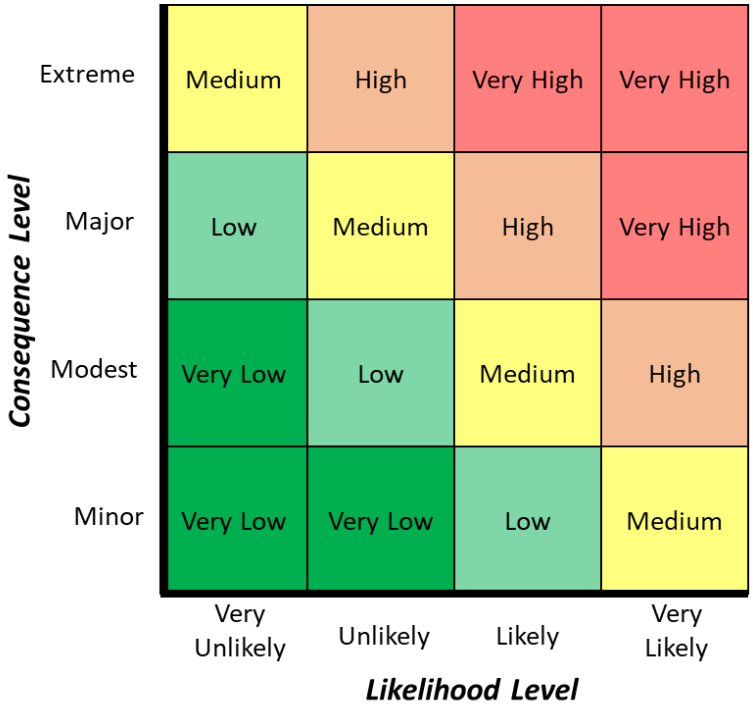


Figure 6. *Risk Matrix*

As described in Section 2.C.3, two *Likelihood Levels* are determined for a given *Use of Concern*: one under the assumption that the strategic competitor does not possess the *Dataset in Question* and again under the assumption that they do. These two *Likelihood Levels* are used to generate two corresponding assessments of *Risk*, whose difference reflect the change in *Risk* associated with the strategic competitor's access to the *Dataset in Question*.

Figure 7 illustrates an example of evaluating the change in *Risk*. In this example, if the strategic competitor is assumed to not possess the *Dataset in Question*, then the *Likelihood Level* is Very Unlikely. However, if the strategic competitor is assumed to have access to the *Dataset in Question*, then the *Likelihood Level* increases to Likely. In this example the *Consequence Level* is Major, so the acquisition of the *Dataset in Question* increases the *Risk Level* from Low to High.

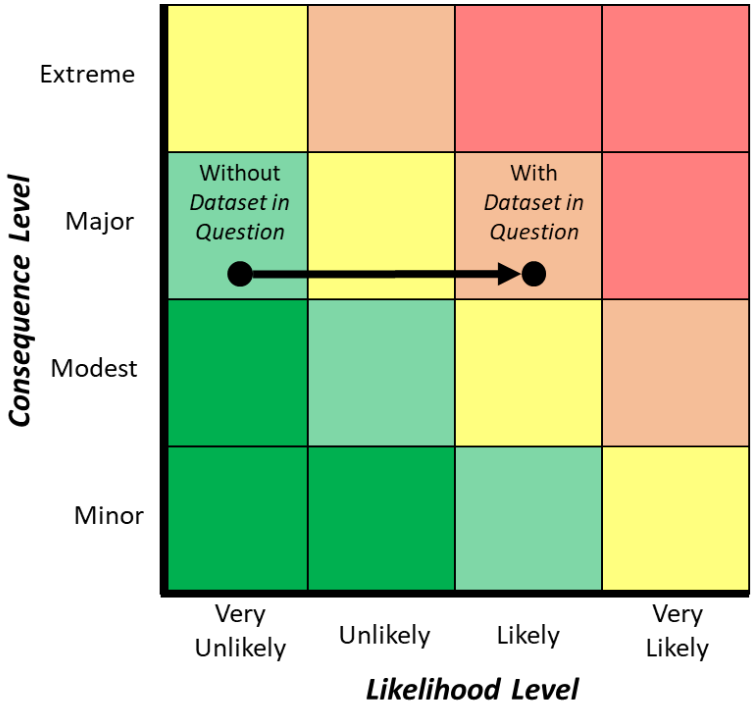


Figure 7. Example Illustrating Change in *Risk* due to the Strategic Competitor's Access to the *Dataset in Question*

As described in Section 2.D, the *Consequence* assessment is predicated on the assumption that the strategic competitor successfully achieves the *Use of Concern*—regardless of how likely it is. Therefore, the presence or absence of the *Dataset in Question* does not change the *Consequence* of the *Use of Concern*. As a result, all changes in *Risk* are based solely on changes in *Likelihood* and are represented as horizontal displacements in the *Risk* matrix.

As discussed in the previous section, the IDA team did not consult with operational stakeholders to determine an expertly informed assessment of *Consequence Level* for the *Consequence Metric* in Chapter 3. However, one objective of the case study was to illustrate the methodology. Therefore, we provide a notional *Consequence Level* that allowed us to assess the associated *Risk Level* and present a *Risk Matrix*. Given the reliance on a notional *Consequence Level*, the *Risk Level* presented in the case study should similarly be viewed as notional. That being said, the *Risk* reported in the case study (i.e.,

the *Likelihood Level* and the *Consequence Metric*) is not notional and reflects the IDA team's assessment.

F. Identify Drivers of Risk

In addition to providing the user with an overall *Risk* and *Risk Level*, the methodology can be used to identify *Drivers of Risk*, which are *Enabler(s)* or other capabilities that the *Likelihood* of the strategic competitor achieving the *Use of Concern* is highly sensitive to. *Drivers of Risk* can provide insight on data, capabilities, and technologies that can strongly impact a strategic competitor's ability to achieve the *Use of Concern.* In some cases, the *Dataset in Question* may also be a *Driver of Risk*. Some *Enablers* identified as *Drivers of Risk*, particularly information or technologies held by U.S. companies or individuals, may provide a user with candidate *Enablers* to consider safeguarding.

As all *Enablers* are, by definition, required to achieve a successful *Use of Concern*, *Drivers of Risk* can be determined by examining which *Enablers* the strategic competitor is least likely to possess within the *Timeframe of Interest*. There can often be more than one *Driver of Risk* for a *Use of Concern*, in which case obtaining a *Driver of Risk* alone may not change the *Likelihood* of successful use. Taken together, the *Drivers of Risk* can be considered the "bottlenecks" for achieving a *Use of Concern*. Examination of *Drivers of Risk* is not a required step of the methodology, but can provide useful insight with regards to potential risk mitigation.

G. Limitations

This final section of the chapter details the limitations of the risk assessment methodology. The primary limitation of the risk assessment method is that usability beyond the IDA research team has not been assessed. The methodology is likely highly dependent upon the user's expertise, access to information, and available time for the analysis. We do not know the minimum amount of time or type and level of expertise required to generate accurate results. Nor do we know the sensitivity of the methodology to the use by someone with constrained time or expertise. A lack of understanding of the methodology's usability presents a challenge to characterizing the implications of many of the method's other limitations, such as its reliance on users to generate *Uses of Concern* and *Operations of Interest* for a given *Dataset in Question*.

The methodology involves characterizing the *Likelihood* for individual *Uses of Concern* for a given *Dataset in Question*. Assessing a single *Use of Concern* may produce an incomplete depiction of the range of possible applications of the data. There is no guarantee that the user-selected *Use of Concern* will generate the greatest or even a representative *Risk*. This may be ameliorated through the consideration of multiple *Uses of Concern*, but users may lack the time to repeatedly apply the methodology. Additionally,

a user's lack of familiarity with the *Dataset in Question* and its potential applications may result in them failing to identify the most meaningful *Uses of Concern*.

Similar to identifying meaningful *Uses of Concern*, users also must select meaningful *Operations of Interest*. Ideally, the vulnerabilities of selected *Operation of Interest* will result in an informative characterization of *Risk*. The greatest *Risk* posed by a strategic competitor's access to some dataset may be associated with *Operations of Interest* that were not considered, or may be associated with the cumulative effect on multiple operations. Again, repeated applications of the methodology to a variety of *Operations of Concern* may ameliorate this limitation given the user has sufficient time and knowledge available.

The expertise required to assess the *Likelihood* of a given *Use of Concern* is likely different from the expertise required to assess the *Consequence* of that use. Accurately assessing *Likelihood* likely requires familiarity with both the scientific phenomenon associated with the biological data and technical details of its application. Accurately assessing *Consequence* likely requires operational-relevant expertise. Ideally, users of the methodology will engage with stakeholders of the *Operation of Interest* throughout the *Consequence* assessment process.

Given the time and resources available, the IDA team did not validate the risk assessment methodology. Likewise, we did not assess reproducibility – that is, if execution of the methodology by two different users generates the same result. If different users arrive at different results, the methodology may require multiple individuals to conduct parallel assessments, resulting in additional cost and time requirements. Follow-on efforts assessing the methodology's sensitivity and reproducibility in the face of diverse users would improve our understanding of the informativeness and resource requirements of the generated risk assessments.

Finally, we did not assess the parsimony of the risk assessment methodology. That is, are all of the methodology's inputs required to generate an informative risk assessment? Or, could a methodology with fewer inputs, or inputs that require less specificity, generate a similarly informative risk assessment? For example, the methodology currently uses three user-specified estimates on each *Enabler's Time until Available* to generate an assessment of that *Enabler's Availability*. An alternative approach could involve users directly assessing each *Enabler's Availability* to be a binary value, as either likely to be available or unavailable within the *Timeframe of Interest*.

If all *Enablers* were likely to be available to the strategic competitor, then it would similarly be likely that the strategic competitor could achieve the *Use of Concern*. IDA did not assess the utility of such a simplification by considering the potential tradeoff between the time required for execution and accuracy or reproducibility.

3. Agricultural Bioprocessing Data Illustrative Case Study

A. Summary

Strategic Competitor Objective

The *Strategic Competitor Objective* is to prevent loss of pork supply in China due to Porcine Reproductive and Respiratory Syndrome (PRRS) with genetically engineered (GE) disease-resistant pigs.

Dataset in Question

The *Dataset in Question* is comprised of a complete description of the tools, methods, and observations of the research published by 2016 Whitworth et al. on the two generations of PRRS virus (PRRSV)-resistant pigs produced by knocking out the CD163 gene. This includes description of the gene target for disease resistance, the design of the CRISPR-Cas9 system (i.e., single guide RNA sequence with transactivating CRISPR RNA duplex), detailed description of the CRISPR-Cas9 delivery method, and full characterization of the observed-off target effects.

Use of Concern

The *Use of Concern* is to develop genetically engineered PRRSV-resistant pigs using CRISPR-Cas9 to offset the number of pigs lost to PRRS in China.

Enablers

Table 3 summarizes our assessment of the *Enablers* associated with the *Use of Concern* under the assumption that the strategic competitor does and does not possess the *Dataset in Question*. Values for the *Time until Available* are presented as *Best Guess* [*Lower Bound*, *Upper Bound*], unless all three estimates are zero, in which case the *Time until Available* is simply presented as 0 years. Blue text denotes *Enablers* that are part of the *Dataset in Question*.

Table 3. Enabler Availability for Agricultural Illustrative Case Study

#	Category	Process Step	Description	Time until Available w/o Dataset in Question (years)	Availability w/o Dataset in Question	Time until Available w/ Dataset in Question (years)	Availability w/ Dataset in Question
1	Data	1, 2	Gene target	0	1	0	1
2	Data	1, 2	Resistance against Chinese strains[A]	3.3 [2.8, 6.8]	0.94	2.8 [2.8, 6.8]	0.98
3	Technological Capabilities	2, 3	CRISPR-Cas9 system + design	0	1	0	1
4	Scientific Knowledge	3	SCNT editing expertise	0	1	0	1
5	Scientific Knowledge	4	Expertise in genomics/genotyping for large scale breeding	0	1	0	1
6	Infrastructure	4, 5	Large testing facility	0	1	0	1
7	Data	5	Demonstration of large-scale success	5 [4, ∞]	0.47	5 [4, ∞]	0.47
8	Infrastructure	5	Commercialization approval	7 [6, >11][B]	0	7 [6, >11] *	0
9	Scientific Knowledge	5	Commercial-scale manufacturing	>20 [10, ∞] *	0	>20 [10, ∞] *	0
	***Likelihood Metric* (minimum)**				0		0

[A] In addition, the *Dataset in Question* includes information on observed off-target effects that, while not an *Enabler* itself because it is not required for achieving the *Use of Concern*, affects the *Time until Available* for other *Enablers*.

[B] Given that the *Lower Bound* of these *Enablers* is greater than five years, their assessed *Availability* will be zero. Therefore, imprecise estimates for the *Upper Bound* and the *Best Guess* as ranges instead of a single value (i.e., >11 years or >20 years) do not impact the assessed *Availability*.

Likelihood

Table 4 shows the assessed *Likelihood Level* for the case study.

Table 4. Likelihood Level within 5 Years for Agricultural Bioprocessing Case Study

Likelihood of Successful Use	
With access to dataset	Very Unlikely
Without access to dataset	Very Unlikely

Operation of Interest

The *Operation of Interest* is the U.S. Department of Agriculture's Foreign Agricultural Service's mission to protect the U.S. pork export market.

Consequence

Assuming that each disease-resistant pig replaces a pig that would have been lost to PRRS and imported from the U.S., the *Consequence* of the *Use of Concern* would result in an estimated $1.3 billion USD loss in U.S. pork exports to China, or 18.3 percent of the value of global U.S. pork exports.

Risk

Both with and without the *Dataset in Question*, it is **Very Unlikely** that the strategic competitor will be capable of reducing 18.3 percent of the value of global U.S. pork exports by producing sufficient GE PRRSV-resistant pigs for domestic consumption in the next five years.

Drivers of Risk

There are two *Enablers* that behave as *Drivers of Risk* for China's ability to achieve the *Use of Concern* in the next five years: approval of GE pigs for commercialization and commercial-scale manufacturing capability for production of PRRSV-resistant pigs. Regarding the former, China's regulatory environment and bureaucratic process to approve the commercialization of genetically modified organisms (GMOs) for food is deleteriously stringent. Unless the regulatory process is changed significantly, it would require two to three years to undergo one round of review for a production safety certificate after a GE product has completed four stages of R&D approval by the Administration Office for Bio-safety of Agricultural GMOs (which itself could take more than five years).

Second, commercial-scale production of PRRSV-resistant pigs has yet to be demonstrated due to the relatively intensive technical skills and labor requirements

involved in successful genetic engineering of large animals. In order to successfully breed the founder GE pigs once they have been produced, significant expertise in genomics and genotyping is required to ensure that the genetic modifications are carried through subsequent herds. In addition, the life-cycle of the pigs requires significant time for scale-up even without the difficulties in genetic engineering.

B. Detailed Analysis

1. Strategic Competitor Objective

China is the country with the highest pork demand globally and is one of the top export markets for U.S. pork.[32] Given the recent efforts by the Chinese Communist Party (CCP) to reduce agricultural import reliance and projected increase in domestic pork consumption,[33] it is sensible that China would be interested in reducing the incidence of porcine disease, particularly one that has been noted to have high prevalence in China over the past 20 years.[34]

To determine the number of GE pigs needed to achieve the use, we estimated the number of pigs lost in China to PRRS. Given the data gap on number of PRRSV infections in China, the incidence rate for the disease (14.04 percent) was applied to the domestic supply, assumed to be equal to Chinese pig production values. For this study, we used a five-year average of Chinese pig production data in number of whole pigs (see Table 5).

---

[32] U.S. Department of Agriculture Foreign Agricultural Service (FAS), *2021 United States Agricultural Export Yearbook*, (Washington, DC: USDA FAS, April 14, 2022), https://www.fas.usda.gov/data/2021-agricultural-export-yearbook.

[33] Lauren Greenwood, *China's Interests in the U.S. Agriculture: Augmenting Food Security through Investment Abroad*, U.S.-China Economic and Security Review Commission Staff Research Report, (Washington, DC: U.S.-China Economic and Security Review Commission, May 26, 2022,): 9, https://www.uscc.gov/research/chinas-interests-us-agriculture-augmenting-food-security-through-investment-abroad.

[34] Kegong Tian et al., "Emergence of Fatal PRRSV Variants: Unparalleled Outbreaks of Atypical PRRS in China and Molecular Dissection of the Unique Hallmark," *PLoS One* 6, no. e526 (June 2007): 1-10, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1885284/; Jie Song, Di Shen, Jie Cui, Baohua Zhao, "Accelerated evolution of PRRSV during recent outbreaks in China," *Virus Genes* 41 (2010): 241-245, https://pubmed.ncbi.nlm.nih.gov/20652733/; Xiaoxiao Zhang and Chunhe Guo, "Recent advances in inhibition of porcine reproductive and respiratory syndrome virus through targeting CD163," *Frontiers in Microbiology* 13 (September 16, 2022), https://doi.org/10.3389/fmicb.2022.1006464.

Table 5. Chinese Domestic Pig Production 2017-2021

Year	Number of Pigs
2017	433,250,000
2018	428,170,000
2019	310,410,000
2020	406,500,000
2021	449,220,000
Five-Year Average	**405,510,000**
14.04% loss to PRRSV	**56,933,604**

Source: "China: Annual pig census," Pig333.com Website, accessed May 17, 2023, https://www.pig333.com/pig-production-data/china_11/.

Notably, there were discrepancies amongst three data sources on Chinese pig production numbers; namely, Pig333, Statista, and Organization for Economic Cooperation and Development Principal Accounting Officer. The source used for this study was selected based on its independent data collection method using industry sources as compared to the others, which relied on Chinese government reported values.

2. Dataset in Question

In a previous IDA study for the OUSD Strategic Intelligence and Analysis Cell, we found that the production of disease-resistant GE animals could yield significant benefits to various sectors (e.g., pharmaceuticals, medicine, agriculture).[35] As part of this effort, we found that in 2019, Genus, a British firm, licensed its PRRSV-resistant GE pigs to Beijing Capital Agribusiness (BCA) Co Ltd.[36] Notably, Genus first developed these GE pigs in partnership with American researchers at University of Missouri and Kansas State

---

[35] Joseph C. Hamill, Ashley Farris, Janet C. Marroquin Pineda, Jay S. Shah, and Katherine M. Sixt, (U) *In-Depth Technology Review for Biotechnology Net Technical Assessment*, P-32974, (Alexandria, VA: Institute for Defense Analyses, 2022): 76. TOP SECRET//HCS-P/SI//FGI DEU//OC-USGOV/NOFORN/FISA//LES. Only unclassified information is referenced in this document.

[36] Reuters Staff, "Genus Shares Surge on Deal to Market Gene-Edited Pigs in China," *Reuters*, May 16, 2019, accessed October 4, 2021, https://www.reuters.com/article/us-china-genus-plc/genus-shares-surge-on-deal-to-market-gene-edited-pigs-in-china-idUKKCN1SM121?edition-redirect=uk.

University (KSU) in 2014, and continues to work with KSU to explore countermeasures against swine influenza and African Swine Fever.[37,38]

Therefore, Chinese acquisition of American research data from University of Missouri and/or KSU could serve as a potential real-world technology transfer case study. BCA's partnership with Genus may provide some information that would be included in the *Dataset in Question*, such as the gene target, but the scope for this project is focused on the acquisition of U.S.-held data.

For the purposes of this case study, Genus PLC and its global porcine genetics business arm, Pig Involvement Company (PIC), own the dataset, but the dataset would likely also be held and managed by the lead research team at KSU and University of Missouri.[39] Therefore, the *Dataset in Question* is a real-world dataset that includes a complete and detailed description of the tools, methods, and observations of the research published by 2016 Whitworth et al. on the two generations of PRRSV-resistant pigs produced by knocking out the CD163 gene.[40]

Specifically, the *Dataset in Question* includes information on the gene target that informs the design of the CRISPR expression system used and includes full characterization of off-target effects that are alluded to in the publication but not described. Data on off-target effects are not essential to achieving the *Use of Concern*, so they are not considered an *Enabler*. However, such data would likely expedite experimentation for other *Enablers,* so they are considered when estimating *Time until Available* with the *Dataset in Question*.

Genus is committed to pioneering the introduction of PRRSV-resistant GE pigs into the global market.[41] Therefore, acquisition of the company itself would likely have technical and infrastructure capabilities to produce GE pigs at industrial-scale for other applications; this would give China advantages beyond those offered by data acquisition

---

37 Kristin M. Whitworth, Kiho Lee, Joshua A. Benne, Benjamin P. Beaton, et al., "Use of the CRISPR/Cas9 System to produce Genetically Engineered Pigs from In Vitro-Derived Oocytes and Embryos," *Biology of Reproduction* 91, no. 3 (September 2014): 1-13, https://pubmed.ncbi.nlm.nih.gov/25100712/.

38 "Genus R&D: Our Strategic Progress," Genus PLC Website, accessed September 26, 2022, https://www.genusplc.com/about-us/our-strategic-progress/genus-rd/.

39 The team being those involved in Whitworth et al. 2014, "Use of the CRISPR/Cas9 system" and Whitworth et al., "Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus," *Nature Biotechnology* 34, no.1 (January 2016): 20-2022, https://www.nature.com/articles/nbt.3434.

40 Whitworth et al., "Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus."

41 A. Mark Cigan and Pieter W. Knap, "Technical considerations towards commercialization of porcine respiratory and reproductive syndrome (PRRS) virus resistant pigs," *CABI Agriculture and Bioscience* 3, no. 34 (2022): 1-20, https://doi.org/10.1186/s43170-022-00107-5.

alone. For example, Genus has initiated regulatory approval in the U.S. and is actively collaborating with Beijing Capital Agribusiness to establish in-country lab and production facilities.[42]

3. Use of Concern

Research from the past decade indicates that PRRSV infection is a prime candidate for intervention through genetic modification.[43] Conversely, PRRS vaccines have been found to have inconsistent and waning efficacy due to the virus's highly mutagenic and recombinant nature.[44] Therefore, there has been rising interest and continuing advancements in the development of PRRSV-resistant GE pigs as a disease prevention strategy.[45]

As of 2019, there was a 14.04 percent prevalence rate of PRRSV infection in randomly tested pigs in China (100 of 712 samples from various regions).[46] Based on a five-year average of pig production in China, more than 56 million disease-resistant GE pigs would need to be produced in order to offset the 14.04 percent prevalence of PRRS.[47]

4. Enablers

a. Selection of Enablers

To identify the data, infrastructure, scientific knowledge, and technological capabilities enabling the *Use of Concern* for the *Dataset in Question* (i.e., *Enablers*), it is necessary to first understand the scientific process underlying such a *Use of Concern*. The

---

42 Ibid.

43 Zhang and Guo, "Recent advances in inhibition of porcine reproductive and respiratory syndrome virus through targeting CD163"; Hongming Yuan et al., "Current Status of Genetically Modified Pigs That Are Resistant to Virus Infection," *Viruses* 14, no. 417 (February 2022), https://doi.org/10.3390/v14020417.

44 I. Ruedas-Torres et al., "The jigsaw of PSSV virulence," *Veterinary Microbiology* 260 (September 2021): 1-9, https://doi.org/10.1016/j.vetmic.2021.109168; Chanhee Chae, "Commercial PRRS Modified-Live Virus Vaccines," *Vaccines* 9, no.185 (February 2021), https://doi.org/10.3390/vaccines9020185.

45 Kristin M. Whitworth et al., "Improvements in pig agriculture through gene editing," *CABI Agriculture and Bioscience* 3, no.41 (June 2022), https://doi.org/10.1186/s43170-022-00111-9.

46 Nahua Chen, Yanzhao Xiao, Mengxue Ye, et al., "High genetic diversity of Chinese porcine reproductive and respiratory syndrome viruses from 2016 to 2019," *Research in Veterinary Science* 131 (August 2020): 38-42, https://pubmed.ncbi.nlm.nih.gov/32289611/.

47 Chinese pig production data from 2017 through 2021 was collected from pig333.com, an independent source of pork industry data. Alternatively, the OECD-FAO publishes agricultural supply and consumption reports annually based on country-provided data. However, discrepancies between open source reporting of African Swine Fever cases in China and 2018 production numbers provided by OECD-FAO calls into question the reliability of OECD-FAO data.

steps comprising this process were identified from relevant scientific experiments (i.e., production of disease-resistant GE pigs) conducted by academia and industry research groups.

Small-scale studies demonstrating feasibility describe the first four steps of the GE process: identify a gene target, design the gene-editing system, insert the gene-editing system into the organism, and reproduce the GE product.[48] Industry-funded projects are fewer in quantity and typically interim results are not published. Instead, most industry-funded projects appear to be focused on scaling up production and bringing the GE animal product to market.[49] In addition, there are articles that discuss the regulatory requirements for the commercial approval of GE agricultural products.[50] Figure 8 summarizes the general process for genetically engineering an agricultural product at industrial scale.

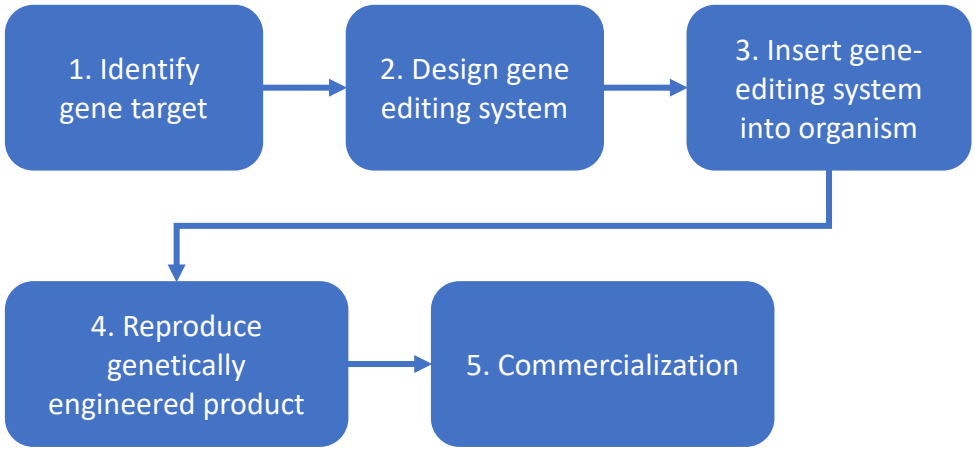


Figure 8. Generalized Process Flow Diagram: Process for genetically engineering an agricultural product at industrial scale

---

48 Examples can be found in the following review articles: Laura Daniela Ratner et al., "Practical Approaches for Knock-Out Gene Editing in Pigs," *Frontiers in Genetics* 11, no. 617850 (March 2021), https://pubmed.ncbi.nlm.nih.gov/33747029/; Zhang and Guo, "Recent advances in inhibition of porcine reproductive and respiratory syndrome virus through targeting CD163."

49 Cigan and Knap, "Technical considerations towards commercialization"; Beate Rielinger et al., "Cas9-expressing chickens and pigs as resources for genome editing in livestock," *PNAS* 118, no. 10 (2021), https://doi.org/10.1073/pnas.2022562118.

50 Examples include: Hamish A. Salvesen et al., "Simulating the Commercial Implementation of Gene-Editing for Influenza A Virus Resistance in Pigs; An Economic and Genetic Analysis," *Genes* 13, no. 1436 (August 2022), https://doi.org/10.3390/genes13081436; Zhihua Xiao and William A. Kerr, "Biotechnology in China- regulation, investment, and delayed commercialization," *GM Crops & Food* 13, no. 1 (May 2022): 86-96, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9090284/.

b. Assessment of Enablers

Identify Gene Target

1) Gene target for disease resistance

A gene target is required to first determine whether disease resistance can be conferred through gene editing and can inform the design of the gene editing system. Ensuring disease resistance may involve more than one gene target, in which case the gene editing system will need to accommodate multigenetic targets. In this case, one gene target has been demonstrated to be sufficient to inhibit PRRSV infection (more below).

Time until Available without Dataset in Question: 0 years

Time until Available with Dataset in Question: 0 years

Several experimental studies, some of which were conducted in China, have demonstrated that targeting CD163 receptor through gene editing is effective in preventing PRRS in pigs.[51] These studies have been published in various academic journals and are publicly available. Thus, acquisition of the *Dataset in Question* does not affect knowledge of the gene target.

2) Disease resistance against most prevalent PRRSV strains in China

Epidemiologic and phylogenetic studies have shown that sub-lineage emergence of new PRRSV variants occurs every 1 to 4 years globally, with primary lineages continuing to diversify and virulent sub-lineages varying their degree of genomic sequence overlap.[52] Given this genetic variability and empirical evidence of strain-dependent efficacy in vaccines,[53] it is necessary to determine whether the current gold standard for targeting PRRSV resistance (i.e., gene CD163) would be effective against the most prevalent viral strains in China.

---

[51] Chinese studies include: Huaqiang Yang et al., "CD163 knockout pigs are fully resistant to highly pathogenic porcine reproductive and respiratory syndrome virus," *Antiviral Research* 151 (2018): 63-70, https://doi.org/10.1016/j.antiviral.2018.01.004; Chunhe Guo et al., "Highly Efficient Generation of Pigs Harboring a Partial Deletion of the CD163 SRCR5 Domain, Which Are Fully Resistant to Porcine Reproductive and Respiratory Syndrome Virus 2 Infection," *Frontiers in Immunology* 10, no. 1846 (August 2019): 1-14, https://doi.org/10.3389/fimmu.2019.01846.

[52] Igor A. D. Paploski et al., "Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States," *Vaccines* 9, no. 608 (June 2021), https://doi.org/10.3390/vaccines9060608.

[53] Chae, "Commercial PRRS Modified-Live Virus Vaccines"; Raymond R.R. Rowland et al., "Effect of the host genotype at a PRRS resistance marker on evolution of the modified-live PRRS vaccine virus in pigs," *Virus Research* 316, no. 198809 (May 2022), https://pubmed.ncbi.nlm.nih.gov/35568091/.

Time until Available without Dataset in Question: 3.3 [2.8, 6.8] years

Time until Available with Dataset in Question: 2.8 [2.8, 6.8] years

PRRSV variants are grouped into two main types: the European (type-1) and North American (type-2).[54] These groupings generally correspond to their geographic prevalence, though East Asia, China specifically, has also seen significant circulation of the virus since 1995.[55] In China, high genetic diversity has been documented in circulating PRRSV since the first major outbreak in 1995, although the most prevalent strains have consistently been type-2.[56] In the past five years, the predominant strains in China have been highly-pathogenic PRRSV (HP-PRRSV, lineage 8) and NADC 30-like (lineage 1) strains.[57] However, GE efforts involving CD163 as a target have only been applied to a select number of PRRSV strains that do not include these sub-lineage variants.

Experiments targeting CD163 have tested the following strains:

- Strains originating in/endemic to North America and the United States: NVSL, 97-7895, VR-2385, MN184[58]
- Strains originating in/endemic to China: JXA1, MY[59]

Therefore, it is reasonable to assume that further research would be required to confirm that the CD163 receptor could still serve as an effective gene target for conferring disease resistance against dominant Chinese PRRSV strains. Based on observed experimental designs, the following formula was used to estimate the time required to complete this research, where time estimates for pregnancy, maturation, and reproductive

---

54 Paploski et al., "Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States."

55 Chen et al., "High genetic diversity of Chinese porcine reproductive and respiratory syndrome viruses from 2016 to 2019."

56 Jian Chen et al., "Genetic Variation of Chinese PRRSV Strains Based on ORF5 Sequence," *Biochemical Genetics* 44, no. 9/10 (October 2006), https://pubmed.ncbi.nlm.nih.gov/17048090/.

57 Kui Fang et al., "Epidemiological and Genetic Characteristics of Porcine Reproductive and Respiratory Syndrome Virus in South China Between 2017 and 2021," *Frontiers in Veterinary Science* 9, no. 853044 (April 2022), https://www.frontiersin.org/articles/10.3389/fvets.2022.853044/full.

58 Calvert et al., "CD163 Expression Confers Susceptibility to PRRSV"; Whitworth et al., "Use of CRISPR/Cas9 System to Produce GE Pigs from In Vitro-Derived Oocytes and Embryos"; Whitworth et al., "Gene-edited pigs are protected from PRRSV"; Burkard et al., "Precision engineering for PRRSV resistance in pigs"; and Prather et al., "Knockout of maternal CD163 protects fetuses from infection with PRRSV."

59 Guo et al., "Highly efficient generation of pigs harboring a partial deletion of the CD163 SRCR5 Domain," and Yang et al., "CD163 knockout pigs are fully resistant to highly pathogenic PRRSV."

test are fixed and discovery of the gene target and *in vitro* experimentation are variable, as indicated by red text:[60]

$$\text{Time to demonstrate resistance} = \text{discovery of gene target} + \textit{in vitro}\ \text{experimentation} + \text{pregnancy} + \text{maturation} + \text{reproductive test}$$

The following times were used:

- Duration of pregnancy: four months
- Maturation time: one year
- Reproductive test (i.e., insemination and pregnancy of GE animal): 0.5 years[61]

Time until Available without Dataset

To calculate the *Lower Bound*, we assume that the CD163 target works and in vitro trials work on the first try. Accordingly, the time required to discover the gene target is 0 years and the time to conduct *in vitro* experimentation is one year. Therefore, the time to demonstrate resistance is:

$$0 + 1 + \frac{4}{12} + 1 + 0.5 = 2.8\ \text{years}$$

To calculate the *Upper Bound*, we assume that the CD163 target does not work so a new gene target is required and *in vitro* trials require some trial-and-error.[62] In this case, the time required to discover the gene target is two years and the time to conduct *in vitro* experimentation is three years. Therefore, the time to demonstrate resistance is:

$$2 + 3 + \frac{4}{12} + 1 + 0.5 = 6.8\ \text{years}$$

To calculate the *Best Guess*, we use timelines of Chinese research articles (e.g., 2019 Guo et al. and 2018 Yang et al.) and assume that the CD163 target works (i.e., has been successful in disease prevention for PRRSV types 1 and 2 and for multiple sub-lineages) and *in vitro* trials require some trial-and-error. In this case, the time required to discover

---

[60] These estimates were derived from several GE pig experiments and published best practices. See Whitworth et al., "Use of the CRISPR-Cas9 System to Produce Genetically Engineered Pigs from In Vitro-Derived Oocytes and Embryos," and Laura Daniela et al., "Practical Approaches for Knock-Out Gene Editing in Pigs."

[61] Because PRRSV infection affects reproduction (e.g., infertility, increased changes of miscarriages), it is vital to evaluate the ability for GE pigs to reproduce viable progeny to verify disease resistance.

[62] If a new gene target is required, it is dubious whether the *Dataset in Question* would still be of interest for this *Use of Concern*.

the gene target is 0 years and the time to conduct *in vitro* experimentation is 1.5 years. Therefore, the time to demonstrate resistance is:

$$0 + 1.5 + \frac{4}{12} + 1 + 0.5 = 3.3 \text{ years}$$

Time until Available with Dataset

We assume that the *Dataset in Question* includes a full characterization of off-target effects, which in turn would inform the development of a more efficient and effective CRISPR-Cas9 system and delivery method as compared to developing the system de novo. Based on this assumption, the *Best Guess* for the *Time until Available* with the *Dataset in Question* is reduced by half a year, the length of time approximated for ex vivo experimentation with a new CRISPR system. All other time estimates remain unaffected by having the *Dataset in Question*.

Design Gene Editing System

3) CRISPR-Cas9 system and design

A gene-editing system is required to modify an organism's genome. The *Use of Concern* for this case study specifies CRISPR-Cas9 as the nuclease system of interest for producing PRRSV-resistant pigs.

Time until Available without Dataset in Question: 0 years

Time until Available with Dataset in Question: 0 years

CRISPR-Cas9 systems are commercially available for purchase and are designed according to user specificity. Based on open source publications Chinese researchers have demonstrated successful design and use of CRISPR-Cas9 to engineer PRRSV-resistant pigs. Therefore, acquisition of the *Dataset in Question* does not change the *Time until Available* for this *Enabler*.

Insert Gene-Editing System into Organism

4) Somatic Cell Nuclear Transferase (SCNT) expertise

The first step to editing the genome of an animal is to introduce the gene-editing system (i.e., CRISPR-Cas9) into the organism. Currently, the most common and accessible (in terms of expertise) method is to generate single-cell edited embryos in vitro through Somatic Cell Nuclear Transfer (SCNT). Alternatively, *in vivo* methods such as parthenogenetic activation or *in vivo* zygote production can be employed, although there is

a lack of empirical data demonstrating success in editing pigs, likely due to high technical expertise and additional laboratory requirements.[63]

Time until Available without Dataset in Question: 0 years

Time until Available with Dataset in Question: 0 years

Scientific publications indicate that Chinese researchers have used SCNT successfully to generate GE pigs, so the *Dataset in Question* would not change the *Time until Available* for this *Enabler*.[64]

Reproduce Genetically Engineered Product

5) Expertise in genomics/genotyping for large-scale breeding

Empirical studies indicate that inheritance of the PRRSV resistance allele (as engineered by knocking out CD163 using CRISPR-Cas9 in this case study) is recessive. Therefore, strategic breeding via genotyping is required to ensure that subsequent generations of pigs carry the genomic modification of interest.[65]

Time until Available without Dataset in Question: 0 years

Time until Available with Dataset in Question: 0 years

Given the close partnership between Genus and Beijing Capital Agribusiness and recent press releases, it appears that China may have in-country expertise or at least access to expertise in genotyping for large-scale breeding of GE pigs.[66] Therefore, the acquisition of the *Dataset in Question* does not affect the *Time until Available* for this *Enabler*.

6) Large testing facility

Achieving the *Use of Concern* within five years would require large-scale production of a nuclear herd that would then be bred to industrial scale.[67] A large testing facility would thus be required to carry out GE procedures in pigs, a relatively large animal.

Time until Available without Dataset in Question: 0 years

Time until Available with Dataset in Question: 0 years

---

63 Ratner et al., "Practical Approaches for Knock-Out Gene Editing in Pigs."

64 Yang et al., "CD163 Knockout pigs are fully resistant to highly pathogenic porcine reproductive and respiratory syndrome virus," and Guo et al., "Highly Efficient Generation of Pigs Harboring a Partial Deletion of the CD163 SRCR5 Domain."

65 Cigan and Knap, "Technical considerations towards commercialization."

66 Genus, "Genus R&D: Our Strategic Progress."

67 Xiao and Kerr, "Biotechnology in China."

Press releases indicate that at least one Chinese company, Beijing Capital Agribusiness, has established in-country infrastructure, including lab facilities to support industrial-scale production of GE pigs.[68] The *Dataset in Question* would not impact this.

Commercialization

7) Demonstration of large-scale success

Scaling up production to meet market demand and to obtain commercialization approval requires a large supply of GE pigs beyond laboratory efforts.[69] The exact number of pigs required to demonstrate large-scale success is unknown. However, we assume, based on industry reporting, that producing a generation of GE pigs with homozygous CD163 alleles is sufficient to engage in the commercialization approval process in China.[70]

Time until Available without Dataset in Question: 5 [4, ∞] years

Time until Available with Dataset in Question: 5 [4,∞] years

To date, there is a dearth of GE experimentation beyond labor- and time-intensive, small-scale studies involving less than 20 embryo transfers each.[71] Therefore, demonstrating large-scale success would require additional efforts, regardless of whether the *Dataset in Question* has been acquired. Thus, the *Time until Available* values are the same with and without the data. The *Lower Bound* is based on Genus estimates for the time required to produce a third generation of GE pigs with homozygous CD163 alleles that can be bred reliably at scale without off-target effects (i.e., a nucleus generation).[72]

---

68 Genus, "Genus R&D: Our Strategic Progress."

69 Xiao and Kerr, "Biotechnology in China."

70 According to Cigan et al., efficient reproduction of PRSV-resistant pigs can be achieved by breeding GE females and males of a first generation with line-identical, wild-type males and females that have undergone genetic testing for "high merit" traits. This mating would produce a second generation of heterozygous alleles that can then be cross-bred to produce a third generation of homozygous CD163 alleles. This strategy bypasses the difficulty of trying to create a founder generation of "pure" GE pigs with full resistance genes, which is complicated by low-frequency clean exon deletion on the CD163 gene, detected and undetected multiple CD163 alleles, off-target effects, and variable litter size. For full discussion of the strategy, see Cigan and Knap, "Technical considerations towards commercialization," 14-16.

71 Whitworth et al., "Use of CRISPR/Cas9 System to Produce GE Pigs,"; Whitworth et al., "Gene-edited pigs are protected from PRRSV"; Burkard et al., "Precision engineering for PRRSV resistance in pigs"; Prather et al., "Knockout of maternal CD163 protects fetuses from infection with PRRSV"; and Yang et al., "CD163 knockout pigs are fully resistant to highly pathogenic PRRSV."

72 Scientific studies indicate that current methods of genetically engineering the CD163 gene for disease resistance frequently results in mosaicism (i.e., progeny having more than two alleles) and that the frequency of alleles of interest is uneven and variable across different tissues. See Cigan and Knap, "Technical considerations towards commercialization," 5-7.

According to the Genus report, the creation of a homozygous CD163 allele generation is sufficient to begin commercial performance equivalency testing and regulatory data submission. The *Upper Bound* reflects the technical uncertainty in being able to produce a large population of GE pigs using current methods given the absence of empirical evidence. The *Best Guess* assumes that it is possible to generate GE pigs at a sufficiently large scale for commercialization but requires some experimental refinements.

8) Approval of gene-edited pigs for commercialization in China

Genetically modified food organisms must undergo a multi-stage approval process for commercialization and registration with The Ministry of Agriculture and National Administration Committee for Biosafety of Agricultural GMOs in order to enter the Chinese domestic market.[73]

Time until Available without Dataset in Question: 7 [6, >11] years

Time until Available with Dataset in Question: 7 [6, >11] years

Press releases indicate that Genus has licensed its pigs from Beijing Capital Agribusiness, which as of 2019 was on track to seek regulatory approval in China. However, the announcement did not indicate that the process had been initiated. Figure 9 shows time estimates for the process of commercialization of agricultural GMOs in China. Notably, this estimate is based on the timelines for the two currently approved GMO products in China, cotton and papaya,[74] which are plant crops. Therefore, it is likely that the approval process for commercialization of GE pigs deviates in the R&D timeline, so adjustments were made to the time estimates for this case study as described below. Because the *Dataset in Question* would not affect the overall commercialization approval process, the *Time until Available* estimates are unaffected.

Moreover, while not a requirement for the approval process itself, public acceptance of GMO products has played a role in relevant political decisions. Historically public skepticism of GMO foods, in part resulting from GMO food controversies in China, has stalled Chinese efforts to approve commercialization of GMO products because of local

---

[73] Xiao and Kerr, "Biotechnology in China," and Zhihua Xiao and William A. Kerr, "The political economy of China's GMO commercialization dilemma," *Food and Energy Security* 11 (July 2022): 1-14, https://doi.org/10.1002/fes3.409.

[74] Joseph Maina, "China pushes ahead with GMO crops to safeguard food security," *Alliance for Science*, January 21, 2022, https://allianceforscience.cornell.edu/blog/2022/01/china-pushes-ahead-with-gmo-crops-to-safeguard-food-security; Alice Yuen-Ting Wong and Albert Wait-Kit Chan, "Genetically modified foods in China and the United States: A primer of regulation and intellectual property protection," *Food Science and Human Wellness* 5, no.3 (September 2016): 124-140, https://www.sciencedirect.com/science/article/pii/S2213453016300076.

government officials' fear of social instability and dissatisfaction.[75] Recent research shows that the majority of online public debate on GMOs in China continues to be negative.[76]

As a result, the CCP is actively trying to shape domestic public opinion on GMOs to enable widespread commercialization of GMO foods and has supported a rise in "citizen science communicators" (i.e., non-scientist citizens actively engaging in science communication) advocating for GMOs since 2018.[77] Should there be widespread public support for GMO foods in China, it is likely that the regulatory process would be overhauled to streamline commercialization for a wide range of products. Figure 9 depicts the commercialization approval process used for cotton and papaya.

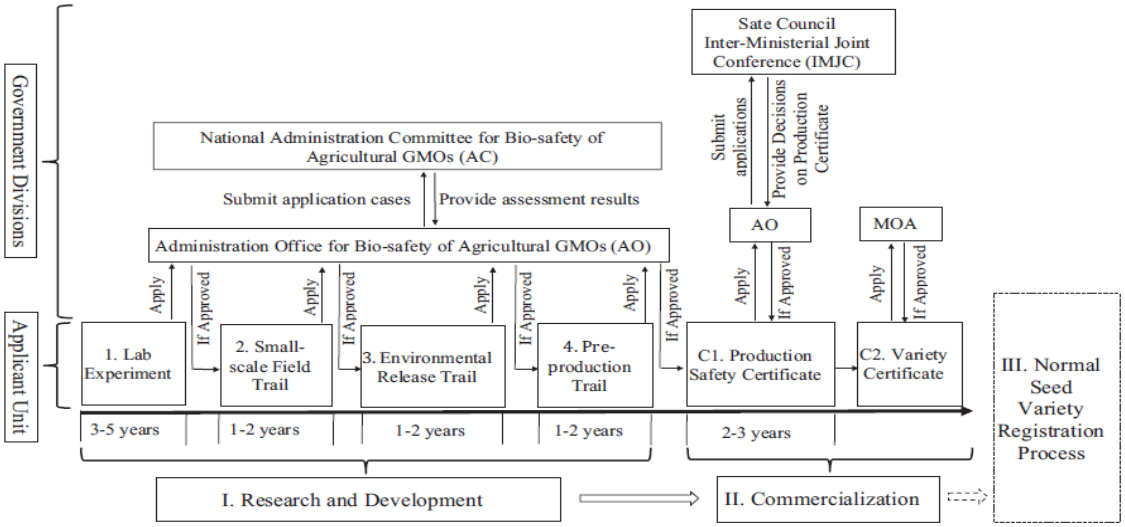


Figure 9. Estimated Timeline for Commercialization Approval of GMOs in China

Source: Xiao and Kerr, "The political economy of China's GMO commercialization dilemma."

Lower Bound: We assume that Chinese approval authorities accept pre-existing experimental results for lab and small-scale trials, the "environmental release trial" is bypassed, and the "preproduction trial" follows the target breeding timeline published by Genus.[78] Increased public support for GMO commercialization (as a result of aggressive

---

75 Xiao and Kerr, "The political economy of China's GMO Commercialization Dilemma," 4-5.

76 Yan Jin, Simon Schaub, Jale Tosun, and Justus Wesseler, "Does China have a public debate on genetically modified organisms? A discourse network analysis of public debate on Weibo," *Public Understanding of Science* (January 2022): 1-13, https://journals.sagepub.com/doi/10.1177/09636625211070150.

77 Maina, "China pushes ahead with GMO crops to safeguard food security"; Zheng Yang, "The new stage of public engagement with science in the digital media environment: citizen science communicators in discussion of GMOs on Zhihu," *New Genetics and Society* (April 2022), https://doi.org/10.1080/14636778.2022.2063826; Wenting Yu et al., "Correcting science misinformation in authoritarian country: An experiment from China," *Telematics and Informatics* 66 (January 2022), https://doi.org/10.1016/j.tele.2021.101749.

78 Cigan and Knap, "Technical considerations towards commercialization," 13.

Chinese propaganda) also leads to a streamlined approval process that enables completion of the review within the *Lower Bound* of the Commercialization step in Figure 9.

- Step I, R&D: Four years to produce three generations of GE pigs for "disease, commercial performance testing, regulatory submissions"
- Step II, Commercialization: Two years for production safety certificate + 0 years for a variety certificate[79]

Therefore, we assume that the *Lower Bound* is 6 years.

Upper Bound: We assume that the approval process requires the company to complete all phases of R&D in-country, breeding timeline for Genus exceeds estimates, and the commercialization approval process remains unchanged from the status quo.

- Step I, R&D: Three years for lab and small-scale trials + >3 years for "preproduction trial"
- Step II, Commercialization: Three years for production safety certificate + >2 years variety certificate[80]

Therefore, we assume that the *Upper Bound* is >11 years.

Best Guess: Based on the 2019 Genus announcement of their intent to initiate the approval process, it can be assumed that licensed pigs can be used to generate a third generation of GE pigs for commercial performance testing and regulatory submissions, thereby following the *Lower Bound* timeline for R&D (i.e., four years). However, given the historical hesitance of local Chinese officials to support GMO food commercialization, it is unlikely that the regulatory process will change in the next couple of years or in time to influence progress within the five-year period of interest. Therefore, we assume that the commercialization step will take at least three years, as shown in Figure 9, leading to a total *Best Guess* of seven years. Likewise, Genus stated in 2019 that it was expected to take "several years" to seek approval for commercial production of their GE pigs.[81]

---

[79] According to Xiao and Kerr, obtaining a production safety certificate is the most onerous step in China's commercialization approval process. Once the production safety certificate is granted, the approval for a variety certificate "does not take long." We assume for the *Lower Bound* that the regulatory approval process is streamlined as a consequence of aggressive government campaigning; therefore, the time to review an application for a variety certificate is negligible. See Xiao and Kerr, "The political economy of China's GMO Commercialization Dilemma," 4.

[80] In the absence of data on the time required to review a variety certificate application, we assume for the *Upper Bound* estimate that the process takes at least two years. This is based on the length of the production safety certificate review process (two to three years).

[81] Reuters Staff, "Genus Shares Surge on Deal to Market Gene-Edited Pigs in China."

9) Commercial-scale manufacturing of GE pigs resistant to PRRSV infection

Time until Available without Dataset in Question: >20 [10, ∞] years

Time until Available with Dataset in Question: >20 [10, ∞] years

Given that engineering the disease resistant CD163 allele of interest is highly labor-intensive and the GE CD163 allele is recessive, Genus has proposed an approach to commercially produce GE pigs by breeding genetically engineered pigs with wild-type line identical mates.[82] However, such a scheme would require a technically skilled team to produce a nucleus generation (as described in the *Time until Available* assessment for *Enabler* 7) and to coordinate with independent producers who would then breed the homozygous GE pigs at a large enough scale for commercial production.

For this scheme to generate the 56 million GE pigs needed for our *Use of Concern* within a five-year period, the nucleus generation would need to be comprised of at least 60,000 GE pigs.[83] As discussed previously, to date, relevant studies have involved less than 20 embryo transfers each and only one research group has published results on a second generation of pigs. That study, the subject of this dataset, disclosed one successful pregnancy by one GE sow yielding eight piglets, only three of which retained the genetic modification.[84] The pregnancy rates of GE pigs is therefore unknown; it is also unknown whether there have been other (unsuccessful) attempts to breed a second generation of GE pigs. Moreover, according to Chinese research publications, only 29.5 to 35.7 percent of GE pigs survive and are healthy into adulthood with current GE methods.[85]

Therefore, we estimated how many disease-resistant pigs could be produced by breeding GE pigs at the current rate of production, starting with 10 GE pigs. For the *Lower Bound* estimate, we assumed that the low survival rate of GE pigs was improved to be comparable to wild-type pigs and that the breeding life of pigs was about five years, with each sow producing an average of 10 piglets per litter (i.e., the average litter size for wild-type pigs). This would result in 21.6 million disease resistant pigs over a 10-year span.[86]

---

[82] For a full description of the breeding scheme, see Cigan and Knap, "Technical considerations toward commercialization."

[83] Cigan and Knap, "Technical considerations toward commercialization."

[84] Kristin M. Whitworth et al., "Gene-edited pigs are protected," 20-22.

[85] Yang et al., "CD163 knockout pigs are fully resistant to highly pathogenic porcine reproductive and respiratory syndrome virus," and Guo et al., "Highly Efficient Generation of Pigs."

[86] While producing 21.6 million disease-resistant pigs is still insufficient to achieving the *Use of Concern*, which calls for 56 million PRRS-resistant pigs, forecasting pig production values beyond 10 years is highly inaccurate based on our assumption that scientific understanding remains unchanged. It is possible, perhaps even likely, that technological advancements improve the production rates of pigs in the future, beyond five years. Therefore, for the purpose of this case study, we used 10 years as the *Lower Bound*.

For the *Best Guess*, we assume that the survivability rate of GE pigs improves but the litter size of disease-resistant pigs remains unchanged (i.e., about three piglets per litter), resulting in 76,000 pigs over a 10-year span and a time estimate of >20 years. Lastly, the *Upper Bound* assumes that commercial scale-up of PRRSV-resistant GE pigs is not technically feasible. Given that the *Dataset in Question* does not offer information for scaling up manufacturing beyond production of lab-scale quantities, the time estimates for this *Enabler* are unaffected by its acquisition.

5. Likelihood

The overall *Likelihood* of China domestically producing GE pigs resistant to PRSV infection at a commercial scale that offsets current loss of pork to the disease in the next five years is **Very Unlikely**. Chinese acquisition of the *Dataset in Question* would not change the *Likelihood* of achieving the *Use of Concern* in this case study. This is due to the two *Enablers* that act as current bottlenecks: approval of GE pigs for commercialization in China and commercial-scale manufacturing of GE pigs. Given that the *Likelihood Metric* of China obtaining these two *Enablers* is assessed to be 0 in the next five years, the overall *Likelihood Level* for having all elements required to achieve the *Use* is **Very Unlikely**.

It is important to note that this case study centers around the use of gene editing as the sole countermeasure to PRRSV infection and assumes that gene editing is conducted *in vitro* using SCNT techniques, as described in Whitworth et al. 2014. This may not be the only approach of interest for China to achieve their objective of combatting PRRSV in domestic pork supply. However, the risk assessment methodology requires an independent risk assessment for each *Use of Concern*, which defines the method of interest.

It is also important to note that the timeline of commercialization approval is based on the current regulatory landscape in China, although the Chinese ministry of agriculture began to revise bureaucratic processes for GMO seeds in late 2021.[87] However, given the uncertainty of the implications for GMO livestock and lack of publicly available details on the revisions, we did not consider the new draft guidelines in this assessment.

6. Consequence

The *Consequence* of the *Use of Concern* was evaluated in the context of impact to the U.S. Department of Agriculture's Foreign Agricultural Service's (USDA FAS) *Operation of Interest* to protect the U.S. pork export market. As a result, the *Consequence Metric* is the quantity of U.S. dollars lost by reduced pork exports to China. We assumed that every

---

[87] Dominique Patton, "China plans overhaul of seed rules to pave way for GMO approvals," *Reuters* (November 14, 2021), https://www.reuters.com/article/china-gmo-regulations/china-plans-overhaul-of-seed-rules-to-pave-way-for-gmo-approvals-idUSKBN2HZ0D3.

GE pig produced commercially in China would replace a pig that would have been otherwise imported from the U.S.

USDA data on the value of pork exports in U.S. dollars from the United States to China over the past five years were averaged to estimate the economic impact of China decoupling from the pork market after achieving their objective of commercially producing GE disease-resistant pigs to replace losses from PRRS.[88] Based on USDA published data,[89] the relative impact of achieving the *Use of Concern* on the value of the overall U.S. pork export market is 18.3 percent loss in U.S. profits according to the following calculation:

$$\$1.302 \text{ billion (U.S. pork exports to China)} / \$7.131 \text{ billion (U.S. total pork exports)} \ x\ 100 = 18.3\%$$

Given China's large domestic market for pork and projected increase in domestic consumption, the USDA FAS would likely consider strategies to mitigate the economic impact of losing its Chinese market. In order to convert the *Consequence Metric* (i.e., 18.3 percent loss in U.S. pork exports) into a *Consequence Level* (i.e., Extreme, Major, Modest, and Minor), a USDA FAS officer should be consulted.

Once having a final determination of *Consequence*, the *Risk* caused by the strategic competitor successfully achieving the *Use* can be calculated using the *Risk Matrix* and *Likelihood Level* with the *Dataset in Question*. A second measure of *Risk* can then be calculated for the *Likelihood Level* without the *Dataset in Question* to assess the change in *Risk*. Because the *Likelihood* was found to be the same with and without the *Dataset in Question*, there is no change in *Risk* associated with its acquisition by the strategic competitor.

7. Notional Risk Level

It is **Very Unlikely** that China can reduce 18.3 percent of the value of global U.S. pork exports by producing sufficient GE PRRSV-resistant pigs for domestic consumption in the next five years using the *Dataset in Question*. Without the *Dataset in Question*, the *Likelihood Level* is still **Very Unlikely**. U.S. agricultural export and pork production experts should be consulted to provide an informed judgement on the appropriate *Consequence Level* associated with this reduction in pork exports. However, for the purpose of illustrating the risk assessment methodology, the IDA team defined the

---

[88] Note, the estimated economic impact does not account for fluctuations in the global pork market value.

[89] "Pork and Pork Products 2021 Export Highlights," U.S. Department of Agriculture Foreign Agricultural Service Website, accessed September 23, 2022, https://www.fas.usda.gov/pork-2021-export-highlights.

Consequence Levels for the impact of reduced pork exports on the USDA's Foreign Agricultural Service's mission to protect the value of U.S. pork exports as follows:

- Minor: <20 percent profit loss; confined damage to the *Operation of Interest*
- Modest: 20-50 percent profit loss; considerable damage to the *Operation of Interest*
- Major: 50-80 percent profit loss; catastrophic damage to the *Operation of Interest*
- Extreme: >80 percent profit loss; existential damage to the *Operation of Interest*

Therefore, we consider the *Consequence Level* for this case study to be **Minor** given the estimated loss of <20 percent (at 18.3 percent). As shown in Figure 10, a Minor *Consequence Level* would result in a notional *Risk Level* of **Very Low** regardless of if the strategic competitor possesses the *Dataset in Question*.

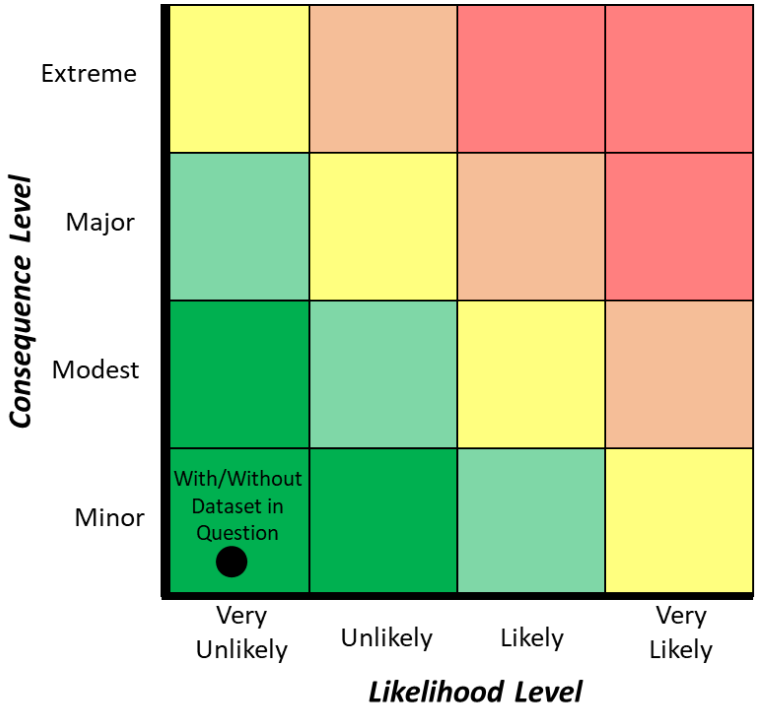


Figure 10. Notional Risk Matrix for Agricultural Case Study

8. Drivers of Risk

This case study highlights two significant *Drivers of Risk* to China's ability to commercially produce GE pigs that are resistant to PRRSV infection at a scale that offsets current losses to the disease in the next five years. The first is approval of gene-edited pigs for commercialization in China. The second is the ability to successfully produce GE pigs

at the commercial scale, due both to lacking technical capability for commercial-scale reproduction of GE mammals and to the natural life-cycle of pigs and their maturation timeline that cannot be expedited (at this time).

China's lengthy and cumbersome regulatory approval processes for GMO food commercialization is a significant hindrance to bringing PRRSV-resistant GE pigs to its domestic market. The bureaucratic process alone could exceed the five-year *Timeframe of Interest*. As mentioned previously, it would take two to three years to undergo one round of review for a production safety certificate after a GE food product has gone through four stages of R&D approval by the Administration Office for Bio-Safety of Agricultural GMOs (which itself could take more than five years).

A significant factor for the CCP's historical reluctance in streamlining GMO approval has been local government hesitancy in response to negative public opinion of GMOs.[90] Therefore, this *Driver of Risk* may not be easily influenced by U.S. activities.

Furthermore, commercial-scale production of PRRSV-resistant pigs has yet to be demonstrated. Genus's efforts to create a comprehensive capability that transitions lab and small-scale successes to commercial-scale production offers a promising path forward to commercializing PRRSV-resistant pigs. In light of Genus's existing partnership with Chinese companies and investment in Chinese infrastructure and technical capabilities, there may not be opportunities to limit Chinese progress towards achieving this *Enabler*.

However, the timeline for commercial-scale production of GE pigs is further constrained by the one-year maturation period and pregnancy duration of pigs, thereby placing the burden of scale-up on the number of pigs that can be genetically engineered at once for the founder herd. Significant automation and technology advancements would need to be achieved in GE methods to produce significant founder pigs for this *Use of Concern*.

9. Conclusion

The methodology worked relatively well for this case study, presumably because the *Dataset in Question* and *Use of Concern* were clearly defined and narrowly scoped. There was a quantitative objective that defined the number of disease-resistant GE pigs to be produced within five years. The *Dataset in Question* was also based on a real-world example of research conducted by U.S. scientists, so we were bettered positioned to assume what data were being held privately, given the research results published in the scientific literature.

---

[90] Xiao and Kerr, "The political economy of China's GMO commercialization dilemma."

However, estimating the time required for China to obtain certain *Enablers* was difficult, in particular for those depending upon scientific experimentation, because it necessitated knowledge of the processes involved in generating those *Enablers*. For example, to estimate the time required to identify a gene target for disease resistance, we needed to understand the scientific process involved and be familiar with how much time is required for lab experimentation using the gene-editing techniques of interest and the logistic considerations for designing and ordering a CRISPR system. In addition, we determined that China already had the *Dataset in Question* or a similar dataset, which decreased the utility of this case study.

Conducting the research and analysis for this case study took approximately 140 hours. Ideally, the analyst would have subject matter expertise in gene-editing techniques for large animals, or at a minimum, a strong understanding of current gene-editing techniques and biological research processes. At the onset of this study, the analyst had expertise in cellular biology but limited knowledge of and no experience with large animal gene-editing techniques and breeding practices for commercialization. In addition, rigorous assessment of the *Consequence Level* would also require understanding of the metrics for mission success at the USDA FAS.

This page is intentionally blank.

Appendix A. Definitions

- ***Biological Data***: information, including associated descriptors, derived from the structure, function, or process of a biological system(s) that is measured, collected, or aggregated for analysis.
- ***Consequence***: the impact or resulting harm to the *Operation of Interest* if the strategic competitor successfully achieves the *Use of Concern.*
    - ***Operation of Interest***: a broad term to describe one of the diverse range of activities whose success is of concern to national security.
    - ***Consequence Metric***: a unit of measure, ideally quantitative, for assessing the extent of the impact of the *Use of Concern* on the *Operation of Interest*
    - ***Consequence Level*:** a categorical measure of the level of harm to the *Operation of Interest*—categories are: Extreme, Major, Modest, and Minor.
- ***Dataset in Question***: the collection of U.S. biological data whose national security risk associated with strategic competitor acquisition is to be assessed.
- ***Enabler*:** capabilities and information that are required to successfully achieving the *Use of Concern*
    - ***Availability***: likelihood that the strategic competitor will possess the *Enabler* within the *Timeframe of Interest*.
    - ***Time until Available***: how long it will take the strategic competitor to possess a given *Enabler*.
        - ***Lower Bound***: a reasonable estimate for the soonest time at which the strategic competitor would possess the *Enabler*
        - ***Upper Bound***: a reasonable estimate for the latest time at which the strategic competitor would possess the *Enabler*—if there is a chance the strategic competitor may never possess the *Enabler* (e.g., requires a yet-to-be discovered scientific phenomenon that may not exist), the *Upper Bound* is ∞

        - ***Best Guess***: a reasonable best estimate for when the strategic competitor would possess the *Enabler*—must lie within the inclusive range of the *Lower* and *Upper Bounds.*
    - ***Drivers of Risk***: *Enabler(s)* or other capabilities that the *Likelihood* of the strategic competitor successfully achieving the *Use of Concern* is highly sensitive to.
- ***Generalized Process Flow Diagram***: a schematic depiction of the major process steps associated with successfully achieving the *Use of Concern*.
- ***Likelihood***: the chance that the strategic competitor successfully achieves the *Use of Concern* within the *Timeframe of Interest*
    - ***Likelihood Metric*:** numeric measure of *Likelihood* that ranges from 0 to 1.
    - ***Likelihood Level***: a categorical measure of *Likelihood*—categories are: Very Likely, Likely, Unlikely, and Very Unlikely.
- ***Risk***: the *Likelihood* and *Consequence* of the strategic competitor successfully achieving the *Use of Concern*.
    - ***Risk Level***: a categorical measure of the level of *Risk* posed by the *Use of Concern* to the *Operation of Interest*—categories are: Very High, High, Medium, Low, and Very Low.
    - ***Risk Matrix***: a two-dimensional mapping of each combination of *Consequence Level* and *Likelihood Level* to an associated *Risk Level*
- ***Strategic Competitor Objective***: a goal the strategic competitor wishes to achieve.
- ***Timeframe of Interest****:* the number of years from the present the user is assessing the *Likelihood* of a strategic competitor achieving a *Use of Concern*. For the presented case study, the *Timeframe of Interest* was set at 5 years.
- ***Use of Concern***: the specific application of the *Dataset in Question* whose risk is to be assessed.

Appendix B. Program Evaluation and Review Technique (PERT) Distribution Calculations

The following details how to implement the calculation of an *Enabler's Availability* from its assessed *Time until Available* using Microsoft Excel. The calculation is demonstrated for two *Enablers*, but the process can be repeated for any number by applying the equations to additional rows. Users enter the name of the *Enabler* and their assessed values for the *Best Guess*, *Lower Bound*, and *Upper Bound* in columns A-D. If an *Enabler* does not have a finite *Upper Bound*, then "infinity" (without the quotation marks) should be used in column D. The *Timeframe of Interest* is specified in cell H1. The calculated *Availability* for each *Enabler* is displayed in column E. Columns F and I are intentionally empty and used to delimitate inputs from calculated values. Columns J-U are for intermediate calculations.

Table B-1. PERT Distribution Calculations

	A	B	C	D	E	F	G	H	I
1	**Enabler**	**Best Guess**	**Lower Bound**	**Upper Bound**	**Probability of Availability in Timeframe**		**Timeframe of interest (yrs)**	5	
2					=IF(D2="infinity",U2,O2)		**Probability of achieving use in timeframe**	=MIN(E:E)	
3					=IF(D3="infinity",U3,O3)		**Likelihood of achieving use in timeframe**	=IF(H2<0.2,"Very Unlikely",IF(H2<0.5,"Unlikely",IF(H2<0.8," Likely","Very Likely")))	

	J	K	L	M	N	O
1	Adj. BG	Adj. LB	Adj. UB	alpha	beta	Probability
2	=IF(B2=C2,B2+0.01,B2)	=C2	=IF(B2=D2,D2+0.01,D2)	=1+(4*(J2-K2)/(L2-K2))	=1+(4*(L2-J2)/(L2-K2))	=IF(C2>$H$1,0,IF(D2<=$H$1,1,BETA.DIST($H$1,M2,N2,TRUE,C2,D2)))
3	=IF(B3=C3,B3+0.01,B3)	=C3	=IF(B3=D3,D3+0.01,D3)	=1+(4*(J3-K3)/(L3-K3))	=1+(4*(L3-J3)/(L3-K3))	=IF(C3>$H$1,0,IF(D3<=$H$1,1,BETA.DIST($H$1,M3,N3,TRUE,C3,D3)))

	P	Q	R	S	T	U
1	Adj. BG	Adj. LB	z	sigma*	mu*	Probability
2	=IF(B2=0,0.02,IF(B2=C2,B2+0.01,B2))	=IF(C2=0,C2+0.01,C2)	=IF(3^2/4+LN(Q2/P2)<=0,SQRT(-4*LN(Q2/P2))+10^-3,3)	=R2/2-SQRT(R2^2/4+LN(Q2/P2))	=LN(Q2)+R2*S2	=LOGNORM.DIST($H$1,T2,S2,TRUE)
3	=IF(B3=0,0.02,IF(B3=C3,B3+0.01,B3))	=IF(C3=0,C3+0.01,C3)	=IF(3^2/4+LN(Q3/P3)<=0,SQRT(-4*LN(Q3/P3))+10^-3,3)	=R3/2-SQRT(R3^2/4+LN(Q3/P3))	=LN(Q3)+R3*S3	=LOGNORM.DIST($H$1,T3E)

Appendix C. Illustrations

Tables

Figures

This page is intentionally blank.

Appendix D. Abbreviations

BCA	Beijing Capital Agribusiness
CCP	Chinese Communist Party
CDF	Cumulative Density Function
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
EO	Executive Order
FAS	Foreign Agricultural Service
GE	Genetically Engineered
GMO	Genetically Modified Organisms
HP-PRRSV	Highly Pathogenic Porcine Reproductive and Respiratory Syndrome Virus
IDA	Institute for Defense Analyses
JRAM	Joint Risk Assessment Methodology
KSU	Kansas State University
MTA	Maintaining Technology Advantage
OUSD	Office of the Under Secretary of Defense
OUSD(R&E))	Office of the Under Secretary of Defense for Research and Engineering
PDF	Probability Density Function
PERT	Program Evaluation and Review Technique
PRRS	Porcine Reproductive and Respiratory Syndrome
PRRSV	Porcine Reproductive and Respiratory Syndrome Virus
R&D	Research and Development
RNA	Ribonucleic Acid
SCNT	Somatic Cell Nuclear Transfer
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
US	United States
USD	United States Dollars
USDA	United States Department of Agriculture

This page is intentionally blank.

Appendix E. References

Classified References:

Joseph C. Hamill, Ashley Farris, Janet C. Marroquin Pineda, Jay S. Shah, and Katherine M. Sixt. (U) *In-Depth Technology Review for Biotechnology Net Technical Assessment*, P-32974. Alexandria, VA: Institute for Defense Analyses, 2022. 76. TOP SECRET//HCS-P/SI//FGI DEU//OC-USGOV/NOFORN/FISA//LES. Only unclassified information is referenced in this document.

Unclassified References:

Chae, Chanhee. "Commercial PRRS Modified-Live Virus Vaccines." *Vaccines* 9, no.185 (February 2021). https://doi.org/10.3390/vaccines9020185.

Chen, Jian, Tao Liu, Cheng-Gang Zhu, Yong-Feng Jin, and Yao-Zhou Zhang. "Genetic Variation of Chinese PRRSV Strains Based on ORF5 Sequence." *Biochemical Genetics* 44, no. 9/10 (October 2006). https://pubmed.ncbi.nlm.nih.gov/17048090/.

Chen, Nahua, Yanzhao Xiao, and Mengxue Ye, et al. "High genetic diversity of Chinese porcine reproductive and respiratory syndrome viruses from 2016 to 2019." *Research in Veterinary Science* 131 (August 2020): 38-42. https://pubmed.ncbi.nlm.nih.gov/32289611/.

Cigan, A Mark, and Pieter W. Knap. "Technical considerations towards commercialization of porcine respiratory and reproductive syndrome (PRRS) virus resistant pigs." *CABI Agriculture and Bioscience* 3, no. 34 (2022): 1-20. https://doi.org/10.1186/s43170-022-00107-5.

Department of Defense. *Joint Risk Analysis Methodology*. Chairman of the Joint Chiefs of Staff. CJSM 3105.01A. Washington, DC: CJCS, October 12, 2021.

Fang, Kui, Shudan Liu, Xiangmin Li, Huanchun Chen, and Ping Qian. "Epidemiological and Genetic Characteristics of Porcine Reproductive and Respiratory Syndrome Virus in South China Between 2017 and 2021." *Frontiers in Veterinary Science* 9, no. 853044 (April 2022). https://www.frontiersin.org/articles/10.3389/fvets.2022.853044/full.

Genus PLC Website. "Genus R&D: Our Strategic Progress." Accessed September 26, 2022. https://www.genusplc.com/about-us/our-strategic-progress/genus-rd/.

Greenwood, Lauren. *China's Interests in the U.S. Agriculture: Augmenting Food Security through Investment Abroad.* U.S.-China Economic and Security Review Commission Staff Research Report. Washington, DC: U.S.-China Economic and

Security Review Commission, May 26, 2022. https://www.uscc.gov/research/chinas-interests-us-agriculture-augmenting-food-security-through-investment-abroad.

Guo, Chunhe, Min Wang, Zhenbang Zhu, Sheng He, Hongbo Lu, Xiaofeng Liu, Xuan Shi, Tao Tang, and Piao Yu, et al. "Highly Efficient Generation of Pigs Harboring a Partial Deletion of the CD163 SRCR5 Domain, Which Are Fully Resistant to Porcine Reproductive and Respiratory Syndrome Virus 2 Infection." *Frontiers in Immunology* 10, no. 1846 (August 2019): 1-14. https://doi.org/10.3389/fimmu.2019.01846.

Ji, Yan, Simon Schaub, Jale Tosun, and Justus Wesseler. "Does China have a public debate on genetically modified organisms? A discourse network analysis of public debate on Weibo." *Public Understanding of Science* (2022): 1-13. https://journals.sagepub.com/doi/10.1177/09636625211070150.

Maina, Joseph. "China pushes ahead with GMO crops to safeguard food security." *Alliance for Science*, January 21, 2022. https://allianceforscience.cornell.edu/blog/2022/01/china-pushes-ahead-with-gmo-crops-to-safeguard-food-security.

Mohan S., M. Gopalakrishnan, H. Balasubramanian, and A. Chandrashekar. "A Lognormal Approximation of Activity Duration in PERT Using Two Time Estimates." *Journal of Operational Research Society* 58, no. 6 (December 21, 2017): 827-831. https://www.tandfonline.com/doi/full/10.1057/palgrave.jors.2602204.

Paploski, Igor A. D., Nakarin Pamornchainavakul, Dennis N Makau, Albert Rovira, Cesar A Corzo, Declan C Schroeder, Maxim C-J Cheeran, Andrea Doeschl-Wilson, Rowland Kao, Samantha Lycett, and Kimberly VanderWaal. "Phylogenetic Structure and Sequential Dominance of Sub-Lineages of PRRSV Type-2 Lineage 1 in the United States." *Vaccines* 9, no. 608 (June 2021). https://doi.org/10.3390/vaccines9060608.

Patton, Dominique. "China plans overhaul of seed rules to pave way for GMO approvals." *Reuters*, November 14, 2021. https://www.reuters.com/article/china-gmo-regulations/china-plans-overhaul-of-seed-rules-to-pave-way-for-gmo-approvals-idUSKBN2HZ0D3.

Pig333.com Website. "China: Annual pig census." Accessed May 17, 2023. https://www.pig333.com/pig-production-data/china_11/.

Ratner, Laura Daniela, Gaston Emilio La Motta, Olinda Briski, Daniel Felipe Salamone, and Rafael Fernandez-Martin. "Practical Approaches for Knock-Out Gene Editing in Pigs." *Frontiers in Genetics* 11, no. 617850 (March 2021). https://pubmed.ncbi.nlm.nih.gov/33747029/.

Reuters Staff. "Genus Shares Surge on Deal to Market Gene-Edited Pigs in China." *Reuters*, May 16, 2019. Accessed October 4, 2021. https://www.reuters.com/article/us-china-genus-plc/genus-shares-surge-on-deal-to-market-gene-edited-pigs-in-china-idUKKCN1SM121?edition-redirect=uk.

Rielinger, Beate, Hicham Sid, Denise Duda, and Benjamin Schusser. “Cas9-expressing chickens and pigs as resources for genome editing in livestock.” *PNAS* 118, no. 10 (2021). https://doi.org/10.1073/pnas.2022562118.

Rowland, Raymond R.R., Tyler Doerksen, Andrea Lu, Maureen Sheahan, Joan Lunney, Jack Dekkers, and Rachel M. Palinski. “Effect of the host genotype at a PRRS resistance marker on evolution of the modified-live PRRS vaccine virus in pigs.” *Virus Research* 316, no. 198809 (May 2022). https://pubmed.ncbi.nlm.nih.gov/35568091/.

Ruedas-Torres, I, I.M. Rodriguez-Gomez, J.M. Sanchez Carvajal, F. Larenas-Munoz, F.J. Pallares, L. Carrasco, and J. Gomez-Laguna. “The jigsaw of PSSV virulence.” *Veterinary Microbiology* 260 (September 2021): 1-9. https://doi.org/10.1016/j.vetmic.2021.109168.

Salvesen, Hamish A., Timothy J. Byrne, Bruce A. Whitelaw, and Fiona S. Hely. “Simulating the Commercial Implementation of Gene-Editing for Influenza A Virus Resistance in Pigs; An Economic and Genetic Analysis.” *Genes* 13, no. 1436 (August 2022). https://doi.org/10.3390/genes13081436.

Song, Jie, Di Shen, Jie Cui, and Baohua Zhao. “Accelerated evolution of PRRSV during recent outbreaks in China.” *Virus Genes* 41 (2010): 241-245. https://pubmed.ncbi.nlm.nih.gov/20652733/.

The White House. “Executive Order on Advancing Biotechnology and Biomanufacturing Innovation for Sustainable, Safe, and Secure American Bioeconomy.” (Presidential Action, Washington, DC: The White House, September 12, 2022). https://www.whitehouse.gov/briefing-room/presidential-actions/2022/09/12/executive-order-on-advancing-biotechnology-and-biomanufacturing-innovation-for-a-sustainable-safe-and-secure-american-bioeconomy/.

Tian, Kegong, Xiuling Yu, Tiezhu Zhao, Youjun Feng, Zhen Cao, Chuanbin Wang, Yan Hu, Xizhao Chen, Dongmei Hu, and Xinsheng Tian, et al. “Emergence of Fatal PRRSV Variants: Unparalleled Outbreaks of Atypical PRRS in China and Molecular Dissection of the Unique Hallmark.” *PLoS One* 6, no. e526 (June 2007): 1-10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1885284/.

U.S. Department of Agriculture Foreign Agricultural Service Website. “Pork and Pork Products 2021 Export Highlight.” Accessed September 23, 2022. https://www.fas.usda.gov/pork-2021-export-highlights.

U.S. Department of Agriculture Foreign Agricultural Service. *2021 United States Agricultural Export Yearbook.* Washington, DC: USDA FAS, April 14, 2022. https://www.fas.usda.gov/data/2021-agricultural-export-yearbook.

Wenting, Yu, Fei Shen, and Chen Min. “Correcting science misinformation in authoritarian country: An experiment from China.” *Telematics and Informatic*s 66, (January 2022). https://doi.org/10.1016/j.tele.2021.101749.

Whitworth, Kristin M., Kiho Lee , Joshua A. Benne, Benjamin P. Beaton, Lee D. Spate, Stephanie L. Murphy, Melissa S. Samuel, Jiude Mao, Chad O'Gorman, Eric M.

Walters, and Clifton N. Murphy, et al. “Use of the CRISPR/Cas9 System to produce Genetically Engineered Pigs from In Vitro-Derived Oocytes and Embryos.” *Biology of Reproduction* 91, no. 3 (September 2014): 1-13. https://pubmed.ncbi.nlm.nih.gov/25100712/.

Whitworth, Kristin, et al. “Gene-edited pigs are protected from porcine reproductive and respiratory syndrome virus." *Nature Biotechnology* 34, no.1 (January 2016): 20-2022. https://www.nature.com/articles/nbt.3434.

Whitworth, Kristin M., Jonathan A. Green, Bethany K. Redel, Jonathan A. Green, Bethany K. Redel, Rodney D. Geisert, Kiho Lee, Bhanu P. Telugu, Kevin D. Wells and Randall S. Prather. “Improvements in pig agriculture through gene editing.” *CABI Agriculture and Bioscience* 3, no. 41 (June 2022). https://doi.org/10.1186/s43170-022-00111-9.

Wong, Alice Yuen-Ting, and Albert Wait-Kit Chan. “Genetically modified foods in China and the United States: A primer of regulation and intellectual property protection.” *Food Science and Human Wellness* 5, no. 3 (September 2016): 124-140. https://www.sciencedirect.com/science/article/pii/S2213453016300076.

Xiao, Zhihua, and William A. Kerr. “Biotechnology in China - regulation, investment, and delayed commercialization.” *GM Crops & Food* 13, no. 1 (May 2022): 86-96. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9090284/.

Xiao, Zhihua, and William A. Kerr. “The political economy of China’s GMO commercialization dilemma.” *Food and Energy Security* 11 (July 2022): 1-14. https://doi.org/10.1002/fes3.409.

Yang, Huaqiang, Jian Zhang, Xianwei Zhang, Junsong Shi, Yongfei Pan, Rong Zhou, Guoling Li, Zicong Li, Gengyuan Cai, and Zhenfang Wu. “CD163 knockout pigs are fully resistant to highly pathogenic porcine reproductive and respiratory syndrome virus.” *Antiviral Research* 151 (2018): 63-70. https://doi.org/10.1016/j.antiviral.2018.01.004.

Yang, Zheng. “The new stage of public engagement with science in the digital media environment: citizen science communicators in discussion of GMOs on Zhihu.” *New Genetics and Society* 41, no. 2 (April 2022): 116-135. https://doi.org/10.1080/14636778.2022.2063826.

You, Edward H. “Safeguarding the Bioeconomy: U.S. Opportunities and Challenges.” (Testimony for the U.S.-China Economic and Security Review Commission: March 16, 2017). https://www.ehidc.org/sites/default/files/resources/files/Ed_You_Testimony_USCC.pdf.

Yuan, Hongming, et al. “Current Status of Genetically Modified Pigs That Are Resistant to Virus Infection.” *Viruses* 14, no. 417 (February 2022). https://doi.org/10.3390/v14020417.

Zhang, Xiaoxiao, and Chunhe Guo. “Recent advances in inhibition of porcine reproductive and respiratory syndrome virus through targeting CD163.” *Frontiers in*

Microbiology 13 (September 16, 2022).
https://doi.org/10.3389/fmicb.2022.1006464.

This page is intentionally blank.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE (DD-MM-YY)	2. REPORT TYPE	3. DATES COVERED (From – To)
XX-08-2023	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NO.
Methodology to Assess Risk from Strategic Competitor Acquisition of U.S. Biological Data and Application to an Agricultural Bioprocessing Case Study	HQ0034-19-D-0001
	5b. GRANT NO.
	5c. PROGRAM ELEMENT NO(S).

6. AUTHOR(S)	5d. PROJECT NO.
Robert Cubeta Kristen Bishop Janet Marroquin Pineda Ashley Farris Clay Hamill Jay Shah	5e. TASK NO. AI-6-5283
	5f. WORK UNIT NO.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NO.
Institute for Defense Analyses 730 E.Glebe Rd Alexandria, VA 22305	IDA Paper P-33619 Log: H 23-000308

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR'S / MONITOR'S ACRONYM(S)
OUSD(R&E) Pentagon, Arlington, VA	OUSD(R&E)
	11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Strategic competitors have used a variety of licit and illicit methods to acquire U.S. biological data. The Director, Science & Technology Exploitation and Analytics, Maintaining Technology Advantage of the Office of the Undersecretary of Defense for Research and Engineering (OUSD(R&E)) is interested in assessing the potential risk resulting from strategic competitor acquisition of biological datasets and communicating the risk associated with that acquisition. This paper describes the repeatable methodology the Institute for Defense Analyses developed to assess the national security risk posed by the acquisition of U.S. biological datasets by strategic competitors and illustrative case studies that demonstrate the use of this methodology. This abridged paper documents a single illustrative case study pertaining to agricultural bioprocessing data. Specifically, the example examines China's use of privately held bioprocessing data to develop genetically engineered disease-resistant pigs. The methodology described in this study will be used as a framework to characterize the relationship between biological datasets, applications of those datasets, and national security risk in follow-on analyses.

15. SUBJECT TERMS

biological data; Biotechnology; bioeconomy; bio-cybersecurity; bioinformatics; genomics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Patrick Lee
U	U	U	U	84	19b. TELEPHONE NUMBER (Include Area Code) (703) 571-4028

This page is intentionally blank.