Portfolio Modeling and Forecasting of Single-Use Rare Disease Treatments

by

Olivia Claire Goldman
BS Supply Chain Management, Arizona State University, 2016
BS Marketing, Arizona State University, 2016

and

Jordan Michael Leising
BSIE, Purdue University, 2015

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Signature of Author: _____

<div align="right">

Department of Supply Chain Management
May 14, 2021

</div>

Signature of Author: _____

<div align="right">

Department of Supply Chain Management
May 14, 2021

</div>

Certified by: _____

<div align="right">

Dr. Jarrod Goentzel
Director of MIT Humanitarian Logistics Lab
Research Scientist at MIT Center for Transportation and Logistics

</div>

Accepted by: _____

<div align="right">

Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

</div>

Portfolio Modeling and Forecasting of Single-Use Rare Disease Treatments

by

Olivia Claire Goldman

and

Jordan Michael Leising

Submitted to the Program in Supply Chain Management
on May 14, 2021 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

Abstract

Biopharmaceutical companies are increasingly exploring cutting-edge novel gene therapies (GTs) in an effort to cure rare diseases. This capstone develops and tests a practical forecasting framework for sharing capacity across Roche's evolving GT portfolio and driving strategic global supply chain network design. Our problem is challenging, even by the highly regulated pharmaceutical industry standards, with: (1) substantial R&D and mergers and acquisitions investments, (2) some of the world's smallest disease populations, (3) one-time patients, (4) lacking commercial infrastructure, and (5) scarce historical or long-term pipeline data. We created three forecast types based on the target disease state knowledge available to predict an asset's prevalence and incidence patient adoption curves. The resulting asset forecasts are also aggregated into a comprehensive portfolio dashboard. Our user-friendly point model enables stakeholders to market size the prospective current pipeline and risk pool portfolio capacity by clinical phase. We then applied simulations to illustrate long-term product launch scenarios. These tools cater to various stakeholders helping address the key GT production planning and asset targeting problems. Roche has already began utilizing our capstone to methodically consider unknown future assets, with unknown orphan disease severity or populations, in their strategic make vs. buy GT network design decisions.

Capstone Advisor: Dr. Jarrod Goentzel
Title: Director of MIT Humanitarian Logistics Lab and Research Scientist at MIT Center for Transportation and Logistics

Acknowledgements

We first would like to extend our gratitude to Dr. Jarrod Goentzel. Thank you for the opportunity to work on a fascinating capstone under your advisement. As students, we sought feedback from "the academic in the room" and found your insight essential to creating a final product that we can all be proud to share. Pamela Siska helped ensure that our report transcends siloed supply chain management or pharmaceutical industry audiences. Our MIT SCM family, including classmates and CTL staff, were exceptional resources throughout the entirety of this project.

We of course would like to thank the Roche supply chain organization, particularly Pierce Kent and Nicholas Wingdahl, for entrusting us with this important project. The larger Roche team provided invaluable feedback and coaching along the way. The cohesive effort from our sponsor made the quality of research and mutual learning possible.

Finally, a special recognition to our families for their continued support throughout this project and our time at MIT. We would not have been successful without Jennifer Leising, Gina Relkin, and other family and friends.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Problem Background

Gene therapies (GTs) are cutting-edge treatments forming the next generation of biopharmaceuticals. These assets utilize innovative technology to manipulate the human genome in a revolutionary effort to be a single time use cure for diseases. Ongoing clinical trials continue to grow the potential disease use cases for GTs. Rare diseases present some of the most exceptional application opportunities for GTs, many which do not have a single approved medical treatment available (Dunbar et al., 2018). Unfortunately for supply chain experts in this field, there are many challenges with GTs. Five key features make our capstone problem unique: (1) small patient populations, (2) single-dose nature of GTs, (3) theorized unique adoption curves, (4) level of investment in GTs, and (5) lack of current commercial manufacturing footprint. This capstone explores how pharmaceutical companies developing GTs can forecast individual assets, and then leverage a portfolio approach to better estimate the volume that their supply chain networks will need to serve.

1.1.1 Demand: Novel Treatments for Rare Genetic Diseases

Genetic diseases are caused by gene mutations. Rare diseases can stem from the abnormality of a single gene (monogenic) or a set of abnormalities in several genes (polygenic). Given the relationship between genes and rare diseases, it is clear how GTs seem to target rare disease treatment needs. What may not be as immediately apparent is the breadth of potential GT impact on rare disease patients: over 80% of all rare diseases have been traced back to only one causal gene (TRND, 2017). The criteria for classifying a disease as "rare" differs by country. Rare diseases have been defined as disorders affecting fewer than 200,000 patients in the United States (US) since the Orphan Drug Act of 1983, a US government-supported initiative aimed at encouraging rare disease drug development with incentive programs (GARD, 2021). The European Union quantifies conditions as rare diseases when fewer than 1 in 2,000 individuals are affected (GARD, 2021). Gene therapies help fill a service gap for these smaller, and historically disproportionately underserved, patient populations.

Gene therapy has been the "next big thing" for nearly half a century. In the 1960s, Dr. Lorraine Kraus was the first scientist to place functioning DNA in a mammalian cell. Interest in developing innovative genetic medical treatments drastically increased in the 1980s. Gene therapy and cell therapy (CT) research continued to gain momentum into the next decade. US federal regulators reacted to these significant treatment development efforts with a regulatory notice, recognizing a need to provide scientists statutory guidance for the promising future GT and CT products (FDA Federal Register Part II, 53248, 1993).

Traditional drugs typically focus on symptom management over a patient's lifespan. More advanced drugs can proactively slow the onset or reduce the severity of rare disease symptoms, but still often require long-term treatment without curing the root cause. Gene therapies instead focus on attempting to cure chronic conditions by treating the mutated genetic source with one dose (TRND, 2017).

Relating GTs and CTs to common traditional medicines, both treatments use genetic material as the "active ingredient" to combat rare gene abnormalities (NHLBI, 2018). Research over the past several decades has spanned ex vivo and in vivo therapies. CTs frequently target oncologic conditions by using a patient's own somatic cells. These selected cells are extracted to create personalized vectors that are introduced outside of the human body in an ex vivo therapy method before ultimately delivering to patients in vitro. In contrast, general GTs deliver in vivo vectors of general genetic material per disease state to treat conditions. Figure 1.1 details adenovirus vector components in a visual overview of how gene therapies deliver specific DNA to patient cells. Although technological advances have enabled viral vectors to package genetic material, these delivery vehicles also bring use and cost implications.

Figure 1.1

An Illustration of How Gene Therapy Works

A new gene is inserted directly into a cell. A carrier called a vector is genetically engineered to deliver the gene. An adenovirus introduces the DNA into the nucleus of the cell, but the DNA is not integrated into a chromosome.
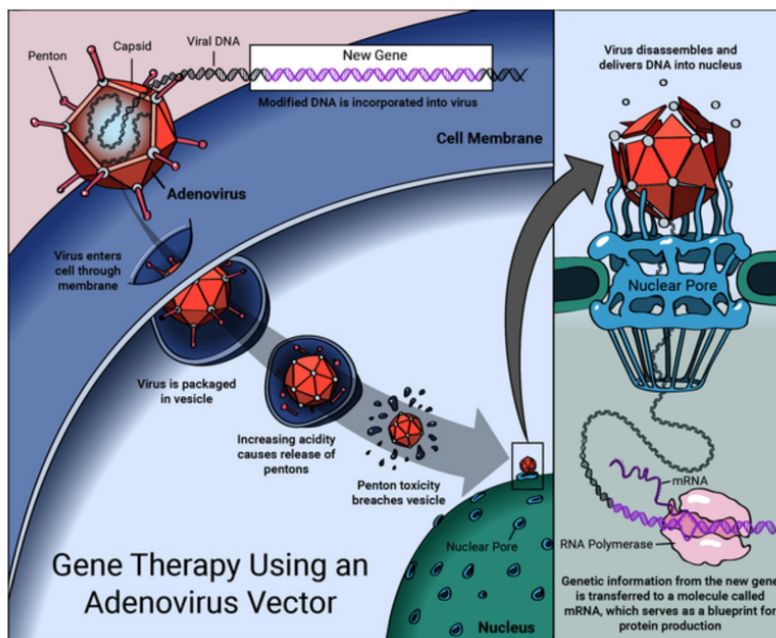
From U.S. National Library of Medicine (NLM). (2020, September 17). *How does gene therapy work?: MedlinePlus Genetics*. Medlineplus.Gov; National Institutes of Health (NIH). https://medlineplus.gov/genetics/understanding/therapy/procedures/

A key ramification of using viral vectors is the inability to administer more doses. After exposure to a viral vector, an immune system develops an immunogenic response to further doses (Sack & Herzog, 2009). Henceforth, patients who have been administered a gene therapy are unable to receive the

treatment again. Superior treatment efficacy is critical to establishing the value of expensive GTs for eligible patients (Science Direct, 2018) with current technology limitations for one-time administration. Most rare diseases do not have existing treatment options (TRND, 2017) that risk dampening gene therapy adoption curves. For example, Luxturna has been able to maximize eligible inherited retinal diseases (IRDs) market access with no viable competing treatment value propositions. Luxturna has been able to capitalize on offering not only superior efficacy between absent medicinal alternatives, but also has driven user adoption by providing better efficacy than currently offered by the standard of care (Science Direct, 2018). While continued gene therapy research progress supports efforts to make these breakthrough treatments part of the future standard of care (Dunbar et al., 2018), we anticipated gene therapy adoption rates to be higher among individuals affected by the minority of rare diseases with traditional symptom management treatments (TRND, 2017) and the greater rare disease population without specific orphan treatments available.

1.1.2 Commercialization Risk: Gene Therapy Clinical Trials

Clinical trials traditionally consist of three distinct phases: I, II, and III. During each clinical phase, clinicians, researchers, and regulatory agencies attempt to discern different aspects of the drug. In Phase I, a small trial of healthy individuals is used to evaluate and confirm how a drug works in the human body, confirm the safety of the drug, and understand dosage impact. During Phase II clinical trials, researchers identify any side effects and potential efficacy of the drug on patients that have the disease. Phase III is a large-scale trial to look for any additional adverse reactions to the drug and confirm efficacy endpoints (Commissioner, 2020). GTs are but a subset of products in the larger biopharmaceutical industry. Safety of a new drug, efficacy, investment costs, approval lead time, and patient population sizes are barriers to the general biopharmaceutical drug commercialization process. Aside from the average commercialization success rate for any new clinical asset of ~12%, Figure 1.2 further quantifies these challenges are further quantified by the average investment $2.6 billion for research and development (R&D) costs, over 10 years pursuing U.S. Food & Drug Administration (FDA) approval, and difficulties of recruiting trial volunteers. (PhRMA, 2015, pp. 13, 19). Figure 1.2 depicts the substantial difficulty of securing FDA approval to commercialize any biopharmaceutical drug.

Figure 1.2

Biopharmaceutical Clinical Development Process

THE BIOPHARMACEUTICAL RESEARCH AND DEVELOPMENT PROCESS
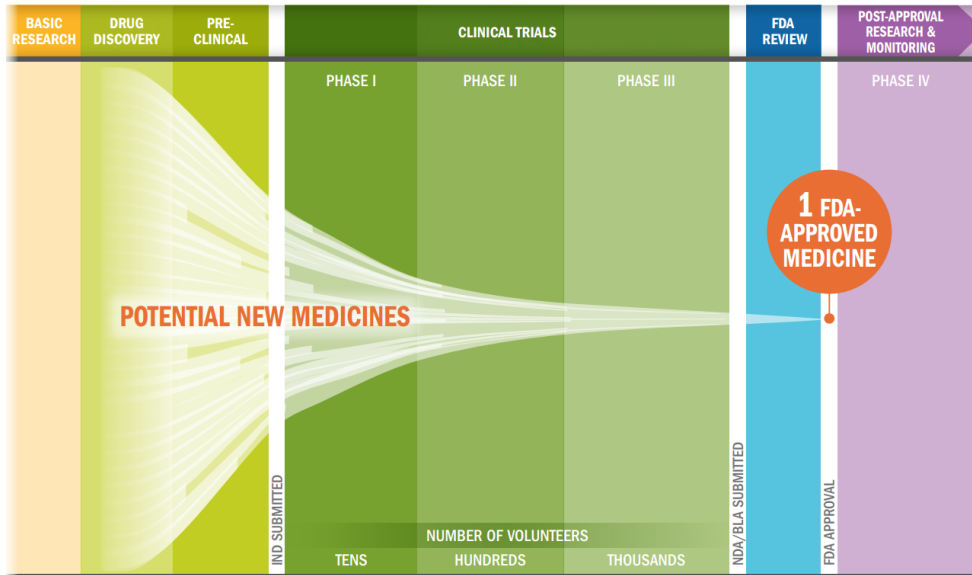
Note. The narrowing white area representing potential new biopharmaceuticals, such as GTs, shows how these assets are continuously removed from the pipeline throughout the clinical trials process. From PhRMA. (2015). *Biopharmaceutical Research & Development: The Process Behind New Medicines*. PhRMA.

At the conclusion of each phase, researchers and regulatory agencies work together to identify if an asset is ready to move forward and when. Regulatory agencies are often very concerned about the safety profile of the drug, whereas pharmaceutical manufacturers are additionally concerned about the efficacy. Clinical trials come with a hefty price tag, with ranges by therapeutic area of $1.4M - $6.6M for Phase I to $11.5M - $52.9M for Phase III (Sertkaya et al., 2016). In addition to the safety and efficacy outcomes, the high cost of clinical trials force companies to evaluate the business case throughout the process, as they look at the competitive landscape. Clinical trials typically require a considerable time investment, as well. Zivin (2000) estimates an average of seven years for a product, frequently referred to as an "asset" in the pharmaceutical industry, to conclude the clinical trials process. Once, and if, a new medicine has concluded the three phases, each local regulatory agency, such as the FDA or European Medicines Agency (EMA), makes independent reviews and approvals for new medicine (Gene Therapy Net & Bleijs, n.d.). Therefore, it is often the case where a new drug is being commercialized in one country but has yet to reach commercial status in another country.

Gene therapies do not always follow the traditional clinical trials process. First, considering the scope of rare diseases, accumulating hundreds to thousands of patients ranges from incredibly challenging to impossible (PhRMA, 2015, p. 13). Drug intensity and administration also affect the ability to run the industry standard randomized clinical trial. For example, Luxturna is delivered to the patient as an

injection to the back of the eyeball (Dunbar et al., 2018). Administering a control group in the trial can be considered both unethical and dangerous (PhRMA, 2015, pp. 11-13). Therefore, GTs do not always follow the standard Phase I, Phase II, Phase III clinical trial process. Per the American Society of Gene and Cell Therapy (ASGCT) as of October 2020, nearly a quarter of the ongoing USA Phase I, Phase II, or Phase I/II clinical trials (215/876) are in combined Phase I/II trials, rather than in either of the two independent stages.

1.1.3 Supply Chain Network Planning: Uncertain New Product Manufacturing and Distribution

To understand the quandary supply chain networks face trying to deliver gene therapies to commercial orphan disease patients, it is important to define pharmaceutical companies' current business model. The commercialization of large molecule assets, such as new oncology medication, have rarely caught pharmaceutical companies flat-footed. Molecules in the pipeline are known and understood by scientists and regulatory bodies alike (FDA, 2020), creating a solid understanding of what is needed to get the asset to market. With this understanding and history of similar assets, companies have been able to create models predicting the likelihood of commercialization of an asset by clinical trial stage. Such models enable epidemiologists and demand planning teams to obtain solid understanding of patient eligibility by molecule. Additionally, current assets often target diseases with the large affected populations. Lastly, product demand deterioration is controlled by either competitor entry or end of a patent life. The defined time horizon of patents and publicly available FDA clinical trial status information provides an outlook on the competitive landscape for advanced commercial demand forecasting. Pharmaceutical companies can project and adjust the long-term need of manufacturing plants using a combination of the main current business model factors. These factors culminate to enable pharmaceutical companies to have confidence in strategic decisions for capacity of drug substance, drug product, packaging, and distribution around the globe.

In contrast, GTs that target rare diseases challenge the underlying assumptions of the typical network design strategy. While any new-to-world offering requires some level of network design decisions, the profoundly unique GT treatments require at least some new network development. These costly network development decisions cannot currently be reliably supported by probability of commercialization, which is not fully understood for GTs. Although over many GT treatment clinical trials have been started, only three gene therapy products have been approved by a regulatory agency. More stringent regulations on GT treatment development understandably increase the difficulty to obtain regulatory approval. Layering in the small sample sizes available for targeted rare diseases, standard clinical trial study participant guidelines are further complicated. Supply chain, finance, and other key organizational stakeholders are ultimately left with little historical data to estimate the mere likelihood of a rare disease GT asset entering the market.

Nearly a thousand companies are developing cell and gene therapies (Deloitte et al., 2020). The market continues to grow as the immature GT space evolves, although few firms are financially or physically capable of commercializing products. However, the GT main players have emerged mostly via acquisition of small biotechnology companies or licensing agreements to distribute individual assets. Even by pharmaceutical industry standards, this space is incredibly expensive to operate in and highly

regulated across different federal bodies. Large pharmaceutical companies with global means, such as Roche, Novartis, and Pfizer, have progressively been adding cell and gene therapies (CGTs) to their pipelines. Figure 1.3 shows the increasing consolidation of cell and gene therapy producers over the past five years via M&A, despite the growing number of entering competitors.

Figure 1.3

Pharmaceutical Industry Mergers and Acquisitions for CGTs (2015-2019)

M&A activity in the cell and gene therapy industry rose sharply in 2019
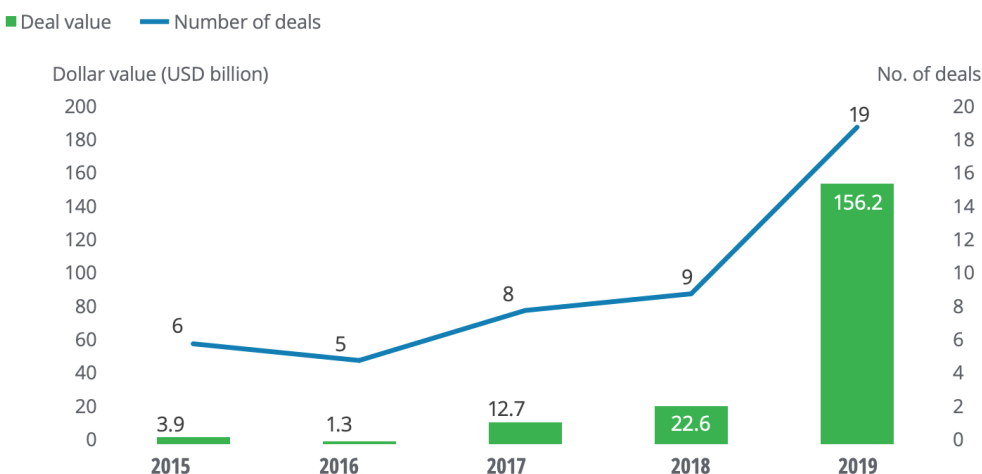
Source: Deloitte analysis of Thomson One Research data.

From Deloitte, Mooraj, H., Kawalekar, O., Gupta, L., & Shah, S. (2020, April 17). *Cell and gene therapies: Delivering scientific innovation requires operating model innovation*. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/life-sciences/operating-models-for-gene-cell-therapy-manufacturing-process.html

Pharmaceutical companies brave and resourceful enough to enter the precarious GT space are forced into a difficult position, considering future scalability and financial value when investing in their new operations. The ability to leverage current biotechnology manufacturing equipment or processes varies by GT treatment type. For instance, different viral vectors require different manufacturing methods. In addition, large pharmaceutical companies focused on global production can struggle to adapt to the substantially smaller batch sizes. The long-term decision factors for GT manufacturing are further exacerbated when serving globally disparate, finite rare disease patient populations.

Few pharmaceutical companies have full-scale GT production. This high-stakes example of a classic "make versus buy" supply chain scenario necessitates substantial investment for any of the primary options: (1) create in-house manufacturing capabilities, (2) contract with one of the select existing GT manufacturers, or (3) acquire a GT manufacturing facility. The latter two options are inorganic growth approaches used more frequently by larger pharmaceutical companies, but they often lead to an unclear understanding of nuanced rare disease GT assets in an already sophisticated pipeline.

1.2 Motivation and Impact on Partner

Our partner for this capstone, Hoffmann-La Roche (Roche), has recently made acquisitions and licensing deals for assets in the rare disease GT space. Currently, some medicine is being manufactured by a selection of vendors. Roche has limited internal capacity to produce drug substance, drug product, or packaging and labeling related to gene therapies, so they are looking to identify the next steps in the future of manufacturing within this space. Due to the unique nature of gene therapies, the approach to network design of gene therapies cannot be based on the commercial success of a single asset. Manufacturers must consider how and whether pipeline assets combine to create an economy of scale.

1.3 Problem Statement

A functional model framework to forecast demand of rare diseases is needed to support network design decisions for manufacturing Roche's growing GT asset portfolio. The diverse set of disease states, existing relevant data, molecules, asset stages, and regional market access determinants require a scalable model that can also be focused, or "sliced", as it is applied to each progressing asset.

The quantity and positions of competitors vary vastly by disease state in the emerging GT "curative" treatment market. Differences between GT assets for each disease state and the respective impact on patient eligibility are uncertain. Several disease states have more than one GT asset in the Roche pipeline alone. However, distinct features or market availability of non-Roche potential GT treatments can threaten competitive advantage. Roche's current GT assets span across the entire pipeline, from preclinical to a single commercial asset. The vast majority of pipeline assets are precommercial.

Which variables and assumptions should be included in this rare disease forecasting model. Furthermore, some attributes that help determine finished good market size and manufacturing requirements are also unknown, such as the treatment drug substance volume, patient eligibility, and insurance/financial coverage per asset. The disparities in epidemiology of each disease state by region are imperative to consider in our model for target market insights. Ultimately, we aim to answer the question: Can a forecasting model support strategic network development for sharing capacity across the evolving CGT portfolio?

To answer this question, this capstone walks through three key steps: asset level forecasting, portfolio aggregation, and scenario planning. First, we developed a novel way to forecast assets. Second, we used these forecasts to estimate the current production planning problem based on Roche's portfolio. Finally, we created a simulation to project the needed volume for several key scenarios. These three scalable tools provide visibility into the short-term commercial capacity needs for Roche's current pipeline, as well as asset selection guidance based on the projected actualization of different annual product launch targets.

2 Literature Review

This capstone explores how pharmaceutical companies developing novel GTs can move from modeling individual assets to the entire product portfolios their supply chain networks will need to serve. To support a multi-asset model, two key aspects of the portfolio breadth must be evaluated: forecasting demand and the likelihood of each asset being FDA-approved. First, research on new product forecasting will be presented. Then, this chapter will consider how to model clinical assets as they move through the portfolio. The niche gene therapy space offers limited commercial market data across 3 total gene therapies approved by either the FDA or the EMA. Model complexity is increased by the problem statement's narrower focus on rare disease gene therapies. Broader biopharmaceutical and supply chain research will also be assessed for consideration of future portfolio network design decisions.

2.1 New Product Forecasting

When forecasting a new product, company launch team stakeholders have two primary considerations: 1) identification of potential market size and 2) projected customer adoption rate of the product. The difficulty of quantifying these considerations varies by industry. Nuanced prescription pharmaceutical markets pose regulatory and patient (customer) eligibility challenges that are largely not applicable to average consumer goods. In market sizing, companies are unable to build new markets. Instead, products are developed to meet a clinically diagnosed patient population (Deloitte et al., 2020). Companies have little to no influence on creating or enabling more potential patients.

Forecasting new categories requires merging the forecasts of each individual product. With product lifecycle closely related to patent lifespan, aggregate multi-product pharmaceutical forecasts are especially sensitive to the model time periods and interrelated variables (Merkuryevaa, Valbergab, Smirnov, 2018). Cook (2006) approaches new pharmaceutical product forecasting with disease state market modeling. This methodology suggests a decision-based approach to define the market and simplified and intently select underlying data attributes (Cook, 2006). Even after defining a market size forecasting algorithm, pharmaceutical product potential markets are subject to direct and indirect dynamic adjustments (Cook, 2006).

Switching to new product adoption curves, the three currently approved gene therapy products (FDA, 2020, EMA 2019) provide little data on actual GT demand observations and commercial production. Quinn et al. (2019), Touchot and Flume (2015), and Mullin (2017) all recognize the GT pricing anomalies for particular disease states with limited populations. Domestic and global healthcare policy pose significant adoption rate implications, such as affecting market access, within the problem scope. Furthermore, there is a lack of clarity around current commercial processes. Before 1985, companies often published peak sales forecasts for individual molecules. However, due to a plethora of erroneous forecasts, companies no longer publish asset level projections (Cook, 2006). This review will start by looking at market sizing of new pharmaceutical products that will enable shaping of the new product adoption curves.

2.1.1 Market Sizing New Pharmaceuticals

Pharmaceutical companies employ qualitative and quantitative forecasting methods to build dynamic new product forecasts for the nuanced commercial gene therapy market sizes. This capstone will consider the merits of two key market sizing strategies highlighted by Cook (2006): patient models and prescription models. As shown in Figure 2.1, patient models start with a top-down approach. Pharmaceutical forecasting teams incorporate various data input on incidence, prevalence, diagnosis, market access, and patient eligibility (e.g. label indication for patients older than 18) to identify the maximum potential population. Such top-down approaches are more consistently used in rare genetic applications than the prescription-based approaches.

Figure 2.1

Top-Down Forecasting Funnel Process
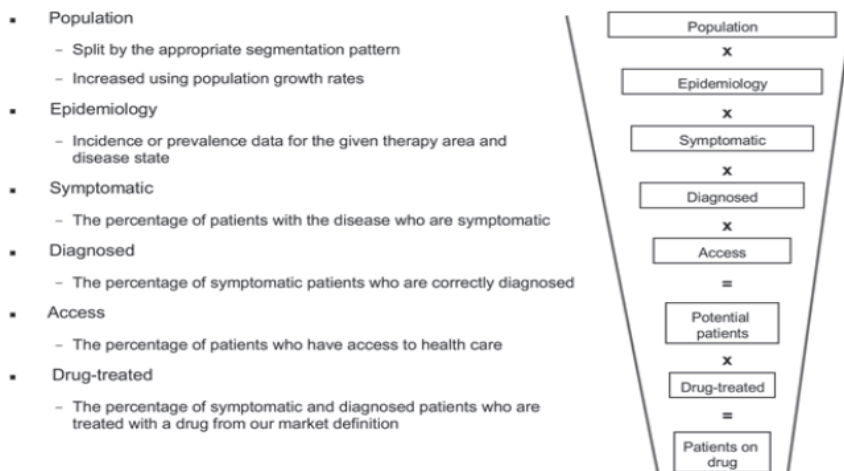
Note. The above market sizing model depicts Cook's (2006) process to develop patient-based algorithms for determining the total potential market of new pharmaceutical assets. From Cook, A. G. (2006, p.38). *Forecasting for the Pharmaceutical Industry: Models for New Product and In-market Forecasting and how to Use Them*. Gower. https://books.google.com/books?id=5IaddXKSNTkC

As an alternative to patient models based on epidemiology, companies can consider utilizing a prescription model. Prescription models leverage data surrounding the volume of written and filled prescriptions for a disease state. Grabowski et al. (2007) use a prescription model to estimate the market share of generic biologics market entry. Prescription model methodology relies on a medication available in the market, which data can be used as the basis for generating new pharmaceutical product forecasts. This model does not align with the realities of the rare disease gene therapies market, where therapies are sparse. Pursuing a patient-based algorithm is more suitable for this capstone because it will enable us to capitalize on the limited patient data available for each disease state.

2.1.2 Gene Therapy Adoption Curves

Once a total market size has been identified, companies must evaluate how much market share a medicine can obtain along with the product life cycle. This capstone will investigate models that project market share along with estimated cumulative time to peak demand. Pausing for a second, let us define two key epidemiological terms: prevalence and incidence. Prevalence is the is a population that currently has the disease (e.g., 1 in 10,000 people). Alternatively, incidence is the rate at which people are diagnosed with the disease in question (e.g., 1 in 10,000 births, 1 in 100,000 people per year).

Teixeira et al. (2019), along with our host company, theorize that rare disease gene therapies will have intense adoption rates, with steep curves after the prevalence population has been addressed. Quinn et al. (2019) share this belief, expecting 70% prevalence adoption two years after a market launch. Product demand will then be mainly driven by incidence rate after the peak patient adoption is actualized. As indicated by Vakratsas and Kolsarici (2008) and Quinn et al. (2019), effective models need to separate the rates of prevalence adoption (previously diagnosed patients) and incidence adoption (newly diagnosed patients). This approach seems particularly logical when the medicine is either substantially more efficacious than the standard of care or the medicine is new-to-world for treating a certain disease state. The latter is expected for many gene therapies. Although individual rare diseases populations are small, prevalence population is limited to under ~200,000 people in the US and ~246,000 people in the EU (Puiu & Dan, 2010), less than 13% of rare diseases in the US have FDA-approved drugs or biologics (Hahn & Abernethy, 2020).

In contrast to more subjective market share and time to peak models, we will explore objective mathematical adoption modeling options. Vakratsas and Kolsarici (2008) propose a two-phase generalized diffusion model. The authors recognize that innovative products will launch with accumulated patient demand waiting as untreated members of the prevalence population. One of the most defining features of this capstone is that rare disease states have finite patient populations at any given time, unlike long-term infectious diseases such as HIV/AIDS. After treating most of the cumulative prevalence population, the incidence population subset will drive demand as onset occurs and patients qualify for treatment. This fraction of the total population will represent most future market share and merits an independent incidence adoption curve. Vakratsas and Kolsarici (2008) have success forecasting spectrum diseases (e.g., depression) but not in dichotomy (e.g., HIV). Ding and Eliashberg (2008) approach new product forecasting via a Markov model. Additionally, although they address two of the three key decision makers (patient, payer, provider) by including important clinical and patient decision variables, Ding and Eliashberg (2008) fail to address the power that insures have in pharmaceutical uptake. Conversely, diffusion modeling is not well suitable for branded pharmaceuticals. Unlike standard consumer goods, pharmaceuticals must prevail through multiple stakeholder decision points to be filled for patient consumption. Cook (2006) suggests that the combination of three key decision makers creates a difficult modeling environment.

2.2 Commercialization Likelihood

Introducing a new medicine to market for commercial availability is a long, structured process. Once discovered, pre commercial medicines are segmented into two phases: preclinical assets and clinical

trials. Given the problem scope, this capstone will largely focus on the process and strategy to model gene therapies through the arduous clinical trial Phases I-III.

Assets are scrutinized for efficacy and safety by each clinical trial phase before receiving FDA approval, as introduced in the Introduction chapter. The corresponding probability that a gene therapy will successfully obtain approval for serving the commercial patient population is partially determined by its clinical stage. To address commercialization likelihood, Quinn et al. created a segmentation of four categories to forecast success based on historic clinical trial performance. The authors combine this approach with a Monte Carlo simulation to arrive at a probabilistic number of launches products (2019). Alternatively in a thesis focusing on early state portfolio, Heyman (2010) leverages company information around duration within each step. This data is used to develop a Beta distribution to project success likelihood paired with a Monte Carlo simulation to estimate the total output. Key assumption to both simulations is using historic performance as an indication of future success.

Additional complexities exist when considering the relative lack of commercial success GTs have had throughout their history. It was not until 2017 that the FDA approved the first commercial GT for the US, Kymriah (FDA, 2017). This creates a fundamental disconnect, as little data is available to predict the future likelihood of an asset making it through the clinical trial process. Nearly 60 years of research have led to numerous molecules that fail to accomplish what was expected during the clinical trials process. To overcome this, one might assume in their model that historic developed technology platform (i.e., small molecule) data will be reflective of future GTs success. Whereas that might be a fair assumption, biostatisticians disagree on the likelihood of pipeline assets advancing through each phase. Wong et al. (2019) compares different literature and gives likelihoods of success by clinical phase for various factors (e.g., indication, biomarker utilization).

2.3 Summary

Selecting the methodology for demand forecasting the rare disease gene therapy asset portfolio entails two primary potential modeling frameworks: eligible patient sizing and quantifying adoption curves. This capstone will first work via a top-down forecasting approach to identify the eligible patient pool who will become consumers of our asset. We will then work to quantify the shape of the adoption curve driven by various disease state inputs. This capstone ultimately combines these frameworks in a single model to provide a basis for supporting supply chain network design decisions. Researchers have been able to successfully develop such comprehensive new product demand forecasts in various industries. However, pharmaceutical new product forecasting proves to be distinctly challenging due to the industry's renowned high level of complexity. Pharmaceutical supply chain researchers caution that this complexity is not only a performance barrier for manufacturing and distribution, but also impacts crucial new product forecasting accuracy upstream (Merkuryevaa, Valbergab, Smirnov, 2018, p.4). The core framework should forecast an individual asset's lifecycle over time and then combine these assets to arrive at a gene therapy portfolio forecast.

Developing the forecasting model for this capstone requires methodology that carefully evaluates a greater margin of error risk, driven by the substantial uncertainty around likelihood of commercialization and competition for each asset. While it would be useful to develop a robust

simulation for more targeted stage-gate probabilities of commercial success, focusing on maximizing robustness of our market sizing and adoption curve forecasting with accepted success probability assumptions will deliver the most value to the capstone host company.

3 Methodology

Our capstone models the potential demand forecasts for a pharmaceutical company's GT asset portfolio. These GT assets can be in any clinical trials phase described in section 1.1.2, or even in an approved state. This process is broken into two distinct steps: asset level forecasting and portfolio analysis.

Starting with asset level forecasting, we have developed a forecasting methodology that considers two key challenges: varying extent of knowledge and various clinical stages of assets. It is difficult to assess historic performance with just three approved in-vivo GTs, the first of which was approved in December 2017 (FDA 2020). We instead concentrated on making the point model robust for future adaptability, which will provide the most realistic value long-term. Additionally, we recognize that not all assets in a portfolio will achieve clinical success, therefore the model enables volume discounting based on a success likelihood. The model was built for our corporate partner to incorporate new data as their niche GT portfolio and the overall market simultaneously mature. The forecasting model can then be used to either project current pipeline volume, or to design and run potential scenarios considering the number of assets to be commercialized.

3.1 Comparing Asset and Portfolio Level Views

To match the uniqueness of each disease state and asset that comprise a portfolio at any given time, our methodology enables assets to be evaluated individually and then combined. Figure 3.1 illustrates the top-down methodology we created by adopting elements of Cook's (2006) patient-based algorithms described in section 2.1.1. The methods used to forecast assets are dependent on the current knowledge level of an asset's targeted disease state. With this in mind, we grouped assets into the following categories, in descending order: specific knowledge (Type 1), moderate knowledge (Type 2), and minimal knowledge (Type 3). The different asset forecast types allowed us to vary the key input variables based on what information is available for the targeted disease state. The five Type 2 assets currently in pipeline served as the basis for developing our initial framework and then adjustments made to the other forecast types for our model.

Figure 3.1

Asset Level to Portfolio Modeling
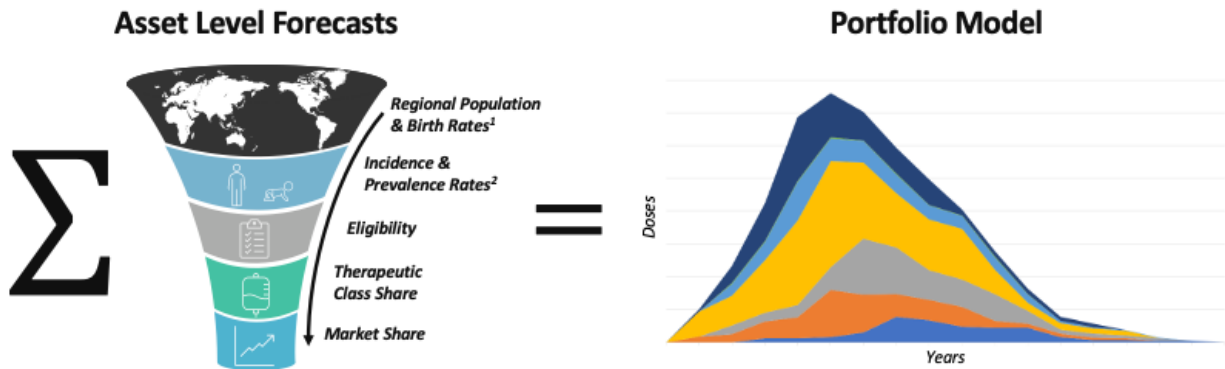
At an asset level, the model first seeks to evaluate the asset's targeted disease state. Disease state attributes were used to model the prevalence (previously diagnosed) and incidence (newly diagnosed) populations. Once combined, the model can risk-adjust the projected volume based on the likelihood that the asset is commercialized. The output forecasts patient demand over time, which directly related to the number of doses that must be produced to serve.

Once per-patient volumes are finalized, our partner can use either total volume or risk adjusted volume to create various scenarios to predict commercial manufacturing production. Taking a holistic approach, the model can help support strategic long-term manufacturing and distribution decisions to serve the projected addressable market size. The above process will need to be repeated for a variety of assets within a portfolio, at which time we can summarize the entire addressable pipeline to project a total volume number. Each asset level forecast has unique intricacies that complicate forecasting.

3.1.1 Forecast Types

Every asset forecast is unique. Pharmaceutical companies have dedicated forecasting teams for effectively sizing the market and launching new product. Additionally, large pharmaceutical companies have many assets in their pipeline, with assets constantly coming in and dropping out. With the intention of not investing time, effort and money on assets that might not make it to market, companies spend the majority of their forecasting knowledge on assets that are either launched or very close to market launch. Given that many of the assets we are considering are prelaunch, this creates an issue. We grouped assets based on the general current knowledge level for developing the following three forecast types (compared in Table 1) to address the varying levels of public data, variables, and assumptions:

Type 1: Specific knowledge asset forecasts, where our partner company forecasting team has data on the specific disease state and volume projections for the asset. Most of these specific knowledge assets are closer to market, currently moving through Phase II or III clinical trials.

Type 2: Moderate knowledge asset forecasts, where it is known which disease state the asset was developed to treat. Publicly available resources can be used for epidemiological inputs. Most of these moderate knowledge assets have the potential to go to market within the next10years and are currently moving through Phase I or II clinical trials.

Type 3: Minimal knowledge asset forecasts, where the future asset's disease state is unknown. These are forecasts for potential preclinical research or M&A deals. Publicly available resources must be used for epidemiological inputs of various disease states.

Table 1

Forecast Types

Type	Asset Knowledge	Target Disease State	Epidemiology and Market Data	Typical Clinical State	Typical Launch Range
1	Specific	Confirmed	Internal and public specific data on the disease state	Phase II/III	0-3 years
2	Moderate	Confirmed	Public general data on the disease state	Phase I/II	3-5 years
3	Minimal	Unknown	Public data on types of rare diseases	Any	0-15 years

3.2 Base Use Case: Forecasting a Type 2 Asset

To demonstrate how the forecast methodology works, we will first walk through a Type 2 forecast. Once we have walked through this foundational use case, we will compare the different variable input options and assumptions made to develop the Type 1 and Type 3 forecasts of model. Within the Type 2 forecast, we expect that our partner will have sufficient knowledge of the asset for entering needed inputs. Otherwise, the forecast should be modeled as a Type 3. Table 2 shows all the needed inputs that are used within the model and if it requires a user input.

Table 2

Type 2 Model Input Variables

Type 2 Forecast Variable	Input Method	Values
ph: Clinical phase of asset	User	PI, PII, PIII
r_{prev}^{mrk}: Baseline point prevalence rate	User	Percent
$r_{inc}^{mrk,type}$: Annual incidence rate and type	User	Percent
$share_{class}^{mrk}$: Class share of therapeutic	User	Percent
$share_{prev}^{mrk}$: Prevalence market share	User	Percent
$share_{inc}^{mrk}$: Incidence market share	User	Percent
elg: Estimated eligibility	User	Percent
sev: Severity	User	High, Medium, Low
ea: Early Adopter window	User	Integer
own^{mrk}: Applicable regional market	User	Yes/No
yr^{mrk}: Launch year of regional market	User	Year
yr^{pc}: Patient cliff year	Calculated	Year
pop^{mrk}: Population in a region	Systematic	Integer
$brth^{mrk}$: Annual births in a region	Systematic	Integer

Figure 3.2 shows an input dashboard for Europe Type 2 forecast. The pink tinted cells show key inputs that we will walk through to estimate Europe's volume for this asset. We then combine the various global markets to arrive at an asset level volume that will be risk adjusted for capacity planning.

Figure 3.2

Type 2 Market Level Forecast Example

Europe		Prevalence Data			Incidence Data		Total Volume	
		Population	1,059		Population	662		
		Build-Up Effect	1,059					
Baseline Prevalence	0.0028%		%.Year	Treatments		Treatments		Treatments
Annual Incidence Rate	0.0210%	2025	25%	265	2025	497	2025	761
Estimated Eligibility	30%	2026	25%	265	2026	497	2026	761
Estimated Launch	2025	2027	25%	265	2027	497	2027	761
Class Share	40%	2028	25%	265	2028	497	2028	761
		2029	0%	-	2029	497	2029	497
Prevalence Uptake		2030	0%	-	2030	497	2030	497
MRK Share	60%	2031	0%	-	2031	497	2031	497
Time to Share	4	2032	0%	-	2032	497	2032	497
		2033	0%	-	2033	497	2033	497
		2034	0%	-	2034	497	2034	497
Incidence Uptake	75%	2035	0%	-			2035	-
		2036	0%	-			2036	-
		2037	0%	-			2037	-
		2038	0%	-			2038	-
		2039	0%	-			2039	-

3.2.1 Epidemiology Data (Prevalence and Incidence)

In order to forecast how many does will be administered to patients, we need to understand how many people have a disease in a given geographical area. The model utilizes publicly available population and birthrate data at a country level to help solve this problem. For simplicity, the model will take inputs at the market level rather than the country level modeling. Our partner has classified 129 marketed countries into seven distinct regions, shown in Table 3.

Table 3

Market Classifications

Market	Label Regions
Europe	Western Europe
APAC	Asia Pacific
LATAM	Latin and South America
CEETRIS	Central & Eastern Europe, Turkey, Russia, and India
ME	Middle East
Africa	Africa
N America	US and Canada

With markets now defined, the model needs to understand how large the potential population in a market a disease. There are two main subsets of a disease state's patient population:

1. Prevalence (r_{prev}^{mrk}): Pool of existing population of individuals who have the disease when the asset is commercialized in a certain market
2. Incidence ($r_{inc}^{mrk,type}$): Rate at which individuals are diagnosed with the disease, who can potentially become new addressable GT patients after the asset is commercialized (newly diagnosed)

Prevalence concerns the current population with a disease. The model will use this proportion as the percent total population of the region that has a disease. In contrast to prevalence, there is incidence rate. Here we look to estimate how many new patients are diagnosed (or born) with a disease. An input is required for Type 2 forecast for annual incidence for every geography, but we do allow incidences rates to be measured in two ways: population based $r_{inc}^{mrk,pop}$ or birthrates $r_{inc}^{mrk,brth}$, where inc^{mrk} in Equation 3.7 would take the form of pop^{mrk} and $brth^{mrk}$, respectively. The model uses birthrates by region within the birthrate incidence to estimate the new patients born with a disease or a diagnosis rate to understand how many new patients are diagnosed each year. Matching inputs for each region, a calculation can project the current population within each region and the annual population that are diagnosed for a specific disease.

3.2.2 Eligibility and Class Share

Not everyone who has a disease will be prescribed a potential treatment and, although far less common in the rare disease space, not every disease will have only one available treatment option. The first two adjustments made to the population set are: (1) which patients are eligible to be addressed by the treatment and (2) the likelihood that a gene therapy will be the selected treatment modality.

Investigating patient eligibility, two major points arise. First, a significant portion of the rare diseases that gene therapies address onset in adolescence (shown in Appendix A). In turn, these disease treatments are developed for adolescent patients. FDA clinical trials have additional restrictions for testing medicines on children (FDA, 2015). Therefore, as an asset comes to market, the label claim age range might not be exhaustive of the potential patients. Age-driven eligibility restrictions are common even with over-the-counter pharmaceuticals, such as with Ibuprofen and other nonsteroidal anti-inflammatory drug (NSAID) capsules/tablets that are not suggested for children under seven years old (NHS, 2019). In addition to age, certain underlying conditions or comorbidities may make patients within the qualifying age range ineligible for receiving the gene therapy, thus, preventing therapeutic adoption. Revisiting the NSAID capsules/tablets example, packaging labels clarify eligibility restrictions to advise against adults who have uncontrolled high blood pressure or are pregnant from taking the product (NHS, 2018). This reasoning contributes to why the model requires a user input for the eligibility (elg) that will be used in owned markets.

Class share is another significant cut to the potential addressable population set. Again, this variable can be impacted by two key factors. If a disease state has a new gene therapy which must compete for patients against a non-gene therapy, it is highly likely that there will be differences in safety profiles and efficacy (Dunbar et al., 2018). This exact situation is unfolding today between three Spinal Muscular Atrophy therapies (Spinraza, Zolgensma, and Evrysdi) (Talbot & Tizzano, 2017, p. 529–533). Alternatively, a GT asset might have to compete not against other potential therapeutic classes, but palliative care. The most economical gene therapy treatment currently available in the US costs $425,000 (Jackson et al., 2020). Due to economic hurdles in various counties, it is likely that gene therapies are either not likely to be launched into the commercial market or have significantly smaller addressable commercial market populations. Thus, our model requires the user to input class share for each global market ($share_{class}^{mrk}$) to combat the challenges outlined.

These two key variables, eligibility and class share, are essential to help narrow down the total population that has a disease to the potential patient set. From here, the model will consider how effectively our partner will be able to convert potential patients to patients.

3.2.3 Market Share

The last numerical input the Type 2 forecast needs as an input to model an asset's population size is the projected conversion rate for patients who will undergo this GT to treat their disease. Building on the total disease population split between prevalence and incidence patients described in section 2.1.2, we theorized that the willingness or desire to receive a therapy will be different for each population subset.

Therefore, the model user must project both a market's prevalence conversation rate ($share_{prev}^{mrk}$) and incidence conversation rate ($share_{inc}^{mrk}$).

In contrast, the incidence population uptake rate is a single direct input. This conversion rate of the ongoing newly diagnosed patients then becomes a multiplier to systematically dampen the annual incidence population for addressable GT volume.

Our model separates the prevalence and incidence uptake user inputs to enable different population conversion rates. Furthermore, patient populations vary globally. To account for variability across applicable global markets, both population type uptake inputs are built into the model at the market level.

3.2.4 Prevalence Patient Population and Adoption Curve

Building our model with a top-down approach, it now has all the inputs and assumptions needed to identify the patient population. Equation 3.1 lays out how we use these variables to identify the total prevalence population that will become gene therapy patients.

$$patients_{prev}^{mrk} = pop^{mrk} * r_{prev}^{mrk} * elg * share_{class}^{mrk} * share_{prev}^{mrk} \qquad (3.1)$$

Whereas the calculation is rather straightforward, the complexity arises when considering the timeline on which this patient set will convert from potential patients to GT asset patients. After discussions with our corporate partner, we created a disease state "severity" variable (sev). We classified disease states into high, medium, and low severity based on the level of impact on a patient's quality of life. We will use this severity score to separate our prevalence population into two categories:

- Early adopter patients
- Non-early adoption patients

Every prevalence patient must be either an early adopter or not. We must assess how much of the prevalence population will rush to become patients. Our high, medium, and low severity classifications result in three allowable values for severity, or percent of patients as early adopters: 100%, 50%, and 0%, respectively. Conversely, that means that the non-early adopter population will either be 0%, 50%, or 100% of the prevalence population for the high, medium, and low severity scores.

Elaborating on the extremes, considering other industries, it might be naive to think that 100% of potential customers would rush to become adopters; however, considering the healthcare industry, for long-underserved rare disease populations facing a lack of alternatives and terminal diagnosis, 100% is not an unrealistic assumption for patients fighting for their lives. At the other extreme, assuming 0% of early adopters may be too low. Thinking about adulthood-onset diseases that are not life-threatening, patients could conceivably lack an early adopter mindset to attack less severe diagnoses with highly invasive, expensive, and relatively new gene therapies. With an understanding of how model will break the patient population into these segments, additional insights are to be gained with estimating how quickly the early adoption population will be serviced. The model accepts "early adopter timeline"

(ea^{mrk}) as a variable input for the number of years until the early adopters are served, shown in Equation 3.2.

$$yr_{ea}^{mrk} = yr_{launch} + ea^{mrk} \qquad (3.2)$$

For simplicity, we will cover only the uniform distribution assumption here. The normal distribution option will be addressed in 3.2.6, as our model theorizes that the conversation rate of early adopters will follow a uniform distribution.

$$f(t)_u^{mrk} = \begin{cases} \dfrac{1}{yr_{ea}^{mrk} - yr_{launch}}, & yr_{launch} \le t \le yr_{ea} \\ 0 \quad , & otherwise \end{cases} \qquad (3.3)$$

Equation 3.3 shows probability of conversion from potential patient to patient for year t for market *mrk* following the uniform (*u*) distribution.

If a patient is not going to be an early adopter, he or she will fall into the "normal adopter" category. In order to consider how these patients are going to be spread out, we must consider patent cliffs in pharmaceuticals. Two assumptions are built into the next steps here. First, the model assumes that these gene therapy assets are subject to an intense patent cliff. Second, as shown in Equation 3.4, the model assumes the patent cliff to be 10 years globally.

$$yr_{pc} = yr_{launch} + 9$$

$$(3.4)$$

With a patent cliff understood, the model theorizes that "normal adopters" are uniformly distributed between the market launch year and the asset's patent cliff, as in Equation 3.5.

$$g(t) = \begin{cases} \dfrac{1}{yr_{pc} - yr_{launch} + 1}, & yr_{launch} \le t \le yr_{pc} \\ 0 \quad , & otherwise \end{cases} \qquad (3.5)$$

Once the patient population is appropriately separated, the model can now estimate the total prevalence population in Equation 3.6.

$$prev(t)^{mrk} = f(t)_u^{mrk} * patients_{prev}^{mrk} * sev + g(t) * patients_{prev}^{mrk} * (1 - sev)$$

$$(3.6)$$

3.2.5 Incidence Patient Population and Adoption Curve

When considering the incidence rate adoption curve, the model makes no assumption in the ability to influence any of the user entered rates over time. Therefore, a top-down calculation of the incidence rate, incidence rate driver, eligibility, class share, and incidence rate uptake can project the annual incidence rate population. This annual number is expected to run from launch to expected patent cliff. Equation 3.7 shows the expected incidence population in year x.

$$inc(t)^{mrk} = \begin{cases} inc^{mrk} * r_{inc}^{mrk,type} * elg * share_{class}^{mrk} * share_{inc}^{mrk}, & yr_{launch} \leq t \leq yr_{pc} \\ 0 & , \quad otherwise \end{cases}$$

(3.7)

Recalling from section 3.2.1, incidence rate driver in a unique variable. It can either take the form of a regions annual birth rate or population, based on "type" input on variable $r_{inc}^{mrk,type}$.

3.2.6 Market Ownership and Summation

Most of these breakthrough drugs will have demand in several global markets. Compiling across regions, we must understand how launch years come into the model. Assets neither simultaneously nor equally become commercially available in all global markets. Therefore, we modeled the asset adoption curves specifically to the applicable markets where they are intended to launch. The regional market adoption curves are summed to provide the asset's global demand forecast over time.

Methodology steps described in sections 3.2.2-3.2.5 are each repeated for the markets noted by the capstone partner. All user input variables can be specified at the market level, except for disease severity. Before adjusting the available variables for individual markets, the user must first confirm the applicable regions and commercialization timing.

Figure 3.3

Market Ownership and Launch Table from Model

Region	MRK Ownership?	Launch Year
Europe	Y	2025
APAC	Y	2025
LATAM	Y	2026
CEETRIS	Y	2026
ME	Y	2026
Africa	Y	2027
N America	N	2024

Shown in Figure 3.3., a binary yes/no (mathematically 1,0), variable exists for each region to select where Roche will have market ownership. The launch year of the commercialized asset in each market region then must be identified. This open input variable enables the model to forecast assets over the full time horizon, regardless commercialization is spaced equally across market regions or if an asset is launched simultaneously in multiple markets.

The model can use this variable in equation 3.7 to calculate the total volume for a year, in a market.

$$vol(t)^{mrk} = own^{mrk}(inc(t)^{mrk} + prev(t)^{mrk})$$

(3.7)

3.2.7 Normal Distribution Option

The Type 2 forecast model has two adoption distribution types available for user selection. In addition to the uniform distribution discussed in section 3.2.6, users can alternatively choose to model the early adopter uptake following a normal distribution. Here, we theorize that the time at which the eligible early adopter population will convert to GT asset patients following a normal distribution between the launch year and the early adoption year cutoff. Equation 3.9 notes the distribution and how the model estimates the year when an early adopter will convert to a patient.

$$Year_{conversion} \sim N(\frac{yr_{launch} + yr_{ea}}{2}, 1) \qquad (3.9)$$

Slight adjustments must be made in practical use to the left tail of this adoption curve. When used within the model, it does not allow for an early adopter to become a patient before an asset launches. Therefore, values of the distribution prior to the launch year assumes that these affected prevalence population members will become patients during the launch year.

$$f(t)_n^{mrk} = \begin{cases} 0, & t < yr_{launch} \\ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \frac{yr_{launch} + yr_{ea}^{mrk}}{2})}, & t \geq yr_{launch} \end{cases} \qquad (3.10)$$

We expect no other changes for equation 3.11 when calculating the market prevalence for year x besides substituting the $f(x)_n$ for $f(x)_u$ equations (3.10 for 3.3).

$$prev(t)^{mrk} = f(t)_n^{mrk} * patients_{prev}^{mrk} * sev + g(t) * patients_{prev}^{mrk} * (1 - sev) \qquad (3.11)$$

3.3 Differences for Type 3 Forecast

With a firm understanding of the build-up on a Type 2 forecast, let us consider what assumptions and changes must be made for a Type 3 forecast. Unlike the vast amount of information needed for Type 2, Type 3 forecast will require users to input only five key pieces of information: disease prevalence range, disease severity, first launch year, market ownership, and clinical state. In this section we will walk through the underlying assumptions built and methodology used to build up forecast unknown asset portfolio. We start by showing in Figure 3.4 the input dashboard for a Type 3 asset. Comparing to Figure 3.2 and 3.4, there are significantly less possible inputs.

Figure 3.4

Type 3 Forecast Dashboard for an Asset

SPK 8832	PI	13.6%			Adoption Distribution Type	Straight Line			Incidence Driver?	Population	
MC Output	13.6%										
					Prevalence Data				Birth Driven Incidence		Total Volume
Disease Prevalence	6-9 / 10 000				Population	26,986			Population	11,259	
Uptake Estimate	Low										
First Launch	2043										

				Year	Treatments		Year	Treatments		Year	Treatments
				2043	898		2043	374		2043	1,272
Region	MRK Ownership?	Launch Year		2044	2,040		2044	850		2044	2,890
Europe	Y	2044		2045	2,620		2045	1,092		2045	3,712
APAC	Y	2044		2046	2,702		2046	1,126		2046	3,828
LATAM	Y	2045		2047	2,702		2047	1,126		2047	3,828
CEETRIS	Y	2045		2048	2,702		2048	1,126		2048	3,828
ME	Y	2045		2049	2,702		2049	1,126		2049	3,828
Africa	Y	2046		2050	2,702		2050	1,126		2050	3,828
N America	Y	2043		2051	2,702		2051	1,126		2051	3,828
				2052	968		2052	1,126		2052	2,094
				2053	315		2053	752		2053	1,067
				2054	77		2054	276		2054	353
				2055	14		2055	34		2055	48
				2056	2		2056	-		2056	2
				2057	0		2057	-		2057	0

3.3.1 Underlying Assumptions

A few assumptions were made to simplify the Type 3 forecast. These assumptions primarily center on market launch year sequence, market class share, and eligibility. Table 4 shows the sequence at which Type 3 forecasts are assumed to launch in various markets around the world. Additionally, it shows the estimated class share of these assets and eligibility rates within a market. Class share estimates primarily focus on a market's willingness to adopt and pay for GT technology. Eligibility has been held constant across all markets, driven by the assumption that label claims will be nearly identical.

Table 4

Type 3 Forecast Market Launch Sequence

Market	Launch Sequence	Class Share	Eligibility
Europe	*n+1*	40%	30%
APAC	*n+1*	2%	30%
LATAM	*n+2*	4%	30%
CEETRIS	*n+2*	4%	30%
ME	*n+2*	4%	30%
Africa	*n+3*	2%	30%
N America	*n*	50%	30%

3.3.2 Epidemiology Adjustments

When modeling an asset when the diseases state is uncertain, it makes it impossible to research the epidemiology. To give a better frame, global point prevalence is broken down into six categories shown in Table 5. This information derives from Orphanet (2021) classification of global point prevalence.

Table 5

Disease State Population Inputs by Orphanet Point Prevalence Classification

Point Prevalence	Prevalence Used	Incidence Used
<1 in 1,000,000	1 in 1,000,000	1 in 30,000,000
1-9 in 1,000,000	5 in 1,000,000	1 in 6,000,000
1-9 in 100,000	5 in 100,000	1 in 600,000
1-5 in 10,000	2.5 in 10,000	1 in 120,000
6-9 in 10,000	7.5 in 10,000	1 in 40,000
>1 in 1,000	1 in 1,000	1 in 30,000

These same inputs were used to enable inputs for a disease point prevalence. Turing to incidence rate, there is a more simplified approach to estimating patients. Type 3 forecasts derive prevalence by acknowledging most of the patients with these diseases do not have standard lifespans. Therefore, prevalence is found by dividing incidence by 30 and the incidence rate driver is population.

3.3.3 Adoption Curve and Uptake Estimate

Similar to Type 2 forecasts, Type 3 forecasts have a severity impact within our model. Unlike before, Type 3 forecast will use this variable to drive give different inputs to the mathematical formulation. As seen in Table 6 the severity impact ratings control the early adopted impact for the prevalence population. We will expect 100% of the prevalence population addressed in four years for "High" severity disease state products. We expect half of the prevalence population to rush for "Medium" disease states, with the remaining 50% evenly spread over the 10-year patent life. A "Low" severity input will have uniform prevalence adoption throughout the 10-year patent life. We will use this variable to also address uptake rates.

High, medium and low severity disease states will respectively have 75%, 50% and 25% incidence uptake rates. For the prevalence population uptake, the adoption will be 60%, 40%, and 20% for high, medium, and low severity disease states. We believe the assumed uptake rate by population type to be reasonable, considering our knowledge and research on the various disease severity levels. These assumptions are detailed in Table 6. Disease states that are likely to have a rush of prevalence population are also much more likely to have higher uptakes.

Table 6

Type 3 Forecast Disease Severity Mode Impact

Severity Impact Rating	High	Medium	Low
Early Adopter Population Impact	100%	50%	None
Prevalence Uptake	60%	40%	20%
Incidence Uptake	75%	50%	25%
Early Adopter Time	4	4	10
Adoption Curve	Normal	Normal	Uniform

3.3.4 Asset Forecasting

With the adjustments to the Type 3 model inputs defined, the model can use the same mathematical formulas used in Type 2. $prev(t)^{mrk}, inc(t)^{mrk}$ do not vary mathematically between the two models. The above assumptions and variables defined in 3.3 create all needed inputs to use the same formulas.

3.4 Adjustments for Type 1 Forecast

Roche has a team of market access, epidemiology, and forecasting experts. It would be naive for our model to make assumptions to build a forecast when it is more appropriate to include their team's forecast. Therefore, our model allows our partner to drop in the forecast from their internal team.

3.5 Combining Assets to Build a Portfolio

Let us now look at all these difference assets together. We will bring the assets together in two different ways. First, we will walk through an unadjusted volume, or the projected volume in all assets in the portfolio make it to market. Then we will consider the risk adjusted volume. The risk adjustment volume discounts the total volume by a probability of success, or better named the expected total volume. Table 7 show the definitions and use cases for each of these scenarios.

Table 7

Forecasted Demand Volumes and Uses

Demand Volumes	Forecast Output (# Treatments)	Business Application
Unadjusted	Potential total commercial volume	Market sizing supply chain needs for the asset if successfully commercialized
Risk Adjusted	Manufacturing volume based on an asset's current pipeline state Risk Adjusted = P(Success)* Unadjusted	Planning future supply chain capacity for only the volume currently likely to be commercialized – this creates an expected value of volume in a year

3.5.1 Total Volume Projection

Looking at the total volume for any year, the total volume can be calculated by summing all assets and markets from Equation 3.7, to create Equation 3.12.

$$vol(t) = \sum_{asset} \sum_{mrk}^{7} vol(t)^{asset,mrk}$$

(3.12)

3.5.2 Risk Adjusted Volume

Alternatively, to considering the total volume, the model can calculate the risk adjusted volume for a year. As noted, the risk adjustment considers that an asset in the pipeline is not certain to make it to market. Therefore, to estimate the value of ra^{asset}, Table 8 shows how we arrive at the various values.

Table 8

Risk Adjustment by Clinical State

Clinical Phase Input	ra^{asset} Value
PI	13.6%
PII	23.8%
PIII	66.3%

The model can use the values to arrive at the total risk adjusted volume for any year, as shown in Equation 3.13.

$$risk\_adj(t) = \sum_{asset} \sum_{mrk}^{7} ra^{asset} * vol(t)^{asset,mrk}$$

(3.13)

With a complete understanding of how this forecasting model works, let us move to potential applications of the model.

4 Results and Discussion

Our methodology represents a novel way to rapidly forecast rare diseases demand for GTs. For results, we will demonstrate and evaluate various use cases for forecasting these assets. The results section will walk through a model of a single assets, a known asset problem, and then a scenario planning problem. These three demonstrations will help show the usefulness of this forecasting model for our partner company.

4.1 Single Asset Forecast

Section 4.1 will focus on results from a single asset forecast for both Type 2 and Type 3 assets. Recall from 3.1.1 that Type 2 are moderate knowledge forecasts where a significant number of inputs are required, and Type 3 require significantly fewer model inputs. Type 1 assets will not be covered, as they are already forecasted by a team at Roche and simply used to enable a portfolio level view within the model.

4.1.1 Type 2 Asset

Let us start results by evaluating the output of a single Type 2 forecast. We created inputs for a normal adoption curve, birth-driven Phase III asset. Table 9 also details the other model inputs for an example Type 2 asset with anticipated ownership of all regional markets.

Table 9

Single Asset Model Inputs by Market

	Europe	APAC	LATAM	CEETRIS	ME	Africa	NA
Ownership	Y	Y	Y	Y	Y	Y	Y
Launch Year	2025	2025	2026	2026	2026	2027	2024
Prevalence	.0028%	.0028%	.0028%	.0028%	.0028%	.0028%	.0028%
Incidence	.021%	.021%	.021%	.021%	.021%	.021%	.021%
Eligibility	30%	2%	10%	4%	4%	2%	50%
Class Share	40%	40%	40%	40%	40%	40%	40%
Prev. Mrk. Share	60%	60%	60%	60%	60%	60%	60%
Inc. Mrk. Share	75%	75%	75%	75%	75%	75%	75%
Early Adopter Time	4	4	4	4	4	4	4

With all inputs defined, Figure 4.1 shows our results for the example asset by year and market level, both in unadjusted and risk adjusted volumes. Additionally, the rightmost column the figure allows us to see the total number of patients that this asset will serve over a 15-year horizon.

Figure 4.1

Results Table Output for a Single Asset

		2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	TOTAL
Unadjusted Volume	Europe	–	566	955	955	566	498	497	497	497	497	497	–	–	–	–	6,025
	APAC	–	140	221	221	140	126	126	126	126	126	126	–	–	–	–	1,477
	Latam	–	–	306	441	441	306	283	282	282	282	282	282	–	–	–	3,189
	CEETRIS	–	–	442	614	614	442	412	411	411	411	411	411	–	–	–	4,578
	ME	–	–	40	53	53	40	38	38	38	38	38	38	–	–	–	413
	Africa	–	–	–	145	179	179	145	139	139	139	139	139	139	–	–	1,480
	N America	739	1,108	1,108	739	674	673	673	673	673	673	–	–	–	–	–	7,732
	TOTAL	739	1,814	3,073	3,168	2,667	2,264	2,172	2,165	2,165	2,165	1,492	870	139	–	–	24,894
Prob. Adj. Volume	Europe	–	375	633	633	375	330	329	329	329	329	329	–	–	–	–	3,994
	APAC	–	93	147	147	93	84	83	83	83	83	83	–	–	–	–	979
	Latam	–	–	203	292	292	203	188	187	187	187	187	187	–	–	–	2,115
	CEETRIS	–	–	293	407	407	293	273	272	272	272	272	272	–	–	–	3,035
	ME	–	–	27	35	35	27	25	25	25	25	25	25	–	–	–	274
	Africa	–	–	–	96	118	118	96	92	92	92	92	92	92	–	–	981
	N America	490	735	735	490	447	446	446	446	446	446	–	–	–	–	–	5,126
	TOTAL	490	1,203	2,037	2,100	1,768	1,501	1,440	1,436	1,436	1,436	990	577	92	–	–	16,505

One key observation from reviewing Figure 4.1 is the lack of regional volume minimums. The model is concerned with the ability to commercialize in a region, not the desire. The results for the Middle East (ME), where the maximum commercial demand population in a year is 53 patients, warrants discussion if a company would view the volume as an attractive enough market to pursue regional regulatory approval. Another interesting point to review when forecasting at this level is regional mixes. Currently, this current forecast has 55% of the volume coming from Europe and North America. This might be a little light considering the historic breakdown of regional volumes and should be further evaluated. This demonstrates how the inputs required for Type 2 requires a few iterations.

4.1.2 Type 3 Asset

Evaluating the different potential results for Type 3 forecast is very important to understanding future application of the model. Rather than reflecting on a single example result, we present all potential results of the Type 3 forecasts. Shown in Table 10, you can see the output for all 18 combinations of prevalence and severity, assuming an asset is commercialized globally.

Table 10

Total Volume Outputs from Type 3 Forecasts

Prevelance	Serverity	Years Post Launch															Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
<1 / 1 000 000	High	4	22	41	37	19	8	5	5	5	5	3	1	0	-	-	153
	Medium	3	11	19	17	11	8	7	7	7	7	4	2	0	-	-	102
	Low	2	4	5	5	5	5	5	5	5	5	3	1	0	0	-	49
1-9 / 1 000 000	High	19	110	206	185	95	38	24	23	23	23	15	6	1	-	-	765
	Medium	15	56	93	87	57	38	33	33	33	33	22	8	1	-	-	510
	Low	8	21	24	24	24	24	24	24	24	24	16	4	1	0	-	243
1-9 / 100 000	High	192	1,097	2,062	1,846	951	375	237	225	225	225	150	55	7	-	-	7,649
	Medium	149	558	935	870	572	380	334	330	330	330	221	81	10	-	-	5,102
	Low	85	208	235	243	243	243	243	243	243	243	158	35	8	0	-	2,429
1-5 / 10 000	High	962	5,486	10,308	9,229	4,755	1,876	1,187	1,127	1,126	1,126	752	276	34	-	-	38,245
	Medium	744	2,792	4,673	4,352	2,861	1,901	1,672	1,652	1,651	1,651	1,103	405	50	-	-	25,509
	Low	424	1,039	1,176	1,214	1,214	1,214	1,214	1,214	1,214	1,214	791	176	39	0	-	12,144
6-9 / 10 000	High	2,885	16,459	30,923	27,688	14,266	5,628	3,562	3,381	3,378	3,378	2,256	828	102	-	-	114,734
	Medium	2,233	8,376	14,020	13,057	8,583	5,704	5,015	4,955	4,954	4,954	3,308	1,215	150	-	-	76,526
	Low	1,272	3,116	3,527	3,643	3,643	3,643	3,643	3,643	3,643	3,643	2,372	527	116	0	-	36,433
>1 / 1000	High	3,847	21,945	41,231	36,918	19,021	7,504	4,749	4,508	4,504	4,504	3,007	1,104	136	-	-	152,979
	Medium	2,978	11,168	18,693	17,410	11,445	7,606	6,687	6,607	6,605	6,605	4,411	1,619	200	-	-	102,035
	Low	1,696	4,155	4,703	4,858	4,858	4,858	4,858	4,858	4,858	4,858	3,162	703	155	0	-	48,577

Evaluating the results, a key insight is the volume breath of a patient population. The difference is four orders of magnitude from the highest to the lowest volume forecasts. This is an incredible variation in the number of patients served for a single disease. Additionally, Table 10 shows how the severity input impacts the early adopter population. We can see in the table that high severity peaks three to four years postlaunch with a significant drop year-over-year postpeak. Alternatively, the forecasts for medium severity target disease states do not have nearly as significant of a drop postpeak. The low severity forecasts hit peak volume on year four, postlaunch, but then maintains this volume through year 10. Figure 4.2 compares the projected commercial patient volumes between the three severity factors. Our plotted outputs show the percent of total demand by severity input.

Figure 4.2

Percent of Total Volume over Time by Severity
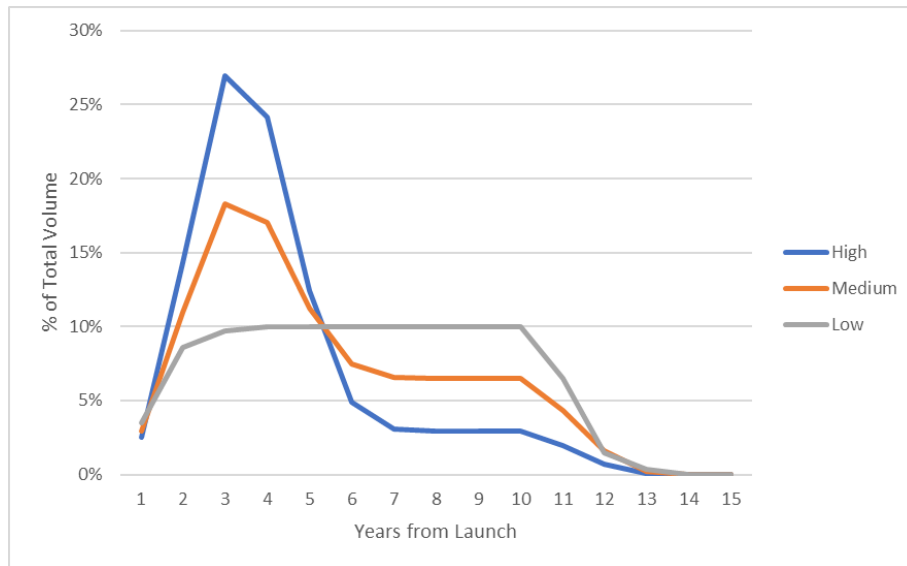
When considering network design, it is incredibly helpful to understand the nature of the demand curve over time. Effectively, these three underlying adoption curves will be scaled based on the prevalence selected. These different curves will have a significant impact on needed capacity and flexibility of capacity. Companies will need to employ unique strategies to serve these various distributions.

4.2 Modeling Known Portfolio of Assets

Now that we have evaluated the results from a single asset, we can consider more complicated portfolio level evaluations. One of the key use cases for the model will be to build a portfolio of the current pipeline assets for our partner company. For this evaluation, we have created a model looking at five known assets. We will walk through how we designed these scenarios, the results, and a discussion around the results.

4.2.1 Portfolio Design

To build a portfolio, we must forecast various individual assets. For the portfolio design results we have created five unique assets using public data to project forecasts for their diseases. These diseases are similar but do not mirror out partner's pipeline, here we will call them Disease A, Disease B, Disease C, Disease D, and Disease E. Figure 4.3 displays the input dashboard for the portfolio we will evaluate.

Figure 4.3

Type 2 Asset Forecasts for Portfolio Model

| FCST Type	Asset	Pipeline State	First Launch
Type 2	Disease A	PIII	2024
Type 2	Disease B	PII	2025
Type 2	Disease C	PI	2027
Type 2	Disease D	PI	2026
Type 2	Disease E	PI	2028

Pausing here for a second, let us elaborate on the use case of a model like this. Roche's currently has a few GT assets within its pipeline, and one commercialized. As with any company developing a new technology, Roche needs to make decisions years in advance how the estimated capacity need for its manufacturing, packaging, and distribution networks for these products. Modeling the current known state and help Roche understand the volume commitments based on their portfolio. We can further include an array of Type 3 assets to help fill in gaps for preclinical assets. Alternatively, unless assets are added via M&A, this result would help our partner project volume needs over the next eight to10years.

4.2.2 Results and Discussion

After all assets are forecasted, the model compiled the forecasts to help allow analysis in a few different ways. The compilation of these volumes might be useful at a regional level or by asset. Additionally, there are different use cases for wanting to understand the total portfolio volume forecast or the risk adjusted forecast. Figures 4.4 and 4.5 below show some example outputs from portfolio planning.
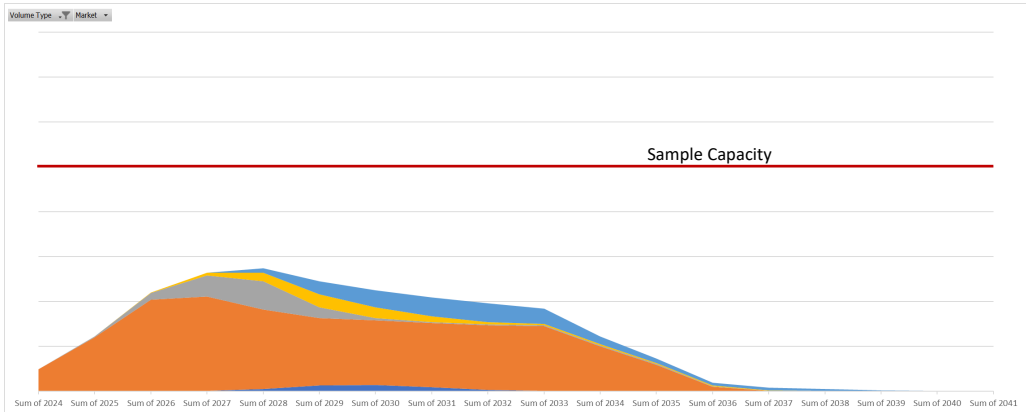
Figure 4.4

Unadjusted Portfolio Volume by Market Region
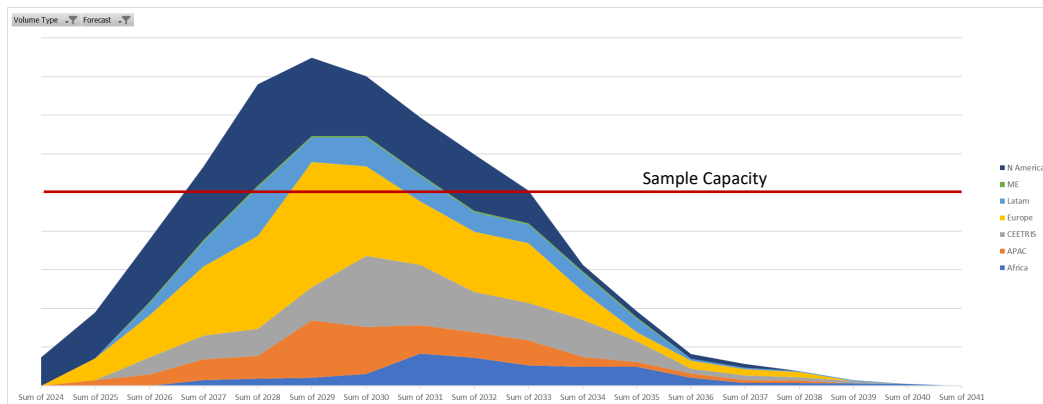
In Figure 4.4 the unadjusted volume by region is shown. Our partner can use these regional outputs to help set direction on the final labeling or distribution capacity for a certain geography. Additionally, Figure 4.4 is the unadjusted volume, indicating this is the point forecast for this portfolio assuming all assets are successfully commercialized. Considering implications of this, a sample capacity line has been placed on Figures 4.4 at fixed capacity line severing a certain number of patients. If capacity is fixed, our partner will have to figure how to address excess demand from 2027-2032, and how to fill the idle capacity before and afterwards.

Figure 4.5

Risk Adjusted Portfolio Level Volume by Asset

In Figure 4.5. we see the same portfolio, but the risk adjusted volume projected identified by asset. A major takeaway here is recognizing how sensitive the risk adjustment volume is to clinical state. Considering the capacity limitation introduced with Figure 4.4, Figure 4.5 shows that when these current assets are adjusted for the probability to make it to market, utilization never exceeds 60%. This highlights that use case for both risk adjusted and unadjusted volume. Risk adjusted volume shows you a probabilistic output, but biopharmaceutical companies cannot be caught short capacity. This helps us understand the range from expected to high for the short-term capacity planning.

Additionally, Figure 4.5 shows the impact that advancing through clinical states can have. Shown in Table 11, Disease B and Disease E have relatively similar total patient volume projections (24,894 to 17,756), but when adjusted you see Disease B is significantly more than Disease E (16,505 to 2,415). That is due to Phase III vs. Phase I clinical states create inputs significantly impacted the risk adjusted volume outcomes.

Table 11

Unadjusted and Risk Adjusted Volume for Portfolio

Year	Risk Adjusted Volume					Unadjusted Volume				
	Disease A	Disease B	Disease C	Disease D	Disease E	Disease A	Disease B	Disease C	Disease D	Disease E
2024	-	490	-	-	-	-	739	-	-	-
2025	-	1,203	19	-	-	-	1,814	79	-	-
2026	-	2,037	152	10	-	-	3,073	638	72	-
2027	7	2,100	470	65	-	51	3,168	1,977	481	-
2028	50	1,768	629	190	98	370	2,667	2,643	1,400	721
2029	126	1,501	244	291	291	925	2,264	1,027	2,137	2,136
2030	143	1,440	47	244	379	1,053	2,172	199	1,793	2,786
2031	86	1,436	23	126	424	631	2,165	99	928	3,115
2032	33	1,436	23	50	424	244	2,165	95	365	3,115
2033	14	1,436	23	27	340	100	2,165	95	196	2,499
2034	11	990	23	24	174	83	1,492	95	176	1,282
2035	11	577	18	24	98	82	870	75	176	717
2036	11	92	5	19	59	82	139	22	143	436
2037	9	-	-	9	59	67	-	-	65	436
2038	4	-	-	2	45	30	-	-	15	331
2039	1	-	-	-	18	7	-	-	-	133
2040	-	-	-	-	6	-	-	-	-	48
2041	-	-	-	-	-	-	-	-	-	-
Total	507	16,505	1,676	1,081	2,415	3,725	24,894	7,043	7,947	17,756

This aligns to the design, and logic, that the expected volume is higher as an asset progress through the clinical stages but calls for caution as out probability of success could be under or overstated to Roche's expectations. Also, the results of modeling the current portfolio show a significant volume drop in mid 2030s. That aligns to a usage of the portfolio model, allowing our partner to project needs from first commercialization for the current known assets. Here we show how we have used our rough forecasting methodology to enable our partner to make more robust capacity planning decisions. Future use of the model to include various Type 3 assets if desired to show volume needs beyond 2030s, or inclusion of earlier Type 3 via M&A. Due to the model complexity of considering what this might look like, instead of using expected volumes within the current portfolio model, we built scenarios to consider longer term planning.

4.3 Portfolio Scenarios

An alternative way to use the model is to consider the long-term projection of how different expected launch scenarios will impact total need over time. This adaptable and interactive model enables users to

enter or adjust various inputs for the evolving gene therapy pipeline. Let us consider a real-life application of the model to the current Roche portfolio – Roche has five different Type 2 assets, but this likely just a starting point to their future in GTs. A key point to this problem is trying to understand how both the current known assets and the unknown assets will combine to arrive at total commercial demand over a long-range time horizon. We used the forecasting model to create 96 different scenarios that illustrate potential commercial demand from randomly generated combinations of currently known and future unknown assets. The design and results of our scenarios are detailed below for evaluating our model.

4.3.1 Scenario Design

To get a better sense of the use of the model use case, we created a wide variety of scenarios attempting to see how the model is useful in supporting commercial volume decisions. We utilized the model to project the volume of the five known assets, along with 55 potential Type 3 assets, for a total of sixty potential assets launched over a 20-year time horizon.

To assess these different situations, we first considered how the current five Type 2 assets might be successful. There are 32 combinations (2^5) of which Type 2 assets might make it to market creating unique total volume scenarios. Like the questions for actualizing current Type 2 assets, we do not know what potential Type 3 assets will be successful and at what projected market demand. Our partner company did not share or have a forecast the number of expected gene therapies they expect to be approved. To consider the extremes that may exist, we developed a three-tiered scenario approach for the number of expected launches per year. The scenario groups are shown in Table 12 and described further below.

Table 12

Definition of 3 Scenario Groupings

Throughput Scenario	A	B	C
Asset Launch Target	3 launches / yr.	1 launch / yr.	1 launch / 3 yrs.
E(Type 2 Launches)	2.5	2.5	2.5
E(Type 3 Launches)	55	17.5	4.17
E(Total Launches)	57.5	20	6.67

At one extreme is a high-volume approach where Roche is launching approximately three assets per year from 2025-2044. At the other extreme, we project volume if Roche is launching only one asset every three years. We also consider a more middle ground: one asset per year. This helps decision makers determine the robustness of their manufacturing strategy.

Finally, we paired the 32 different potential outcomes from the Type 2 forecasts with the three larger buckets of total volume describe in Table 12. This resulted in a total of 96 difference scenarios: 32 Type 2 forecasts paired with Scenario A, 32 Type 2 forecasts paired with Scenario B, and 32 Type 2 forecasts paired with Scenario C. To bring these volume scenarios to fruition, we must define the inputs to the 55 Type 3 forecasts.

4.3.2 Estimating Type 3 Inputs

Considering that the pharmaceutical pipelines are dynamic, nearly all assets of the clinical pipelines will be comprised of Type 3 assets over next 20 years. 55 potential Type 3 assets were randomly selected to illustrate the range of extreme asset throughput rate scenarios based on the estimated probability distribution of a disease's known prevalence class and severity.

In our first estimate of these inputs, we utilized Orphanet (2021) epidemiology and patient age data. For prevalence, we used Orphanet's epidemiology database. Our analysis found worldwide point prevalence for 3,378 diseases. To estimate severity, Orphanet's patient age data set was used. 2,851 had information populated to enable us to classified diseases into traditionally child onset or traditionally terminal. Details of this classification and analysis can be found in Appendix A. We classified diseases as both child onset and terminal as high severity, if a disease was either child onset or terminal as medium, and if it was neither as low. Lastly, for mathematical simplicity, we assumed that onset age, terminal, and prevalence are all independent variables. Combining these aspects together we arrived at column WW Unadjusted Table 13. Upon review of the initial data, three items warranted further investigation. When evaluating worldwide point prevalence, we found a significantly different distribution for diseases with US and EU prevalence noted. Diseases with confirmed point prevalence for the US and/or Europe were not as skewed towards the lowest prevenances. Additionally, we considered that companies might have a selection bias to pursue more prevalent diseases. We found a sampling of clinical assets reported by Curran (2021) (noted in Appendix A) and classified them into point prevalence categories. Lastly, as noted in section 4.1.2, a few of these assets have volume too small to be economical to pursue. As a

result, we adjusted the Type 3 asset probabilities to 0% for four disease states. The excluded disease state classes are those in the two lowest prevalence classes (1-9 / 1,000,000 and <1 / 1,000,000) when paired with a medium or low severity factor. A summary of this analysis can be found in Table 13.

Table 13

Scaling Type 3 Asset Probability by Point Prevalence and Severity Factor

Severity Factor	Known Point Prevalence	WW Unscaled	US+EU Unscaled	Commercial Unscaled	Included?
High	>1 / 1,000	0.0%	0.0%	0.0%	Y
High	6-9 / 10,000	0.0%	0.1%	0.0%	Y
High	1-5 / 10,000	0.3%	5.3%	3.1%	Y
High	1-9 / 100,000	0.6%	8.7%	12.3%	Y
High	1-9 / 1,000,000	0.3%	5.1%	5.4%	Y
High	<1 / 1,000,000	23.3%	5.2%	3.8%	Y
Medium	>1 / 1,000	0.0%	0.1%	0.0%	Y
Medium	6-9 / 10,000	0.0%	0.3%	0.0%	Y
Medium	1-5 / 10,000	0.6%	12.9%	7.5%	Y
Medium	1-9 / 100,000	1.6%	21.3%	29.9%	Y
Medium	1-9 / 1,000,000	0.7%	12.6%	13.1%	N
Medium	<1 / 1,000,000	57.0%	12.7%	9.4%	N
Low	>1 / 1,000	0.0%	0.0%	0.0%	Y
Low	6-9 / 10,000	0.0%	0.1%	0.0%	Y
Low	1-5 / 10,000	0.2%	3.4%	2.0%	Y
Low	1-9 / 100,000	0.4%	5.6%	7.8%	Y
Low	1-9 / 1,000,000	0.2%	3.3%	3.4%	N
Low	<1 / 1,000,000	14.8%	3.3%	2.4%	N

Once the is scaled due to removing the excluded disease sates as noted above, the final distributions are listed in Table 14.

Table 14

Evaluation of Type 3 Asset Distribution

Severity Factor	Known Point Prevalence	WW Scaled	US+EU Scaled	Commercial Scaled
High	>1 / 1,000	0.0%	0.1%	0.0%
High	6-9 / 10,000	0.0%	0.2%	0.0%
High	1-5 / 10,000	1.0%	7.8%	4.3%
High	1-9 / 100,000	2.3%	12.8%	17.1%
High	1-9 / 1,000,000	1.0%	7.5%	7.5%
High	<1 / 1,000,000	85.4%	7.6%	5.3%
Medium	>1 / 1,000	0.0%	0.2%	0.0%
Medium	6-9 / 10,000	0.1%	0.4%	0.0%
Medium	1-5 / 10,000	2.3%	19.0%	10.4%
Medium	1-9 / 100,000	5.7%	31.3%	41.8%
Medium	1-9 / 1,000,000	0.0%	0.0%	0.0%
Medium	<1 / 1,000,000	0.0%	0.0%	0.0%
Low	>1 / 1,000	0.0%	0.0%	0.0%
Low	6-9 / 10,000	0.0%	0.1%	0.0%
Low	1-5 / 10,000	0.6%	4.9%	2.7%
Low	1-9 / 100,000	1.5%	8.2%	10.9%
Low	1-9 / 1,000,000	0.0%	0.0%	0.0%
Low	<1 / 1,000,000	0.0%	0.0%	0.0%

Reviewing the table, it is worth noting that there is an 85% that a disease picked from the WW Scaled distribution has a severity of high and a prevalence of less than 1 in 1,000,000. Rather than trying to reconcile abnormalities and combine these distributions, we will use all three as potential distributions of diseases for our scenarios and evaluate any key differences.

4.3.3 Scenario Planning Results

For results, we will evalute the output from the 96 scenarios itterated 175 times. This rendered 5,600 data points for each of 9 result situations (3 asset distributions x 3 throughput scenarios). We have found that both of these factors signifcantly contribute to the total number of patients served. Displayed in Table 15, we looked at the total number of patients served over the 20 year horizon first. We evaluated the data ranges for each quartile of the 9 results drive. The results show that launch scenario A (3 assets/year) drives the most commercial volume than its corresponding quartiles, whereas B (1 asset/year) enables the second highest, and C (1 asset/3 years) has the fewest. Additionally, Table 15 shows that the US/EU Scaled adoption curve nearly always has the higest volume, Corporate Scaled distribution as the second highest, and the Worldwide Scaled coming having the least volume.

Table 15

Type 3 Results by Total Patients Served (Volume as Multiples of Lowest Value)

Percentile	WW Scaled			US/EU Scaled			Corporate Scaled		
	A	B	C	A	B	C	A	B	C
0%	16	3	1	365	77	14	258	65	16
25%	91	40	26	614	195	70	431	143	56
50%	120	56	38	686	233	91	481	169	73
75%	155	76	51	764	276	115	531	199	93
100%	436	215	162	1,122	601	331	777	354	198

Additional apparent in the data the percent difference in launch rate lessens in each scenario as you move up quartiles. The percentage difference between A, B, and C for each distribution curve are larger in the lower percentiles (0% and 25%) than those in the higher percentiles. Figure 4.6 further illustrates how the distribution curves affect the level of seperability, or differences, between the expected launch rates.

Figure 4.6

Total Patients Served by Launch Rate and Adoption Curve Boxplots
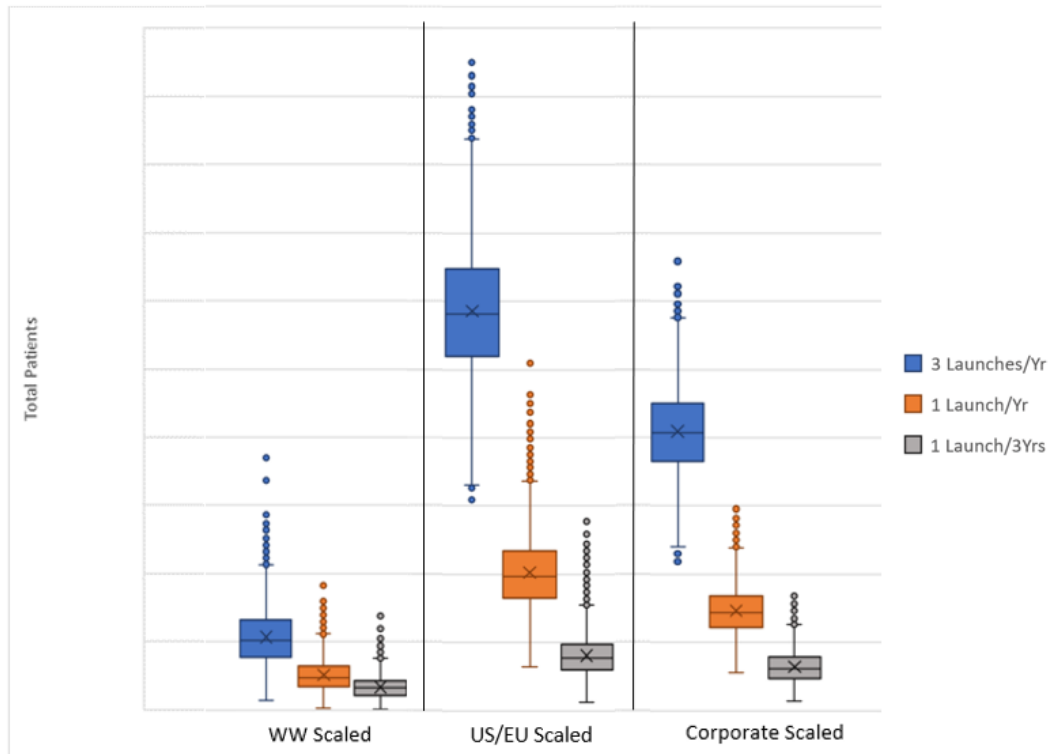
When looking at the worldwide scaled distribution of launch scenarios, there is signifcant overlap between the three launch rates. The US/EU scaled distribtuion have ranges that appear more distinct between the launch rates. Looking at the differerenece between the data, we can see that the WW scaled results is significantly different than the results from the other two options. The launch rate B median for WW scales is less than a third of the US/EU scaled curve B median value. When consdering how the launch rate interacts with the different adoption curves, the results show that as you move up the increasing adoption curves, the impact of launch rate increases.

Where understanding the total patients served is essential, we additionaly evaluated the maximum patients served in a single year in Table 16. Although we shift the perspective from Table 15's cumulative sum of all addressable patients to the maximum servable patients in each scenario here, the same trends persist from the above. Volume increases as you move up the number as asset launch options (A, B, C), but at a smaller magnitude than the increase of the number of assets launched. WW Scaled again is significantly different than the other two results.

Table 16

Type 3 Results by Maximum Patients Served in a Year (Data is Multiples of Lowest Value)

	Max Patients Served								
	WW Scaled			US/EU Scaled			Corporate Scaled		
Percentile	A	B	C	A	B	C	A	B	C
0%	20	3	1	388	94	22	272	78	26
25%	139	82	64	746	276	110	521	182	93
50%	208	105	87	848	334	150	592	252	113
75%	273	140	106	988	407	223	669	300	165
100%	637	439	370	2,565	1,166	980	1,089	565	430

Diving into a median value from Table 16, Figure 4.7 shows the median values for the US/EU Scales asset target with B launch rate (334x). What the figure shows is that there are 3 adoption curves that all had the median value and their volume over time. Even though all three of these adoption curves have the same peak, the occur at different times and rather different overall adoption curves.

Figure 4.7

Portfolio Need of Peak Demand US/EU Scaled B Median Values
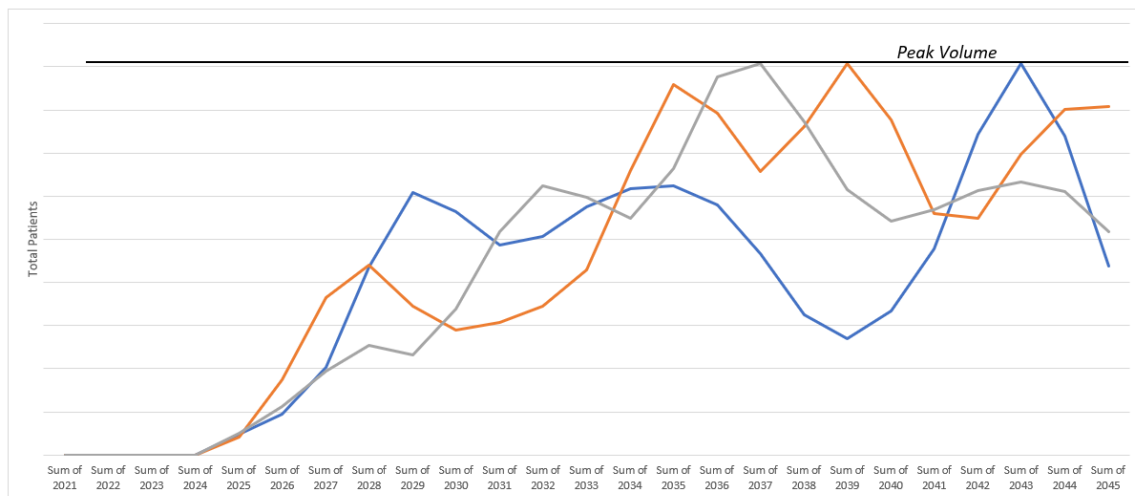
The results of the differences in asset targets, launch rates, and forecast of the median values all have significant impact on the design of this network.

4.3.4 Scenario Planning Discussion

Reviewing the results from the different outcomes of the scenario planning exercise, a few things become very interesting. First, let us consider the difference between the various adoption curves. The worldwide adoption curve is significantly lower than the other two adoption curves. The values for launch rate B are on average 610% lower than the highest value adoption curve, US/EU. Alternatively, corporate distribution is only 36% lower than the US/EU curve on average. This suggests that companies might be targeting rare disease that match the type of disease states that move heavily affect the US and EU, versus the rest of the world. Given the disproportionate amount of revenue that the US/EU represent for pharmaceutical companies, this conclusion is unsurprising. As mentioned in the results, fine tuning the distribution curve between US/EU and Corporate does not lead to much of a difference. For the best practical application here, we would need to understand the expected launch rate throughputs, as they have a larger impact on the US/EU and Corporate distribution curves.

The same trends persist for the maximum patients served that were true for total patients served. As shown in Figure 4.7, the interesting feature of the scenario planning is the various demand curves for similar peak values. This result drives home the need that in order to be able to meet the demand for these products at a portfolio level, the network must remain flexible. Even if we are able to accurately project what the various disease state Roche with target and how successful clinical trials will be – demand could vary widely. Figure 4.7 shows that in 2039, when one of the median results hits the peak, another scenario has volume below 33% of peak patients served. Success in the GT space will depend on flexibility in network design.

4.4 Managerial Implications

Recalling Roche's motivation (Section 1.2) to build one of the first gene therapy global supply chain networks, this capstone helps Roche stakeholders make strategic business decisions for their new infrastructure. Existing supply chain design exists from a recent acquisition (F. Hoffmann-La Roche Ltd, 2019). However, Roche is actively trying to expand their commercial gene therapy portfolio with various assets (Types 1 and 2) currently moving through the clinical pipeline. In addition to the risk that these known pipeline assets are not successfully commercialized, unknown future assets (Type 3) also need to be considered to serve the broader scope of long-term demand. Our point model and scenario planning tools provide current-state insights, as well as adaptable frameworks, for the regional commercial demand. Roche has already begun utilizing the work from this capstone to make decisions around development of the following primary supply chain infrastructure areas illustrated in Figure 4.8:

1. Global viral vector (VV) manufacturing
2. Regional drug product (DP) production and ultra-cold storage
3. Market level patient packing, labeling, and distribution

Figure 4.8

Key Gene Therapy Portfolio Network Design Decisions Scope
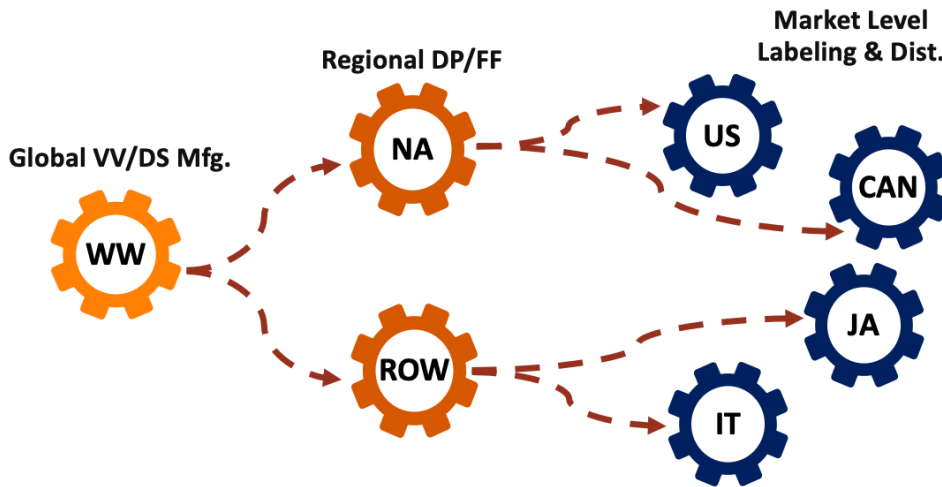
The Type 1 and 2 assets are closest to commercialization, and therefore, are essential for making immediate network decisions. With asset launches expected in early 2020s, it is most cost-effective, convenient, and fastest to leverage any open capacity of the current supply chain network. Roche plans to continue manufacturing drug substance and products in the US facility as known assets are approved to grow the commercial gene therapy portfolio. The current downstream network will also serve the portfolio's global affiliates (patient treatment centers, such as hospitals). Knowing that the breadth and demand volume of gene therapies are increasing, the existing network constraints create two key questions: (1) Which current supply chain operations need to be expanded for serving new commercial demand? (2) At what point will Roche's commercial gene therapy portfolio demand exceed the maximum supply chain capabilities? Roche's long-term strategy is dependent on these misleadingly straightforward questions.

All of the primary network elements listed above are directly impacted by the commercial portfolio demand but are constrained to finite operational capacities. The point model provides a comprehensive and adaptable view of both these supply chain factors. First, the unadjusted volume forecasts commercial demand by region to understand overall annual needs where Roche has market ownership. More specifically for estimating the existing network utilization, our risk adjusted volume indicates what supply chain capacity we believe Roche should plan to allocate for the current clinical stage asset. The US manufacturing facility (primary infrastructure area #1) is expected to be most sensitive to the capacity constraints. To determine the long-term manufacturing levels and capacity inflection point, we value each asset's capacity as the portion of its forecasted commercial volume likely to go to market today (probability of commercial success described in Section 3.5.2). While the risk adjusted Type 1 and 2 asset volumes represent a large amount of Roche's supply chain capacity needs over the next five years, most of the long-term gene therapy portfolio will consist of future unknown Type 3 assets.

Roche stakeholders can use the user-friendly model to decide which Type 3 forecasts they want to consider as part of their current portfolio view. The majority of these forecasts are higher-risk Preclinical or Phase I assets that are essential to realize risk pooling value across the product portfolio. Beyond selecting the number of Type 3 assets and market launch years of interest, the color-coded user inputs enable Type 3 asset variables to be specified for any internal R&D or external M&A efforts. This allows decisionmakers to account for prospective unknown types of assets beyond those early in the development and approval lifecycle. From an asset planning perspective, the point model Type 3 asset forecasts and target throughput scenarios are particularly helpful.

Similar to the objective of consumer products in other industries, the overarching gene therapy portfolio strategy effectively outlines Roche's product line commercialization targets. The three-target throughput scenario planning groups (A, B, C) more clearly illustrate implications for overarching strategy decisions. Higher-level stakeholders can compare the simulated commercial volumes between the target product launch extremes detailed in Section 4.3.1. Additionally, the strategy for selecting future the Type 3 assets, which mainly differentiate the scenarios, can be aligned to the disease state prevalence and severity classifications of the combined US and EU driven market, worldwide epidemiology, or current commercial pipeline pursuits. Type 3 assets will surely cause the global portfolio demand to outgrow the current supply chain infrastructure and partner capabilities.

Roche faces the classic make versus buy supply chain problem for developing the commercial gene therapy portfolio supply chain network. Market and quality regulations drive some of the requirements for new manufacturing, distribution, or local storage facilities. We have been able to see the practical application of this capstone as Roche evaluates their pressing viral vector manufacturing decisions.

4.5 Model Limitations and Improvements

Considering potential improvement and limits to our work, we will separate comments into the point model and the scenario planning. Before exploring them, there a few overarching limitations. First, to maximize user access and compatibility for the future application, our functioning model and scenarios were built in Excel. The capstone deliverable tools are inherently limited by Excel's processing constraints. Also, given the quantity of inputs, limited time was spent working on data validation. The entire model could be improved by improving the user interface and utilization of defensive programing techniques.

4.5.1 Point Model Methodology

When evaluating limitation and potential improvements to the point model forecast, there are three main items that appear as potential shortcomings: clinical probability, commercialization desire, and Type 3 assumptions.

In order to reach a risk adjusted volume, we utilized a research evaluating historical clinical trial success for non-oncologic orphan drugs. This is an overarching set of the diseases that this tool will model, but there are many items other items that impact the probability of commercialization that we did not allow

variation. Complexity of primary endpoints, current standard of care, modality, and company experience all play a role in this probability that were not considered. There is a potentially justification that probability of success should have been an input and not derived from historical performance.

As noted in 4.1.1, another shortcoming is the realization that not every asset is commercialized in every country. Our model assumes that as long our partner has rights to commercialize in a market, they will. This is not accurate. Companies are expected to have minimum volume thresholds to warrant commercialization.

There is an array of shortcomings built into the Type 3 forecasts. Below are three key areas identified as improvement areas or current limitation:

Prevalence Range Mean: The model uses the mean of the prevalence range, or the exact value for the endpoints of the ranges (i.e., >1 in 10,000 and <1 in 1,000,000). There is a significant opportunity to increase the complexity of forecast by identifying the distribution of diseases within the range and selecting from that distribution for forecasting.

Market Access Limits: Impacts of substantial biopharmaceutical commercial product pricing is not thoroughly addressed in our model. Reimbursement is a large factor for pharmaceutical treatments. Especially given the significantly high cost of gene therapies, starting at $425,000 per single eye treatment in the USA (Jackson et al., 2020), the viable payment models will likely have a large impact on the projected patient adoption. The lack of approved gene therapy products, and corresponding historical data, to properly estimate reimbursement for any given asset within each regional market. Our model makes some small adjustments at the class share level for the model's regions, but limited due diligence was completed to arrive at these numbers.

Prevalence to Incidence Lifespan: We assume a thirty-year lifespan for all diseases with connecting prevalence to incidence. There is an opportunity to connect the thirty-year assumption to severity.

Where improvement to any, or all, of these limitations provide opportunity to enhance the model we feel that the justification and assumptions are current appropriate.

4.5.2 Scenario Planning Design

There are two sections of limitations and improvements that need to be addressed for the scenario planning portion of the project: Type 3 inputs and model design.

When considering the Type 3 inputs, we heavily relied on Orphanet (2021) for disease state epidemiological and patient data. Utilizing a single data source is inherently a limitation. Any data quality issues, including missing/unknown, inaccurate, or inconsistent data, from the Orphanet database were carried over to our research. For instance, not all rare diseases identified by Orphanet have known point prevalence. Aside from data sources, the disease states themselves pose limitations to our research. Not all disease states are potentially treatable using the gene therapy technology currently available. Today's GTs are primarily limited to monogenic diseases (Kirschner & Cathomen, 2020). Additionally, the model does not account for potential future multi-disease use cases, where an asset may be able to treat more than one target disease state. Lastly on the Type 3 inputs, in order to use the

data above we heavily relied on assuming independence. In order to calculate our forecast types, our inputs assumed that if a condition is terminal, average age of onset, and point prevalence are not dependent on one another. This is not true. Easiest to see if that if a disease is terminal and child onset, that significantly impacts that global point prevalence of the disease. We view this as an opportunity point for improvement with more robust epidemiological research.

The major limitation to the scenario planning model and work is excel design. Putting aside the current require processing time of to run the model, row limits forced all scenario planning to remove the ability to isolate market demand. In the current design, there is no way to identify the total volume for a single market. There exists a large opportunity to move this work into a simulation software for processing. Future work in this space also exists to identify the most likely adoption curve. This might be by using statistics software to evaluate simulation outputs and identify the most likely annual volume.

5 Conclusion

Pharmaceutical companies are facing new challenges when trying to commercialize novel gene therapies. Each asset is unique, from molecule research and development to the one-time patient administration. Not only are such companies assuming great risk to grow their portfolios, but they are shaping an emerging biotechnology market. Three approved gene therapies provide little historical data, which is often the backbone of new product forecasting. These features of the gene therapy market create a complex challenge when considering demand. This capstone project supports both immediate and long-term global gene therapy supply chain network development decisions with several adaptable tools.

First, we created foundational asset level forecasts of the commercial demand and supply chain capacity volumes for a given gene therapy. These assets are segmented based on the existing available rare disease state knowledge to narrow and enable customization of inputs. The disease-specific forecasts of current mid-to-late-stage assets can be adjusted as more epidemiology or market data become available. The underling framework itself can be applied to forecast future secured assets with confirmed disease states. Our model also serves the need to forecast unknown potential assets, which comprises either preclinical assets or future acquisitions.

The known and unknown asset forecasts include both the total commercial and risk adjusted demand volumes based on an asset's current likelihood of success according to its current clinical stage. Individual asset forecasts are then combined in a portfolio level model. Users are able to choose assets and customize select variables to provide a dashboard of aggregate insights over a 20-year time horizon. The comprehensive supply chain capacity volume forecast allows Roche to project drug substance, drug product, packaging, and distribution across their evolving portfolio for anticipated regional markets.

Beyond the prospective current portfolio addressed in our point model, scenarios support long-term decisions. Three throughput scenarios of three assets per year, one asset per year, and one asset per three years practically apply the model to illustrate the range of commercial demand for target product launch extremes. A simulation was used to randomly generate 96 scenario outcomes across a 20-year time horizon. With only five Type 2 assets currently in the clinical pipeline, future assets targeting unknown disease states will comprise most of Roche's commercial gene therapy portfolio. We analyzed the known epidemiological and average patient age data to create a probability distribution for potentially treatable rare diseases according to the primary US and EU markets point prevalence, worldwide point prevalence, and ongoing pharmaceutical commercial portfolio assets. The variety of scenarios will help higher level Roche stakeholders strategize a supply chain network that can commercialize an expected target number of unknown disease state assets. While all deliverable tools are constrained by Excel's processing capabilities, we suggest that future work focus on improving robustness of the model and building additional scenario simulation iterations.

This capstone presents an adaptable and scalable framework to drive robust decision-making with what limited information is available on the rapidly growing and highly competitive gene therapy global market. Our portfolio model and scenario planning tools will enable Roche strategically make the best decisions for their gene therapy portfolio supply chain network design.

References

American Society of Gene & Cell Therapy (ASGCT). (2020, October). ASGCT Clinical Trials Finder. American Society of Gene & Cell Therapy. https://asgct.careboxhealth.com/

Cook, A. G. (2006). *Forecasting for the Pharmaceutical Industry: Models for New Product and In-market Forecasting and how to Use Them*. Gower. https://books.google.com/books?id=5IaddXKSNTkC

Commissioner, O. of the. (2020, February 20). *The Drug Development Process*. FDA; FDA. https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process

Curran, K. (2021). The science, the history and the future of gene therapy. Rising Tide Biology. Retrieved May 4, 2021, from https://www.risingtidebio.com/what-is-gene-therapy-uses/

Deloitte, Mooraj, H., Kawalekar, O., Gupta, L., & Shah, S. (2020, April 17). *Cell and gene therapies: Delivering scientific innovation requires operating model innovation*. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/life-sciences/operating-models-for-gene-cell-therapy-manufacturing-process.html

Ding, M., & Eliashberg, J. (2008). A Dynamic Competitive Forecasting Model Incorporating Dyadic Decision Making. *Management Science*, *54*(4), 820–834. https://doi.org/10.1287/mnsc.1070.0798

Dunbar, C. E., High, K. A., Joung, J. K., Kohn, D. B., Ozawa, K., & Sadelain, M. (2018). Gene therapy comes of age. Science, 359(6372), eaan4672. https://doi.org/10.1126/science.aan4672

European Medicines Agency (EMA). (2019, March 25). Zynteglo. https://www.ema.europa.eu/en/medicines/human/EPAR/zynteglo

F. Hoffmann-La Roche Ltd. (2019, December 17). *Roche concludes acquisition of Spark Therapeutics, Inc. to strengthen presence in gene therapy*. Roche Media Releases. https://www.roche.com/media/releases/med-cor-2019-12-17b.htm

Gene Therapy Net, & Bleijs, D. A. (n.d.). *Gene Therapy Products on the Market*. Gene Therapy Net. https://www.genetherapynet.com/gene-therapy-products-on-the-market.html

Genetic and Rare Diseases Information Center (GARD). (2021, January 26). *FAQs About Rare Diseases | Genetic and Rare Diseases Information Center (GARD) – an NCATS Program*. National Institutes of Health (NIH). https://rarediseases.info.nih.gov/diseases/pages/31/faqs-about-rare-diseases

Grabowski, H. G., Ridley, D. B., Schulman, K. A., & Philipson, T. J. (2007). Entry and Competition in Generic Biologics [with Commentary]. *Managerial and Decision Economics*, *28*(4/5), 439. https://doi.org/10.1002/mde.1352

Hahn, S. M., & Abernethy, A. (2020). Rare Disease Day 2020: FDA Continues Important Work on Treatments for Rare Diseases. *FDA*. https://www.fda.gov/news-events/fda-voices/rare-disease-day-2020-fda-continues-important-work-treatments-rare-diseases

Heyman, J. B. (Jeffrey B. (2010). *Simulation modeling to predict drug pipeline throughput in early pharmaceutical R&D* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/59167

Jackson, E. A., Kuester, M. K., & Naber, J. M. (2020, November 10). Gene and Cell Therapies. The Actuary Magazine. https://theactuarymagazine.org/gene-and-cell-therapies/#en-6647-1

Kirschner, J., & Cathomen, T. (2020). Gene therapy for monogenic inherited disorders—opportunities and challenges. *Deutsches Aerzteblatt Online*. Published. https://www.aerzteblatt.de/int/archive/article/217169/Gene-therapy-for-monogenic-inherited-disorders-opportunities-and-challenges

Merkuryeva, G., Valberga, A. and Smirnov, A., 2019. Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science*, [online] 149, pp.3-10. Available at: https://www.sciencedirect.com/science/article/pii/S1877050919301061 [Accessed 5 November 2018].

Mullin, E. (2017, October 24). *Tracking the Cost of Gene Therapy*. MIT Technology Review. https://www.technologyreview.com/2017/10/24/148183/tracking-the-cost-of-gene-therapy/

National Heart, Lung, and Blood Institute (NHLBI). (2018, January 12). *The coming of age of gene therapy: A review of the past and path forward*. National Institutes of Health (NIH); U.S. Department of Health & Human Services (HHS). https://www.nih.gov/news-events/news-releases/coming-age-gene-therapy-review-past-path-forward

NHS website. (2020, December 21). Ibuprofen for children. Nhs.Uk. https://www.nhs.uk/medicines/ibuprofen-for-children/

NHS website. (2021, March 2). Ibuprofen for adults (including Nurofen). Nhs.Uk. https://www.nhs.uk/medicines/ibuprofen-for-adults/

Orphanet. (2021) Retrieved May 14, 2021, from http://www.orpha.net/consor/www/cgi-bin/index.php?lng=EN

PhRMA. (2015). *Biopharmaceutical Research & Development: The Process Behind New Medicines*. PhRMA.

Puiu, M., & Dan, D. (2010). Rare diseases, from European resolutions and recommendations to actual measures and strategies. *Maedica*, *5*(2), 128–131.

Quinn, C., Young, C., Thomas, J., & Trusheim, M. (2019). Estimating the Clinical Pipeline of Cell and Gene Therapies and Their Potential Economic Impact on the US Healthcare System. *Value in Health*, *22*(6), 621–626. https://doi.org/10.1016/j.jval.2019.03.014

Sack, B. K., & Herzog, R. W. (2009). Evading the immune response upon in vivo gene therapy with viral vectors. Current opinion in molecular therapeutics, 11(5), 493–503.

Salzman, R., Cook, F., Hunt, T., Malech, H. L., Reilly, P., Foss-Campbell, B., & Barrett, D. (2018). Addressing the Value of Gene Therapy and Enhancing Patient Access to Transformative Treatments. *Molecular Therapy, 26*(12), 2717-2726. doi:10.1016/j.ymthe.2018.10.017

Sertkaya, A., Wong, H. H., Jessup, A., & Beleche, T. (2016). Key cost drivers of pharmaceutical clinical trials in the United States. Clinical Trials, 13(2), 117–126. https://pubmed.ncbi.nlm.nih.gov/26908540/

Song, C.H., Han, JW. Patent cliff and strategic switch: exploring strategic design possibilities in the pharmaceutical industry. SpringerPlus 5, 692 (2016). https://doi.org/10.1186/s40064-016-2323-1

Talbot, K., & Tizzano, E. F. (2017). The clinical landscape for SMA in a new therapeutic era. Gene Therapy, 24(9), 529–533. https://www.nature.com/articles/gt201752

Teixeira, A., Wu, J., Luchi, M., Smith, B., & Choy, M. (2019, November 18). *From Lab to Marketplace, Succeeding With Gene Therapies*. BCG Global. https://www.bcg.com/publications/2019/lab-marketplace-succeeding-gene-therapies

Touchot, N., & Flume, M. (2015). The payers' perspective on gene therapies. *Nature Biotechnology*, *33*(9), 902–904. https://doi.org/10.1038/nbt.3332

Therapeutics for Rare and Neglected Diseases (TRND) Program. (2017, December 15). *Gene Therapy Platform for Rare Diseases*. National Center for Advancing Translational Sciences; National Institutes of Health (NIH). https://ncats.nih.gov/trnd/projects/gene-therapy

U.S. Food & Drug Administration (FDA). (1993, October 14). *Application of Current Statutory Authorities to Human Somatic Cell Therapy Products and Gene Therapy Products; Notice* (Docket No. 93N--0173). Department of Health and Human Services (HHS). *Federal Register, Part II, Vol. 58, No. 197.* Retrieved November 2, 2020, from https://www.fda.gov/media/76647/download

U.S. Food & Drug Administration (FDA). (2015, September 21). Additional Protections for Children. U.S. Food and Drug Administration. https://www.fda.gov/science-research/clinical-trials-and-human-subject-protection/additional-protections-children

U.S. Food & Drug Administration (FDA). (2017, August 30). FDA approval brings first gene therapy to the United States [Press release]. https://www.fda.gov/news-events/press-announcements/fda-approval-brings-first-gene-therapy-united-states

U.S. Food & Drug Administration (FDA). (2020, August 28). FDA Continues Strong Support of Innovation in Development of Gene Therapy Products [Press release]. https://www.fda.gov/news-events/press-announcements/fda-continues-strong-support-innovation-development-gene-therapy-products

U.S. National Library of Medicine (NLM). (2020, September 17). *How does gene therapy work?: MedlinePlus Genetics*. Medlineplus.Gov; National Institutes of Health (NIH). https://medlineplus.gov/genetics/understanding/therapy/procedures/

Vakratsas, D., & Kolsarici, C. (2008). A dual-market diffusion model for a new prescription pharmaceutical. *International Journal of Research in Marketing*, *25*(4), 282–293. https://doi.org/10.1016/j.ijresmar.2008.05.002

Wong, C. H., Siah, K. W., & Lo, A. W. (2019). Estimation of clinical trial success rates and related parameters. *Biostatistics*, *20*(2), 273–286. https://doi.org/10.1093/biostatistics/kxx069

Zivin, J. (2000). Understanding CLINICAL TRIALS. *Scientific American, 282*(4), 69-75. Retrieved November 2, 2020, from http://www.jstor.org/stable/26058674

Appendix A: Severity and Epidemiology Probability

As we started working with the data, we first needed to classify what we determined as terminal based on the average age of death from the data set in Table A1.

Table A1

Terminal Disease Criteria

Average Age of Death	Terminal (T)
Adolescent	Yes
Early Childhood	Yes
Infantile	Yes
Late Childhood	Yes
Young Adult	Yes
Normal Life Expectancy	No
Any Age	No
Adult	No
Elderly	No
% Terminal Orphanet Diseases	**71.83%**
% Non-Terminal Orphanet Diseases	**28.17%**

In Table A2 we also had to classify what counted as child onset.

Table A2

Child/Minor Disease Onset Criteria

Average Age of Onset	Child/Minor (C)
Adolescent	Yes
Antenatal	Yes
Childhood	Yes
Infancy	Yes
Neonatal	Yes
All Ages	No
Adult	No
Elderly	No
% Child/Minor Onset Orphanet Diseases	**77.11%**
% Non-Child/Minor Onset Orphanet Diseases	**22.89%**

Table A3 shows the probability that a disease is either both terminal and child onset, exclusively either terminal and child onset, or neither. In order to arrive at Table B.3 we assumed that the probabilities of these events are independent.

Table A3

Severity Classification Probabilities

Severity Classification	Conditions	Probability
High	$T \cap C$	55.39%
Med	$T \oplus C$	38.16%
Low	$T' \cap C'$	6.45%

Table A4 additionally shows the point prevalence classification from various data sources for orphan diseases.

Table A4

Point Prevalence Classifications and Evaluation of Frequency

Class #	Known Point Prevalence	US+EU Share	WW Share	Commercial Share
1	>1 / 1,000	0.18%	0.00%	0.00%
2	6-9 / 10,000	0.46%	0.03%	0.00%
3	1-5 / 10,000	21.61%	1.07%	12.50%
4	1-9 / 100,000	35.64%	2.61%	50.00%
5	1-9 / 1,000,000	20.96%	1.16%	21.88%
6	<1 / 1,000,000	21.15%	95.12%	15.63%

Table A5 is the data used and the classification used for the to arrive at the "Commercial Share" column in Table A4.

Table A5

Classification of Disease Pursed by Companies (Curran 2021)

Disease	Gene of interest	Company	Prevalence
AADC deficiency (CNS)	AADC	PTC Therapeutics (GT-AADC)	<1 / 1 000 000
ADA-SCID	adenosine deaminase	Orchard Therapeutics (Strimvelis, EMA approved)	1-9 / 1 000 000
Alpha-1 antitrypsin deficiency	A1AT	Adverum	1-5 / 10 000
β-Thalassemia (severe sickle cell)	Hemoglobin (β-chain)	Bluebird Bio (Zynteglo, EMA approved)	1-9 / 100 000
Cerebral ALD	ABCD1	Bluebird Bio (Lenti-D)	1-9 / 100 000
Choroideremia	CHM	Biogen/Nightstar, Spark	1-9 / 100 000
Congestive heart failure	Adenyl cyclase 6	Renova (RT-100)	<1 / 1 000 000
Cystic Fibrosis	CTFR	Vertex, Boehringer Ingelheim	1-9 / 100 000
Duchenne muscular dystrophy (DMD)	Dystrophin	Sarepta, Pfizer, Audentes, Solid	1-9 / 100 000
Fabry disease	alpha-galactosidase A	UniQure, Sangamo	1-5 / 10 000
Glaucoma	BDNF pathway	Astellas	1-9 / 100 000
Hemophilia A	Factor VIII	BioMarin, Spark, Shire, Sangamo, UniQure	1-9 / 100 000

Disease	Gene of interest	Company	Prevalence
Hemophilia B	Factor IX	Spark/Pfizer, UniQure, Sangamo, Freeline	1-9 / 100 000
HIV	CCR5 negative CD4 cells	American Gene Technology	1-5 / 10 000
HoFH (hypercholesterolemia)	LDLR	RegenxBio	1-9 / 1 000 000
Huntington's Disease	huntingtin	UniQure	1-9 / 100 000
Lipoprotein lipase deficiency	Lipoprotein lipase	UniQure (Glybera, EMA approval)	1-9 / 1 000 000
Leber's hereditary optic neuropathy (LHON)	ND4	GenSight Biologics	1-9 / 100 000
Leber's congenital amaurosis (LCA)	CEP290	ProQR	1-9 / 100 000
Metachromatic leukodystrophy	ARSA	Orchard	1-9 / 100 000
MPS I (Hurler syndrome)	IDUA	Sangamo* RegenxBio	1-9 / 1 000 000
MPS II (Hunter's syndrome)	IDS	Sangamo*, RegenexBio	1-9 / 100 000
MPS III (Sanfilippo Syndrome)	SGSH	Abeona	1-9 / 1 000 000
Parkinson's disease	AADC	Voyager	<1 / 1 000 000
Pompe Disease	acid alpha-glucosidase	Sarepta, Audentes	1-9 / 1 000 000
Recessive Dystrophic Epidermolysis Bullosa	Colagen C7	Abeona (EB-101)	<1 / 1 000 000
RPE65 deficiency (vision loss)	RPE65	Spark (Luxturna, FDA approved)	1-5 / 10 000
Spinal Muscular Atrophy (SMA I)	SMN1	Novartis (Zolgensma, FDA approved)	1-9 / 100 000
Wiskott Aldrich syndrome (WAS)	WAS	Orchard	1-9 / 1 000 000
X-linked myotubular myopathy	MTM1	Audentes	1-9 / 1 000 000
X-linked retinitis pigmentosa	RPGR	Biogen/Nightstar	<1 / 1 000 000
X-linked SCID	IL2RG	Mustang Bio	1-9 / 100 000