
AN APPLIEDXL AND KALSHI WHITE PAPER

Biopharma's Public Probability

*The State and Future of Prediction
Markets in Drug Development*

AppliedXL • Kalshi

July 2026

© 2026 AppliedXL, Inc. All rights reserved.

About This Report

This report examines biopharma prediction markets as an emerging field: the opportunities they present, the risks and design challenges they raise, and the standards and processes that would need to exist for them to function reliably. It was written jointly by AppliedXL and Kalshi, and draws on interviews with clinicians, academics, biopharma R&D and strategy professionals, bioethicists, and investors. Quotations from named individuals are drawn from those interviews; being quoted here does not imply that an interviewee or their affiliated institutions endorses prediction markets, AppliedXL, Kalshi, or the report's conclusions.

Disclosures

AppliedXL builds resolution infrastructure for biopharma prediction markets. Kalshi operates a regulated prediction market exchange on which biopharma contracts trade. Both organizations have a commercial interest in the growth and credibility of this market and believe these markets are worth building if built well.

Nothing in this report is investment, legal, or medical advice, or a recommendation to trade any contract or security. Prediction-market prices are aggregated expectations, not statements of fact about any drug, trial, or company, and any specific contracts or prices mentioned are illustrative. Company and program names appear only where they are matters of public record.

Introduction: The Idea Pharma Almost Built

In 2003, Eli Lilly ran a quiet experiment that worked better than almost anyone expected. About fifty employees, chemists, biologists, and project managers with no formal authority over which drugs advanced, traded six of the company's drug candidates through an internal market. The market pulled together what was scattered across the organization, toxicology, clinical, and commercial signals, and ranked the candidates more accurately than the company's existing process; it correctly identified the three that would go on to be most successful.¹ What made the result striking was not just the accuracy, it was what the trading revealed that a survey never could: a willingness to pay \$70 for a candidate expressed a confidence that a \$60 bid did not, a gradient of conviction that disappears the moment you reduce a question to a show of hands. As Alpheus Bingham, then Vice President of Lilly Research Laboratories Strategy, put it in an interview for this report: "When we start trading stock, and I try buying your stock cheaper and cheaper, it forces us to a way of agreeing that never really occurs in any other kind of conversation. That is the power of the market."

The experiment was a success on both outcome and process: it accurately ranked the drugs and generated richer information aggregation and dialogue. What it never became was a permanent fixture. Lilly did not run it at production scale, and no major pharmaceutical company has since. "We had trouble getting a pharma company to go all the way to a full prediction market," Bingham said. The reasons are structural rather than a failing of the companies that tried, a pattern the economist Robin Hanson has documented across corporate experiments at firms from Google to Ford: internal markets tend to be as accurate as or more accurate than management's own forecasts, yet they sit uneasily inside a hierarchy, where a continuously published probability can cut against decisions already made through normal channels. They tend to be wound down within a few years for organizational rather than empirical reasons.² The encouraging part of that finding is the part this report builds on: the markets worked: what defeated them was where they were placed, not what they did.

Three features of corporate life explain why, and none requires anyone to act in bad faith. The first is hierarchy: when a senior leader has championed a drug, subordinates tend to self-censor, and the dissent a market would surface is rarely given room to run. The second is anonymity, or its absence: when individual positions are visible inside the company, when a colleague or a superior can see that you personally bet against your own division's lead compound, people stop trading candidly, or stop trading at all. The third is institutional fit: a live internal probability can sit awkwardly against resource allocations already settled through normal planning, and an instrument that complicates settled decisions is easy to set aside. These are facts about organizations, not flaws in the idea.

The significance is that all three dissolve outside the company. Two decades later, the infrastructure that did not exist in 2003 has been built, and it has been built precisely where those three barriers do not apply. Public exchanges have no internal hierarchy to protect. Anonymity is a design feature, not a concession. No manager's budget can vote them out of existence. Exchanges like Kalshi hold CFTC Designated Contract Market

approval, and millions of individuals participate in markets that span politics, technology, the environment, and more.

These markets already exist, but only sporadically: a scattering of individual contracts rather than a coherent category. On regulated exchanges today, you can already trade questions such as *Will this Phase 3 trial meet its primary endpoint?* or *Will the FDA approve this drug for this indication by year-end?* Each contract resolves to a yes or no and pays out on the answer. What does not yet exist is the depth, the coverage, or the settlement infrastructure to make those scattered prices trustworthy at scale.

This report explores that opportunity, the challenges that come with it, and how they might be addressed. The opportunity is real: a public, continuously updated probability on questions that today only well-resourced insiders can price, and, if the accuracy holds, an incentive-aligned external estimate that could sharpen the capital-allocation decisions on which the pace and cost of drug development turn.

So are the risks: to the integrity of the trials being priced, to patients and the public who may misread a number, and to the markets themselves if they settle in ways participants cannot trust. The promise and the hazards travel together, and the rest of this report examines both, beginning with the problem these markets are meant to solve, turning early to the risks and ethics they raise, and only then to how they might be designed and resolved well enough to be worth the trouble. The honest bottom line up front: the obstacles that confined Lilly's experiment are gone, but an open, liquid, public market presents its own questions, challenges, and opportunities. Those are what this report explores.

Part I: The Information Problem These Markets Are Trying to Solve

What the Industry Already Knows, and Who Gets to Know It

The odds that a drug will succeed are among the most valuable numbers in the economy, and among the least visible. Bringing a single drug to market costs an estimated \$2.3 billion, and a company's fortunes can turn on one trial result.³ Insiders form a view on the odds every day: banks, expert networks, and pharmaceutical companies all produce estimates, but they stay behind closed doors. Even the official public record is incomplete, though not for the reason it first appears. The FDA Amendments Act of 2007 requires most trials to post summary results to ClinicalTrials.gov within twelve months of completion, yet as of the FDA's April 2026 disclosures, roughly 30% of trials highly likely to be subject to that requirement had posted none.⁴ The results often exist elsewhere, in press releases, regulatory filings, and conference presentations, but not in the standardized public channel the law designates, which turns verification into slow, manual work. The scale of the compliance gap is stark: by one analysis from the University of Oxford's Bennett Institute for Applied Data Science, the maximum civil penalties the FDA could in theory have levied against non-reporting sponsors exceed \$120 billion, at a per-day rate that rises with inflation each year.⁵

Bringing a drug to patients is one of the hardest things organized human effort attempts. A single modern Phase 3 program can enroll thousands to tens of thousands of patients across hundreds of trial sites in many countries, coordinated for years against a protocol fixed in advance, all to test a molecule engineered to act on a precise biological target, a receptor, an enzyme, a single misfolded protein, with enough specificity to change the course of a disease without harming everything around it. The people who do this work are attempting something genuinely difficult, and most attempts do not succeed, not because anyone fails at their jobs, but because biology is unforgiving and the bar for proving a new medicine safe and effective is, rightly, very high.

The numbers capture the scale of the challenge. One large study of more than 21,000 compounds found that a drug entering clinical trials had roughly a 13.8% chance of eventually reaching approval, and closer to 3.4% in oncology.⁶ The global clinical trials market was about \$83 billion in 2024, much of it invested in programs that, despite serious science behind them, will not cross the finish line.⁷ This is the nature of frontier research, not evidence of dysfunction; biological uncertainty is high and that is precisely why the work is hard. But within that uncertainty there is a narrower, addressable problem: capital sometimes advances programs that informed observers already doubt, because the mechanisms for surfacing that informed skepticism are private, expensive, or simply unavailable to most of the people who would benefit from them.

And the stakes have never been higher. The global pharmaceutical market reached roughly \$1.7 trillion in 2024 and is on track toward \$2 trillion in the next few years.⁸ The pace of innovation is accelerating on several fronts at once: AI-assisted discovery is compressing development timelines and expanding pipelines; China has risen from a peripheral player to the world's second-largest source of new drug candidates, accounting for roughly

30% of the global innovative-drug pipeline by 2025 and generating record license-out deal value; and regulators are building faster routes to market, including the FDA's new priority pathways aimed at cutting review times from the standard ten-to-twelve months to as little as one to two.^{9,10} More candidates, moving faster, against bigger bets, which makes an accurate, early read on which programs will succeed more valuable than it has ever been.

Probability of success (PoS) estimates for clinical trials and regulatory approvals are foundational to the entire industry. Investment banks publish PoS models for their clients. Expert networks connect institutional investors to physicians and researchers with domain knowledge. Large pharmaceutical companies run sophisticated internal forecasting processes built around exactly these questions. This information exists and is produced with real rigor; the firms that generate it do so for sound reasons, because it is proprietary, competitively sensitive, costly to produce, and in many cases legally constrained in how it can be shared. The point is not that anyone is hoarding it improperly. The point is simply that the public answer does not exist alongside the private one.

Sell-side PoS estimates sit behind subscription paywalls, as commercial research reasonably does. Expert network calls cost thousands of dollars per hour. Internal pharma forecasting models, appropriately, stay inside the company. The result, through no one's fault, is a two-tier information ecosystem: large institutions with the resources to assemble private probability estimates on one side, and smaller investors, researchers, physicians, patient advocates, and journalists working from press releases and public filings on the other.

Prediction markets do not generate new probability information. They make the aggregate of what is known publicly visible, continuously and in real time. Whether that aggregation is accurate enough to be useful is a question addressed later in this report; the structural point is simpler, that the question these markets answer is not new, but the public answer has not existed before.

Markets have two kinds of efficiency: information efficiency and allocation efficiency. Information efficiency means the knowledge is aggregated in some way to represent a more accurate factual, in the case of a stock market, a valuation for the company.

— Alpheus Bingham, former VP, Lilly Research Laboratories Strategy

The Biotech Equity Problem

Biotech is the most event-driven sector in the stock market, which is to say that company values often hinge on single yes-or-no moments rather than on quarterly earnings or gradual growth. A clinical trial result or an FDA decision can make or break a company in a day. These pivotal moments are known as catalysts, scheduled or anticipated events, a trial readout, an advisory committee vote, a regulatory decision date, that the market knows are coming and that can move a stock sharply when they land. A single Phase 3 readout or FDA decision can move a company's stock 50% overnight, and for many smaller biotechs, one trial or one regulatory decision determines whether the company survives at all.

Investors naturally want to position themselves around these catalysts, to take a view on whether a given trial will succeed. But the stock market gives them no clean way to do it. Buying or shorting a biotech's shares means betting on the whole company at once: the drug program, yes, but also its management, its cash on hand, its other pipeline programs, the mood of the sector, and the direction of the broader market. An investor can be exactly right about the drug and still lose money because the company stumbles elsewhere, or wrong about the drug and profit anyway. The share price blends the one question the investor cares about with a dozen they do not.

Prediction markets offer a real-time signal that's independent of biotech company stock prices, which reflect the broader company rather than the specific drug program.

— Haris Vikis, Principal, Oracle Health

There are financial instruments that get partway there. Stock options let an investor bet on how much a stock will move around a catalyst without betting on which direction, useful when you expect a big swing but are not sure which way. But that is still a bet on the size of the share-price reaction, not on the underlying question of whether the drug actually works; the two are related, but not the same thing, and the difference is exactly what gets lost.

A prediction market contract is different in kind: one trial, one approval, one price. An investor with a view on whether a specific drug will meet its primary endpoint can express precisely that view, without taking on all the unrelated variables that come bundled with owning the stock. That is a genuinely new capability. Whether the market's price is accurate enough to rely on as an investment input is a separate question the category is still working to answer.

Sponsor Optimism, Accountability, and Transparency

Two problems sit between a trial's results and an honest public understanding of them: the results may never appear at all, and when they do, they are often framed to look better than they are. Together they define an accountability gap that a public probability could help address.

In late March 2026, federal regulators put the first problem on the record. The FDA sent messages to more than 2,200 companies and researchers, tied to more than 3,000 registered clinical trials, some of them publicly funded, that appeared to have missed their legal obligation to post results to ClinicalTrials.gov.⁴ Federal statute requires most trials to submit summary results within twelve months of their primary completion date; the messages sought voluntary compliance before the agency decides whether to escalate to formal notices of noncompliance and civil penalties that accrue for each day a violation continues. The agency's own analysis found that 29.6% of studies highly likely to be subject to mandatory reporting had submitted no results at all.

The regulators were explicit about why it matters. As FDA Commissioner Marty Makary put it, sponsors "have an ethical obligation to make results public regardless of the data's influence on the company's share price." When trials go unreported, especially those with unfavorable outcomes, the scientific record skews: successes accumulate in the literature while failures quietly disappear, distorting what clinicians, policymakers, and investors believe they know about how drugs actually perform. Independent analyses put the scale of historical non-reporting even higher: one New England Journal of Medicine study found only 38.3% of trials reported results within the period examined, and a later Lancet analysis found roughly 41% reported promptly.¹¹

The results that are published bring the second problem. Spin, the practice of framing neutral or failed outcomes as positive, is not an edge case. A landmark 2010 analysis in the Journal of the American Medical Association examined 72 published trials that missed statistical significance on their primary endpoint and found spin in the conclusions of 42 of their abstracts, presenting the results more favorably than the data supported: titles implying efficacy where none was demonstrated, abstracts omitting the primary outcome, conclusions declaring success on the strength of secondary endpoints that cannot on their own establish clinical benefit.¹²

The mechanisms are well understood and not, for the most part, cynical. Sponsors invested years and capital in a compound, and the pull toward finding something salvageable in a disappointing dataset is human nature under financial pressure; for a small biotech with a single-asset pipeline, spin can even be a survival strategy. The techniques are consistent: mining post-hoc subgroups until something reaches significance, or bundling hard and soft outcomes into composite endpoints so a drug that fails on mortality can still claim success on a lab value, while the primary-endpoint failure sits in the press release's final paragraph.

What makes both problems acute is timing. The existing checks are real but slow: SEC disclosure does not require real-time accuracy in how a program is characterized, FDA labeling review happens after data is submitted, and the literature corrects spin only eventually. Spin lives in that lag, in the press releases, presentations, and abstracts that precede the full record, and a prediction market priced by participants with money at stake occupies

exactly that gap as a live disagreement register. When a sponsor calls a Phase 3 program on track while the market price has fallen from 65% to 18% over six months, that divergence does not prove the sponsor wrong, but it documents that financially motivated participants disagree with the public characterization, and creates a contemporaneous record that did not exist before.

The limit is worth stating plainly: a market can only price a program that throws off some public signal, so what these markets check is the characterization of disclosed results, not the silence of undisclosed ones, and whether the check is meaningful rather than noise depends on price accuracy, participant expertise, and resolution quality, none of it yet settled.

The Track Record

Prediction markets have outperformed alternative forecasting methods in political, financial, and climate forecasting. The question is whether that record extends to biopharma.

The historical evidence is encouraging but limited. The Iowa Electronic Health Markets ran a flu-forecasting market in the 2004–05 season with 61 health care workers from a variety of backgrounds. By the end of a target week the market's forecast was correct about 71% of the time, against roughly 36% for a forecast based on historical averages alone; one week in advance, accuracy was about 50%.¹³ Eli Lilly's internal market correctly ranked drug candidates. Corporate internal markets at firms including Google and Ford outperformed in-house expert forecasts in published accounts, improving on experts by as much as a 25% reduction in mean-squared error, though Google's own researchers found that employees who sat physically close to one another tended to trade in correlated ways, a reminder that aggregation quality depends on the independence of the participants.¹⁴

The accuracy record for commercial biopharma prediction markets, the category this report covers, is thin. Most contracts trade between \$3,000 and \$30,000 of lifetime volume. While large trades on thinner contracts have tended to self-correct toward equilibrium, greater liquidity is still needed before making definitive statements about market accuracy. A price derived from a handful of participants carries different informational content than one derived from thousands. Validation now waits on something specific: contracts deep enough that no single trade can move the price, which means volumes the category has not yet reached.

This points to the category's central unproven claim. Every strong accuracy result above, Iowa, Lilly, the corporate markets, comes from a restricted pool of vetted experts, which is closer to a structured internal forecast than to an open exchange. There is a real case that open markets inherit the accuracy: a larger, financially motivated pool should aggregate at least as well, and public markets have outperformed experts in other domains. The honest position is that the transfer is plausible but undemonstrated in this specific domain, and that the early commercial contracts are the experiment that will settle it. At current low volumes, no one should claim these markets are more accurate than biotech stocks; the report does not. What it commits to instead is verification: as trading volume grows, straightforward accuracy checks, how often contracts priced high resolve YES, how often

the market beats a naive baseline, will be published rather than asserted, so the claim is settled by evidence in the open.

Why these markets have not been applied to science at scale before is the question Part V addresses directly. The short answer is that resolution complexity, regulatory uncertainty, ethical questions, and thin liquidity have each presented obstacles the category is still working through.

From Lilly to Kalshi: A Compressed History

The modern biopharma prediction market category is roughly thirty months old in its current commercial form. The path from the Lilly experiment to today ran through a government cancellation, a pandemic, and two decades of regulatory development.

In July 2003, the same year Lilly ran its experiment, the Defense Advanced Research Projects Agency cancelled the Policy Analysis Market, a project that would have run prediction markets on geopolitical and economic events. Senators Ron Wyden and Byron Dorgan led the backlash, branding it a “terrorism futures market,” and it was scrapped within days.¹⁵ The idea survived in academic venues: the Iowa Electronic Health Markets ran invitation-only health markets through the mid-2000s and a multi-state influenza pilot from 2008 to 2010.

In 2009, the idea moved closer to the clinic. Pharmer’s Market, launched that October by MIT Sloan researchers, Ragu Bharadwaj, a former Vertex Pharmaceuticals research scientist, working with faculty Eric von Hippel and Fiona Murray and advised by Harvard’s Peter Coles, used a broad anonymous community of pharma researchers, chemists, academics, financial analysts, and clinicians to forecast the probability that six breast cancer drugs would clear Phase I, II, and III. Built on Crowdcaster’s platform, its premise was the one this report opened with: that siloed information drives up R&D costs and suppresses approval rates, and that a market could surface distributed knowledge faster and more accurately than any individual expert. It was the clearest early demonstration that the Lilly thesis could be pointed at specific drugs, in public.¹⁶

Anonymous market structures may be the only mechanism that lets certain signals come out. Without anonymity, information stays trapped inside the companies that hold it.

— Ragu Bharadwaj, creator of Pharmer’s Market (2009)

The COVID pandemic was the turning point. For the first time, a broad public wanted to track the same biomedical question at once, whether and when a vaccine would arrive, and prediction markets gave them a place to express it. The vaccine and emergency-use-authorization contracts of 2020 were the category’s first genuinely mainstream moment, and a wave of academic forecasting on pandemic outcomes followed. The drug-specific build-out has been steadily growing since then.

The barrier is no longer whether these markets can exist. It is whether their prices can be trusted when they settle, a question about resolution, and the subject of Part VII.

Part II: The Risks and Objections

Before describing how these markets are built, this report puts the case against them first. A prediction market on a clinical trial raises genuine risks, to the integrity of the science, to patients and the public, to the markets themselves, and it draws principled objections that deserve to be stated at full strength rather than waved away. What follows sets out the most substantive objections in the form a serious critic would put them, with the response the category has developed for each. Where a satisfactory answer exists, it is given; where the concern remains open, that is said plainly. Many of the answers point forward to specific design choices, which the later parts of this report describe in detail; the order is deliberate, the risks come before the solutions, so that the design can be read as a response to them rather than a brochure that never mentions them.

1. “Insiders will capture most of the profits”

The concern. People with privileged access to trial data will trade ahead of outcomes and extract value at the expense of uninformed participants. The concern is felt acutely inside the industry. As Eli Weinberg put it: “Companies will be worried that prediction markets may create a demand for use of confidential information; the risk is that you create a new way for people to monetize information.”

The analysis. The public enforcement record does not support the broadest version of this concern. As Part V details, the canonical biotech insider-trading cases, Martoma, Skowron, Dagar, Catenacci, all involved people with trial-wide visibility: safety-monitoring or steering-committee members, a lead biostatistician, a lead investigator.^{17,18,19,20} Not one involved a site-level nurse or coordinator, whose blinded view of roughly 1% of patients is not a sufficient base to predict an endpoint. The risk is real but identifiable, and the relevant population can be explicitly excluded from trading related contracts. Kalshi’s Source Agency Prohibition, its implementation of the so-called “Eddie Murphy Rule,” is one such mechanism, barring members who hold material nonpublic government information from trading the affected contracts.²¹ The NBA’s Jontay Porter case, which produced a lifetime ban and a federal guilty plea for conspiracy to commit wire fraud after he tipped confidential information to manipulate prop bets, established why positioned participants must be categorically excluded.²² The same logic applies here.

The platform-level safeguards are also more developed than the category’s newness suggests. On a CFTC-regulated exchange, insider trading is not merely discouraged but prohibited under the same federal provisions that govern securities, Section 6(c)(1) and Regulation 180.1, which courts have read to mirror the securities insider-trading rules. Kalshi’s CFTC-certified rulebook extends past that federal floor, defining as prohibited any trading by a person who can access material nonpublic information before its public release, who is an employee or affiliate of a contract’s source agency, or who has any direct or indirect influence over the underlying outcome, alongside broader bars on fraudulent, manipulative, or deceptive activity. Enforcement is not only governmental. The exchange screens defined categories at onboarding, including politicians, government officials, and athletes, either

blocking them or applying special restrictions; it monitors activity through internal surveillance and third-party vendors; it freezes flagged accounts pending investigation; and its disciplinary process can impose fines, disgorgement, and permanent suspension, with suspected unlawful conduct referred to law enforcement.²³ This is the machinery behind the screening that, in the Van Dyke matter noted below, rejected a prospective trader before he could place the trade he later placed on an unregulated venue.

Two honest gaps remain. The subcontracted CRO blind spot is not yet adequately addressed by existing exchange-level policies: a statistician at a firm subcontracted by a CRO holds the same nonpublic information as a sponsor employee, with none of the same compliance visibility. This is addressable at the exchange level, trading prohibitions can be written to explicitly cover subcontracted personnel who hold material nonpublic information, regardless of their direct employer, but the coverage does not yet exist. And the enforcement infrastructure itself, though built on the same legal foundation as securities law, is newer and less tested. CFTC Rule 180.1, adopted under Dodd-Frank in 2011, uses the same misappropriation theory the SEC uses in stock cases, from *United States v. O'Hagan* (1997) and *Salman v. United States* (2016); on February 25, 2026 the CFTC formalized that this authority reaches event contracts, and weeks later put it into practice, charging Gannon Ken Van Dyke, a U.S. Army Special Forces sergeant, with using classified information to buy “Maduro Out” shares on Polymarket for roughly \$400,000, its first insider-trading case in event contracts, alongside a parallel DOJ action.^{24,25} Screening at the regulated venue had rejected Van Dyke months earlier, pushing him to an unregulated platform, and at the platform level the exchange has disciplined members for trading on nonpublic information, in one disclosed case imposing \$20,397.58 in combined disgorgement and penalty plus a two-year suspension.²⁴ The gap narrows with each action but has not closed.

2. “Financial instruments on medical outcomes are illegitimate, it’s gambling, and it commodifies patients”

The concern. Betting on whether a drug will be approved treats patients as variables in a financial equation, and is functionally gambling that should be regulated as such. The view has serious adherents: veteran biotech journalists, among others, have argued that these markets carry no real public benefit and are simply another form of betting dressed up as information. That position deserves to be taken seriously rather than waved away, and part of it is unanswerable on its own terms: if someone believes any financial instrument tied to a medical outcome is inherently wrong, no feature of the design will change their mind.

The analysis. The same outcomes are already traded at far larger scale in biotech equities. Short-sellers profit when trials fail. Options market makers profit from readout volatility. An analysis of 98 products in Phase 3 trials between 1990 and 1998 found biotech stocks moved an average of +27% for eventual winners and -4% for losers in the 120 days before the public announcement, a divergence significant at $p=0.0007$, consistent with informed trading well before disclosure.²⁶ That trading has existed in equity markets for decades, in a less regulated environment, with no position limits. A CFTC-regulated event contract, with

position limits, identity verification, and federal insider-trading law, is a tighter regulatory environment than the equity markets the category's critics accept without comment.

On the gambling question, the legal trend favors the financial-instrument view, though it is not settled. On April 6, 2026, a divided Third Circuit panel ruled 2-1 in *KalshiEX LLC v. Flaherty* that Kalshi's sports-related event contracts are swaps under the Commodity Exchange Act, financial instruments subject to federal regulation rather than gambling subject to state gaming law, and affirmed a preliminary injunction barring New Jersey from enforcing its gambling statutes against them.²⁷ The ruling addressed sports contracts specifically, and as a preliminary injunction it found a likelihood of success rather than deciding the merits. Biopharma contracts sit inside the same federal framework, and the reasoning extends naturally to them, but no court has yet ruled on event contracts as a class. With the Ninth Circuit having heard parallel cases on April 16, 2026, a circuit split could send the question to the Supreme Court.

The behavioral profile reinforces the point. CFTC-regulated Designated Contract Markets apply KYC requirements, position limits, and reporting obligations. Novel drug approval and Phase 3 contracts carry time horizons of twelve to twenty-four months between listing and resolution and require genuine domain expertise to price. A market on a 2027 PDUFA decision has more in common with equity options than with in-play sports betting.

What the regulatory argument cannot do is answer the values objection, that any financial instrument on a medical outcome is inappropriate regardless of governance. The category will need to earn trust through demonstrated behavior, not structural comparisons. The narrower, honest claim the report does stand behind is this: trial outcomes are already traded heavily through biotech stocks, and a regulated prediction market can offer a cleaner, more transparent signal on a specific event than that diffuse trading does. Whether it genuinely improves accountability, rather than simply adding another venue to bet in, remains an open empirical question that only a track record can settle.

A caveat that applies wherever this report invokes the equity-market comparison: three decades of biotech trading show no documented harm to equipoise, enrollment, or review integrity, which is meaningful evidence but not proof of absence for diffuse effects no one has studied directly.

3. “Thin markets are vulnerable to manipulation”

The concern. In thin contracts, a single large trader can move prices and create misleading signals, and in small trials with subjective outcomes, a motivated actor could influence the endpoint determination itself and trade on the result. As Haris Vikis of Oracle Health put it, naming the two risks together: “Insider trading risk is very real, and thin liquidity opens the door to manipulation.”

The analysis. Both versions of the concern are answered by the same scoping decision. The initial scope concentrates contracts on large-company Phase 3 programs with the densest public information ecosystem in the category and heavy institutional equity-market participation. A single large trade is unlikely to persist in a market where hundreds of analysts and investors are independently tracking the same program. Kalshi imposes per-contract position limits, with KYC requirements and third-party monitoring.²⁸ The subjective-endpoint version applies primarily to small Phase 2 trials, which are not in scope; Phase 3 trials have blinded independent central review, pre-specified statistical analysis plans, and DSMB oversight, and the inter-reader variability documented in oncology imaging is precisely what that central review is built to neutralize through independent multi-reader consensus.

Additionally, prediction markets benefit from a robust incentive mechanism to self-correct any price distortion, whether it be manipulation or not. Other asset classes, such as equities and bonds, do not possess a concept of “fair value”, the price of a stock is whatever someone is willing to pay for it. This is not the case on prediction markets, because ultimately the price of a prediction market contract will be either zero cents or one dollar. The fair value exists; it is the likelihood that the event will happen. When prices get distorted, the profit incentive drives other market participants to immediately correct the market back to equilibrium. Evidence from Kalshi has shown this applies even to thinly traded markets.

A related risk is worth naming because it follows directly from putting money on an outcome: once a participant holds a position, the cheapest way to move a binary outcome is sometimes not to trade but to talk. A holder of a NO position can publicly campaign against a trial, amplify unfavorable readings of ambiguous data, or pressure the people whose words move the price, turning passive speculation into incentivized advocacy for a program’s failure. The category already has a preview: in 2026, bettors sent death threats to a reporter whose account was about to settle a large pool, trying to force a different story. The same scoping that blunts manipulation blunts this, large-sponsor Phase 3 contracts resolve against named institutional documents, not sentiment, so a campaign cannot change the outcome unless it changes the underlying filing, and the people most able to influence an outcome are already barred from trading. What that does not touch is diffuse public influence by ordinary position-holders, which is a speech and platform-governance problem more than a contract-design one, and one whose harms land hardest on the individuals who become settlement triggers; it deserves explicit platform conduct rules and watching as volumes grow.

4. “Visible prices will distort the institutions doing the science”

The concern. Publicly visible prices could corrupt the very experiments they track: eroding clinical equipoise by making physicians unwilling to enroll patients in placebo arms, deterring enrollment while a trial is still recruiting, pressuring FDA reviewers on pending decisions, and politicizing the review process. There is also a subtler version, that the signal is easily misread, since a 90% “negative” price still means the drug works in something like one in ten scenarios, a nuance that rarely survives transmission to a lay audience.

The analysis. The interference vector is the one scoping answers most directly: the post-recruitment listing rule and the exclusion of early-phase contracts (Part V) mean a market opens only after enrollment has closed, so it cannot deter enrollment, and the most fragile stages of the science are never listed. On the broader institutional worry, the equity-market comparison applies in its strongest form: biotech stocks have priced trial outcomes and FDA decisions publicly for three decades, and the institutions have held. Equipoise’s structural protections, independent DSMBs, pre-specified analysis plans, blinded adjudication, operate independently of any outside market; FDA reviewers have worked inside continuous stock-market pressure throughout that period with no documented case of market pressure changing a review outcome, and the agency still meets most of its review-performance goals.²⁹ A prediction market price is a smaller signal than the stock moves and analyst coverage that already accompany every major FDA decision: it makes existing information more legible, it does not create it. The misreading risk is real and not fully solved by timing, which is why the report treats endpoint-specific, non-verdict resolution language and plain-language market explainers as commitments rather than options.

5. “Patients will be harmed, they’ll bet against their own survival, or stop enrolling”

The concern. Patients enrolled in trials will short the outcomes of their own treatment, and patients who can follow a public probability market will choose to wait for the approved drug rather than enroll.

The analysis. Patients enrolled in a specific trial are prohibited from trading contracts tied to that trial, the same categorical exclusion that applies to sponsor employees and investigators. The informed population these markets aggregate is different: patients living with a disease who are not enrolled, who follow the research closely and have no ability to influence its outcome. On enrollment, no published study documents the predicted effect; the established barriers are design complexity, geography, eligibility, and insurance, not information. The honest caveat is that the population following a disease area is not always cleanly separable from the population enrolled in its trials, which is why rare-disease contexts warrant ongoing engagement with patient-advocacy communities rather than a blanket assurance.

6. “Markets add nothing, peer review and existing financial tools already do this, and sponsors will game them upstream”

The concern. Peer review, FDA review, and existing financial analysis already process the relevant information rigorously. And to the extent markets matter, sponsors will optimize trial design and public communications for market optics rather than scientific rigor.

The analysis. Prediction markets address a different problem: the absence of a continuously updated, publicly accessible probability estimate on a specific binary outcome. Peer review assesses scientific quality. FDA review assesses safety and efficacy. Neither produces a live approval probability. Existing financial sources, sell-side PoS estimates, expert network calls, internal pharma models, are episodic, paywalled, or proprietary. A prediction market contract isolates a single binary question. That is a structural property, not a claim to superior analytical wisdom.

On upstream gaming: by Phase 3, the room for sponsor distortion is limited, because design, endpoints, and the statistical analysis plan are all locked and registered before unblinding, as Part V details; the flexible earlier phases are not listed.

Part III: The Deeper Ethical Questions

The objections above are the arguments the category meets most often. The ethical questions run deeper, and they do not all dissolve under good design. The framework here draws on Dr. Jonathan Kimmelman, James McGill Professor and Associate Member of the Department of Experimental Medicine at McGill University, who organizes the concerns into a set of bins; several map onto risks already addressed above (interference with the science, and incentivized influence), and the three that raise distinct questions are taken in turn here.

A premise runs underneath all of them. A prediction market on a geopolitical event prices an answer that exists independently of whether anyone is watching. A clinical trial is different: it is actively constructing its answer, and watching it can change it. That asymmetry is what makes biopharma markets ethically distinct from markets on elections or sports.

Regulatory trust and patient equity. A high market price signals broad confidence before the evidence is formally established, creating two harms. The first is an equity grievance: a patient who cannot get into a trial, for reasons of geography, eligibility, or distance, sees a public market implying the drug works and reasonably asks why an effective treatment is being withheld. The second is a regulatory-integrity harm: visible, viral market momentum can generate political and social pressure on regulators that is decoupled from the evidence, especially in a media environment that conflates market confidence with scientific consensus. The late-phase, post-enrollment scope partly addresses the access grievance, since by Phase 3 the drug is being studied in a controlled cohort, not withheld from population use; the more important mitigations are commitments of communication discipline, resolution language that is strictly technical and endpoint-specific rather than a “success/failure” verdict, and a plain-language explainer on every market, with particular caution for conditions with intense advocacy visibility such as ALS or rare pediatric diseases. The deeper version is not a design bug at all: the risk that public sentiment pressures a regulator who should answer only to evidence is a genuine tension between transparency and evidentiary discipline, to be managed through institutional norms and watched over time.

Endpoint mismatch and the paradox of prediction. Markets are structured around endpoints that matter to investors, and communicating those signals to people whose lives are at stake is dangerous. Progression-free survival is the canonical example: a clean, measurable endpoint that drives regulatory and investment strategy but can mean little to a patient for whom overall survival is what matters. A market reading “strong probability of meeting the primary endpoint” will be heard as “strong probability this drug helps me,” and those are not the same statement. There is a related distortion in where capital goes, toward safe, large-market bets rather than true unmet need, which a mechanism that makes sentiment more visible could deepen.

What the regulatory argument cannot do is answer the values objection, that any financial instrument on a medical outcome is inappropriate regardless of governance. The

category will need to earn trust through demonstrated behavior, not structural comparisons, and the gambling question is not legally final until the appellate split resolves. The narrower, honest claim the report does stand behind is this: trial outcomes are already traded heavily through biotech stocks, and a regulated prediction market can offer a cleaner, more transparent signal on a specific event than that diffuse trading does. Whether it genuinely improves accountability, rather than simply adding another venue to bet in, remains an open empirical question that only a track record can settle. A caveat that applies wherever this report invokes the equity-market comparison: three decades of biotech trading show no documented harm to equipoise, enrollment, or review integrity, which is meaningful evidence but not proof of absence for diffuse effects no one has studied directly.

Underlying all of this is a structural observation worth stating in full. As Kimmelman put it: “If your prediction markets are working really, really well, we’re probably not doing the science right, because the whole point of science is to investigate unpredictable questions.” The apparent paradox is that a market predicting trial outcomes with high accuracy implies the answer was knowable in advance, raising the question of why the trial was necessary. But this conflates two different things: whether an outcome is predictable and whether it must be demonstrated. Confirmatory testing in patients is a regulatory and ethical prerequisite for approval regardless of how knowable the result was. No drug is licensed on a forecast, so an accurate market does not make the trial redundant. And there is a second reason the premise doesn’t hold: a Phase 3 prediction market is not pricing the underlying biological unknown, which was investigated in Phases 1 and 2, where this report lists nothing. By Phase 3 the scientific question is largely settled in expectation, and what remains uncertain is execution and regulatory judgment, whether the effect size holds at scale, whether the trial runs cleanly, whether the FDA agrees the endpoint supports approval. A market that prices those well is not making the science redundant; it is pricing the residual uncertainty good science leaves on the table. The paradox bites hardest against markets on early, exploratory science, exactly the category this report excludes, and it remains a standing caution against overclaiming. The design answer is to frame every contract at the endpoint level, never the drug level, to add secondary layers tracking clinical alongside statistical significance where feasible, and to put endpoint glossaries on market pages; the research-capital distortion is only partly answerable by design, and AppliedXL is tracking whether listings in fact cluster around large-market indications over genuine unmet need.

Information cascades and research conservatism. Standard bubble dynamics apply with extra force in a domain already marked by risk aversion: rising prices attract buyers following the price rather than the data, and the downstream decision-makers in medicine are not all sophisticated market participants. Markets could deepen conservatism by pricing high-confidence, late-stage, large-indication trials most readily, making the unconventional bets, the potential breakthroughs, even harder to fund. The early-phase exclusion and the liquidity-before-display threshold blunt the cascade, and the design commits to regular curation audits. This is the concern the category should hold most honestly, because the very choices that make a market trustworthy, late-stage, large-sponsor, liquid, are the ones that bias the listed set toward the already-favored. The right response is external scrutiny: Kimmelman has proposed funding independent scholars to assess whether the presence of

these markets measurably affects research investment, and to publish the findings whichever way they fall. The category should support exactly that.

Why the Risks Come First

The parts that follow describe what these markets resolve on, how they are designed, what they are useful for, and how they settle. Each is, in part, a response to a risk or objection raised in these two parts. Putting the risks and ethics first is deliberate: the value of a public probability and its capacity to mislead travel together, and the design choices only earn their place if the reader has already seen, in full, the harms they are meant to contain. A reader who finishes the next parts unconvinced should be able to point back to the specific concern here that the design failed to answer.

Part IV: What These Markets Resolve On

Two contract types account for the vast majority of informational value and resolution complexity in the category. This report focuses on those two.

Trial Readout Markets

Contract template. *Will [drug] meet the prespecified primary endpoint of [trial], with statistical significance, by [date]? Resolves YES on confirmation that all prespecified primary endpoints were met at the committed level of statistical significance; otherwise NO.*

Start with what an endpoint is, because the whole contract rests on it. Before a trial begins, its designers commit in writing to exactly what they are measuring and what result will count as success, the primary endpoint. It might be how much a tumor shrinks, how long patients live without their disease worsening, how far a symptom score falls. This target is registered publicly, on ClinicalTrials.gov, before any patient is enrolled, which is what makes it useful for a market: the definition of success is fixed in advance and cannot be moved after the fact.

That is also why “did the trial produce positive results?” is the wrong question to build a contract on. “Positive” is a matter of interpretation, and interpretation is exactly what sponsors, with every legitimate incentive to present their program well, will shade favorably. The analytical question, the one with a checkable answer, is narrower: did the trial meet the specific primary endpoint it set for itself, at the level of statistical confidence it committed to? That reframing, from a subjective “positive” to an objective “met the prespecified endpoint,” is the foundation of a contract that can be resolved without dispute.

Three complications make even that narrower question subtle, and a well-written contract has to account for all three. First, a trial can have more than one primary endpoint, two or three, and they are not always reported at the same moment; a contract has to say whether it resolves on one, several, or all of them, and what happens if they arrive separately. Second, meeting an endpoint is a statistical judgment, not a verdict on the medicine: a result can clear the prespecified threshold for statistical significance while the actual effect is modest. Third, and following from the second, a trial can hit its endpoint without the benefit being clinically meaningful, statistically real but too small to change how a patient actually lives. A contract on “meets the primary endpoint” answers the statistical question, not the clinical one, and the difference has to be stated plainly so no one mistakes one for the other.

There's a lot of nuance in clinical trials. Beyond statistical significance, clinically relevant study outcomes are often not binary, and continuous variables like response and survival will likely need to be grouped in ranges. Furthermore, results will likely need comparison to historical benchmarks in many cases.

— Dr. Alexander Drilon, medical oncologist; Clinical Director, Early Drug Development Service, Memorial Sloan Kettering Cancer Center

A further wrinkle is that a single readout is reported in several places, a sponsor press release, a conference presentation, a regulatory filing, a peer-reviewed paper, a ClinicalTrials.gov posting, sometimes with subtly different framings of the same numbers. The resolution standard the category is converging on cuts through this by anchoring to one thing: all prespecified primary endpoints met with statistical significance, confirmed against the registered protocol on ClinicalTrials.gov rather than against the sponsor's characterization of its own results. When the picture is ambiguous, incomplete, or inconsistent with the registered protocol, the default is NO; partial results keep the market open; favorable sponsor language without the statistical evidence does not qualify.

Why meeting the endpoint and winning approval are two different questions.

The most important thing a contract designer can understand is that a trial meeting its endpoint and a drug winning FDA approval are not the same event, and must never be written as one. Lykos Therapeutics' MDMA-assisted therapy for PTSD is the clearest illustration. Its pivotal Phase 3 trials, MAPP1 and MAPP2, did meet their prespecified primary endpoint, the CAPS-5 PTSD severity score, and their key secondary endpoint, with the primary result highly statistically significant and published in Nature Medicine.³⁰ By the endpoint test, the trials succeeded. And yet the FDA did not approve the drug: on June 4, 2024, its advisory committee voted that the data did not establish effectiveness and that the benefits did not outweigh the risks, and a Complete Response Letter followed on August 9, 2024, asking for an additional trial.³¹ The decision turned not on a missed endpoint but on questions about how the trials were conducted, including the difficulty of keeping patients unblinded to a drug with strong perceptual effects.

The lesson sits at the foundation of sound contract design. A contract written on "meets the primary endpoint" would have resolved YES; a contract written on "FDA approval" would have resolved NO; both would have been correct, because they asked genuinely different questions. Conflating the two is the single most common way a biopharma contract produces a result its participants feel cheated by. Endpoint contracts and approval contracts have to be separate instruments, written and resolved against separate documents.

Regulatory Approval Markets

Contract template. *Will the FDA issue a full approval letter for [drug], for [indication], by [date]?* Resolves YES on issuance of a qualifying FDA approval letter by the date; otherwise NO.

The resolution template has converged on a standard now used across most major contracts on the regulated exchanges. It resolves YES only on an actual FDA approval letter (an NDA, BLA, sNDA, sBLA, ANDA, or 351(k), in standard, accelerated, REMS, or indication-limited form) and explicitly excludes the near-misses that have caused disputes: approvable letters, tentative approvals, PDUFA-date extensions, expanded access, EUAs without full approval, and CRLs, with a CRL or withdrawal resolving the market immediately NO. The verbosity of that boilerplate is itself a record of what the category has learned, each exclusion added after an ambiguity produced, or threatened to produce, a contested resolution.

The resolution trigger is a single named document: an FDA approval letter recorded in Drugs@FDA, issued on a public date. Either it exists by the contract end date or it does not. That clarity is why approval contracts are the most liquid and most mature in the category.

The edge cases. What happens when FDA approves a narrower indication than the contract specified? When a PDUFA date is extended by three days into the contract's grace window? When FDA approves on a surrogate endpoint the contract did not anticipate? None of these has yet produced a high-profile public dispute. When one does, it will test whether the resolution infrastructure is adequate, and whether the category's first major controversy arrives before the governance frameworks are ready. A quieter failure mode is the contract that simply goes stale. As Alexa Devos, a Global Strategic Insights Manager at Intuitive Surgical, observed: "One area where estimates feel the least reliable is when a trial is aging with no updates; this becomes another area where insights are missing." A market with nothing new to price drifts on noise, which is its own argument for tight scoping and clear end-dates.

An illustration. Consider, hypothetically, a contract on whether a widely watched obesity drug will win FDA approval for a new use by the end of a given year. As pivotal trial data lands, strong results would tend to push the price up; but if the standard ten-to-twelve-month review clock makes approval within the contract window unlikely absent an expedited pathway, the price can stay well below 50% even on good news, because the contract asks about approval by a date, not about whether the drug works. The two questions, will it work and will it be approved in time, pull the price in different directions, which is exactly the kind of distinction a well-specified contract is built to keep separate. The point of such a contract is not that the market is provably right; it is that the price updates continuously and transparently on public information, in a way no static analyst estimate does.

Other Contract Types

Beyond trial readouts and regulatory approvals, the category includes advisory committee vote contracts, which resolve on the public roll call at FDA panel meetings and have historically functioned as leading indicators for the subsequent approval decision; drug shortage contracts, which track the FDA Drug Shortage Database; label expansion contracts for new indications or patient populations of already-approved drugs; pandemic and emergency-use authorization contracts, which dominated the category in 2020–21; and a small number of drug pricing, IRA negotiation, and M&A contracts. Each has a role in a fully developed ecosystem. None is the focus of this report.

Part V: Market Design as Risk Mitigation

Prediction markets on biopharma outcomes carry domain-specific risks: insider trading from participants with privileged access to trial data, manipulation in thin markets, resolution disputes when endpoints are ambiguous, contamination of the trial whose outcome is being priced, harm to vulnerable issuers, and ethical questions about financial instruments on clinical outcomes. The choices made at the scoping stage determine how serious those risks are in practice.

What follows describes the design choices AppliedXL has made in scoping the initial contracts in this category, the reasoning behind each, and the areas where the design does not yet have satisfactory answers. Each choice is best read as a risk control rather than a product preference, and most of them mitigate more than one risk at once. Restricting the initial scope to Phase 3 readouts addresses resolution-dispute risk (the endpoint is fixed in a public document before listing), insider-trading risk (it narrows the window of pre-decisional non-public information relative to earlier phases), and trial-contamination risk (by Phase 3 the science is largely settled in expectation). Listing only after recruitment has closed is the trial-integrity control. Restricting to large public sponsors is the manipulation control and the issuer-harm control. Requiring a liquidity threshold before a price is displayed is the signal-reliability and cascade control. The subsections below take each in turn; the residual gaps each control leaves open are collected at the end under Open Problems.

The Principle

The most consequential market design decision is upstream: whether a given contract should be listed at all.

Contracts whose resolution criteria tie to a single public institutional artifact, an FDA approval letter, a ClinicalTrials.gov primary endpoint result, resolve with almost no ambiguity. Contracts whose resolution criteria depend on interpretation of a sponsor press release, or on subjective endpoint assessment, resolve with ambiguity as the baseline condition. Filtering out the second class before listing does more to protect market integrity than any enforcement regime applied after a dispute arises.

The initial scope is therefore defined by a single criterion: list contracts that can be resolved correctly against a named public document, and do not list contracts that cannot.

A second screen sits alongside the first: whether a contract is worth listing at all. As Dr. Luiz Diaz, Head of the Division of Solid Tumor Oncology at Memorial Sloan Kettering, framed the listing question: “First: is it of high importance? Is it an important question of medicine? Is it an important question about development? Is it an important question on Wall Street? And is there an algorithm that could be developed to pick that kind of study? The second thing is: will it be positive or not?” The first half of that test governs whether a contract belongs on the board; the second is what the market itself is for.

Listing Phase 3 Readouts Narrows Resolution and Insider-Trading Risk

To mitigate resolution-dispute and insider-trading risk, the initial markets focus on Phase 3 readouts. By Phase 3, the primary endpoint has been specified in the protocol, registered on ClinicalTrials.gov, and in many cases agreed with FDA through a Special Protocol Assessment, so the resolution question maps to a public document fixed before the trial started, which is what makes the contract resolvable without dispute.

Phase 1 and Phase 2 trials frequently involve exploratory endpoints and interim analyses whose interpretation requires judgment beyond reading a registered protocol. They are also the stage at which insider trading risk is most concentrated: a trial investigator's informational advantage is largest when the public evidence base is thinnest. Phase 1 and Phase 2 contracts are outside the initial scope. This will be revisited as resolution infrastructure matures.

The staging logic also tracks how the nature of the prediction changes once a trial is underway. As Eli Weinberg, the healthcare and life sciences partner at Bain & Company, put it: "Before a trial starts, prediction is a scientific question, do we believe in the biology? Once running, it's execution." A contract listed at Phase 3 is pricing execution against a fixed protocol, not betting on the underlying science still in flux.

Listing After Enrollment Closes Protects Trial Integrity

A contract should not be listed while the trial it prices is still recruiting patients. A live, publicly visible probability can flow back into the experiment that generates it: a market showing lukewarm sentiment may discourage physicians from enrolling patients or referring them, a direct operational threat to the trial's completion and to the integrity of its result. This is the cleanest structural answer to the interference concern, and it is the single most important scoping rule the design adopts after the Phase 3 restriction itself.

In practice this means listing a contract only after enrollment has closed. Because some adaptive trials make "closed" ambiguous, the working threshold is a high level of completed enrollment (on the order of 90% or more) before a market may open. By the time a Phase 3 trial has finished recruiting, the population-level question the market prices is no longer one the market's existence can distort: the cohort is fixed, the protocol is locked, and what remains is the readout. This choice, combined with leaving recruitment-sensitive early-phase trials off the board, removes the primary channel through which a market could corrupt the science it is trying to measure.

Requiring Liquidity Before Display Guards Against False Signals

A price is only information if enough independent participants produced it. A market with three traders is not an aggregation mechanism; it is a rumor with a number attached. The design therefore sets a minimum liquidity and participation threshold that a contract must clear before its price is displayed publicly. Below that threshold the contract may trade, but it does not broadcast a probability that downstream readers, physicians, patients, journalists, would reasonably mistake for a settled expert consensus. This guards against the information-cascade dynamic in which buyers follow a price rather than the underlying data, a dynamic that is more dangerous in medicine than in equities because the people acting on the signal are often not sophisticated market participants.

Focusing on Larger Sponsors First Reduces Manipulation and Issuer Harm

The initial scope prioritizes trials from larger, publicly traded pharmaceutical and biotech companies. This is a sequencing decision rather than a permanent boundary, and it reflects three considerations, each of which points to a safeguard that would generally need to mature before the scope broadens.

The first is manipulation. A contract on a small-cap biotech with limited analyst coverage tends to be more exposed to a single motivated actor than a contract on a large-cap sponsor's program, where many analysts and investors are independently tracking the same information. This is not absolute: large caps are not immune to manipulation, and some smaller companies carry meaningful coverage. But it holds often enough to be a reasonable starting filter, and it is largely a function of market depth and surveillance coverage, both of which can be built up rather than being an inherent property of the smaller company.

The second is resolution complexity. When a large company's drug fails, the equity market typically absorbs a bounded move. When a small biotech's single-asset program fails, the company may cease to function, which can create resolution entanglement the category is not yet well equipped to handle. These edge cases appear tractable with clearer resolution criteria and more operating experience.

There is a third consideration, and it is about harm to the issuer rather than risk to the contract. A large company can usually absorb an adverse market signal through its investor-relations function and the depth of its analyst coverage; a small biotech may be less able to. A bad prediction-market price landing on top of a now-public Complete Response Letter, since the FDA has moved toward publishing CRLs, could in some cases move faster than a small issuer's ability to respond and weigh on its investor base before it can put the result in context. Prioritizing large-cap sponsors sidesteps this for now, but the concern does not disappear; it becomes more salient as the scope expands toward the smaller companies that may be most exposed to it. That expansion therefore should be paced against the surveillance and communication safeguards that would protect vulnerable issuers, not only against the manipulation controls that protect the contract.

Small- and mid-cap programs are expected to be added as market depth, resolution governance, and issuer-protection safeguards mature.

Starting With Novel Approvals Keeps Resolution Clean

For regulatory approval contracts, the initial scope covers novel drug approvals: NDAs and BLAs for new molecular entities seeking first approval. Label expansions introduce scope ambiguity around what counts as the indication the contract specified. Some approval types carry political exposure that makes them difficult to price accurately. Niche pathways introduce resolution complexity the initial architecture does not address.

Novel drug approvals on the standard or priority review pathway have the clearest resolution trigger, the deepest public information ecosystem, and the lowest political surface area of any contract type in the category.

Open Problems

Insider trading. The load-bearing point is informational: a site coordinator in a 3,000-patient blinded trial sees roughly 1% of patients and cannot infer the primary endpoint, while a biostatistician or monitoring-committee chair sees the whole trial. The enforcement record is consistent with this. Every major biotech insider trading prosecution over the past two decades involved someone with trial-wide visibility, not site-level access, though that record shows who was charged, not everyone who traded, and should be read as corroboration rather than proof. The tip in the Martoma/SAC case, at roughly \$276 million the largest the SEC had ever charged, came from Dr. Sidney Gilman, who chaired the safety monitoring committee for a Phase II bapineuzumab trial.¹⁷ In the Skowron/Human Genome Sciences case, the tipper Dr. Yves Benhamou sat on the steering committee of a Phase 3 hepatitis-C trial; the funds avoided about \$30 million in losses.¹⁸ Amit Dagar was the senior statistical program lead on Pfizer's Paxlovid trial, realizing roughly \$214,395 by the SEC's civil measure and more than \$270,000 by the DOJ's criminal figure.¹⁹ Dr. Daniel Catenacci was lead investigator on the Phase II FIGHT trial of bemarituzumab and made \$134,142.²⁰ In each case the trader or tipper held a cross-trial role, and in each the wrongdoing was an individual's breach of a duty owed to the company, not conduct by the company itself; the sponsors were the parties whose confidential information was misused.

The design response is categorical exclusion of the cross-trial population. This extends naturally to subcontracted CRO statisticians, who may hold the same nonpublic information as a direct employee while sitting outside the sponsor's compliance framework. The remedy operates at the exchange and contracting level: exchange rules and CRO subcontracts can name this population directly rather than reaching only sponsor employees, supported by MNPI attestations flowed down through subcontracts, restricted-person lists extended to named CRO personnel, and surveillance keyed to CRO affiliation rather than sponsor employment alone. This closes in contracting what categorical exclusion already establishes in principle, and follows a wider pattern of Kalshi's surveillance and enforcement mechanisms that goes well beyond the letter of the law at preventing illicit market activity.

Resolution ambiguity. Even within the constrained initial scope, a recurring set of ambiguous situations, composite endpoints, mixed primary and secondary outcomes, mid-trial protocol amendments, conditional approvals framed as full ones, premature interim

disclosures, are not edge cases but the baseline operating condition of the category. Part VII enumerates them and what reliable resolution requires to handle each.

Ethical questions. The objection that financial instruments on clinical outcomes are inappropriate regardless of governance is not answered by the design choices above. It is a values disagreement the category will need to address through demonstrated behavior over time, not through argument.

Liquidity. At current volumes, the category is too thin to validate its accuracy claims at scale. This is not a design problem but a growth problem.

The Roadmap

The initial scope is a sequencing choice, not a permanent limitation. Early stage trials with well-specified endpoints will be added when the resolution framework for earlier-stage readouts is validated. Small-cap programs will be added when manipulation-surveillance infrastructure is adequate. Label expansions will be added when contract architecture for scope ambiguity is in place.

The expansion is contingent on the track record the initial contracts build. A high-profile disputed resolution in the category's early phase would set back legitimate biopharma prediction markets significantly. The sequencing reflects that risk. It is a deliberate sequencing choice, not a claim that the easy cases are the only ones that matter; the harder contracts are the destination, not an afterthought.

One tension is worth stating where the scoping decisions are made, not only where their ethics are debated. The same criteria that make a contract trustworthy, late-stage, large-sponsor, liquid, also bias the listed set toward programs that are already well-funded and well-covered, and away from the high-need, high-uncertainty research that most needs external signal. This is a genuine cost of the conservative initial scope, not an incidental one. The roadmap's expansion toward smaller and earlier programs is partly an attempt to repay it, and the curation audits described in Part III are how we intend to know whether we are.

Part VI: Applications

Biopharma prediction markets generate one output: a continuously updated public probability on a specific binary outcome. That output is useful in different ways depending on who is reading it. A pattern recurs across every audience below, each currently relies on information that is partial, sponsor-shaped, or paywalled, and a market price adds an incentive-aligned external estimate they did not have before, so the subsections that follow dwell on what is distinct to each rather than restating that common thread. One caution applies throughout and should be read into every use described below: at current volumes the accuracy of these prices in this domain is not yet established, as Part I sets out, so each application treats the price as one input to weigh, not a settled signal to rely on.

Investors

For investors, the contract does what equity cannot: it isolates the one binary question, will this drug clear this bar, from the dozen unrelated variables bundled into a biotech stock.

The practical uses are specific. Before taking a position in a biotech equity around a binary catalyst, an investor can check the prediction market price against their own probability estimate. If the market is at 65% and the investor's model says 40%, that divergence is worth understanding before sizing a position. It does not indicate who is right; it indicates that a participant with money at stake disagrees, which is the most useful piece of information available before a readout. Prices are also being used as a pipeline-monitoring tool, tracking how external participants price each program in a developer's portfolio and comparing those estimates against internal PoS models. The programs where external and internal estimates diverge most are the highest-priority candidates for reassessment.

The current limitation is liquidity. At \$3,000 to \$30,000 of lifetime volume on most contracts, prices are a useful signal but not a definitive one. The practical implication: treat the price as one input, investigate significant divergences from internal estimates, and do not substitute market prices for analyst coverage at current liquidity levels.

However, liquidity may not be as large of an institutional limitation as one would initially think. Even thinly traded contracts can produce enough of a valid price signal to be effective as a reference point for institutional-sized OTC trades. Susquehanna International Group, one of the largest trading firms in the world, has been transparent about the fact they are willing, and often do, price eight and nine figure OTC swaps based on prediction markets with around \$30,000 in volume. While time and scale is needed to truly determine the accuracy of more thinly traded markets, it is a positive and telling sign that institutions believe in the signal of thinly traded markets enough that they are willing to stake tens of millions of dollars on it.

The signal a sophisticated investor wants is also more granular than a binary approve/reject. As Eli Weinberg noted, what an approval is worth depends on its shape: "It doesn't just mean that you get to approval. It means, potentially, that you get to approval at some level of differentiation or subsector of the market that you can go after." The contract types that resolve cleanly today price the binary; the commercial question of which approval often lives in the label and indication detail a single contract cannot yet capture.

Patients and Patient Advocates

Patients and patient communities tracking drugs in development, particularly in rare disease, where a single program may be the only candidate for a condition, have historically had one primary information source: the sponsor, whose press releases and investor materials are built to support confidence and tend to underreport negative or ambiguous results. A market price comes from the opposite incentive: participants who are wrong lose money. The value here is awareness, not advice. A public market can surface which programs an informed community sees as promising, directing attention and scrutiny toward trials that might otherwise be overlooked, and over time it can help make a complex, jargon-heavy field more legible to the people most affected by it. Rare-disease communities, long among the best-informed lay participants in clinical research out of necessity, monitoring registries, attending conferences, knowing the investigators, are well placed to use that kind of signal to follow the science more closely.

Clinical trials are an essential part of our healthcare system and how we as patients get access to the latest innovations. But the clinical trial process is opaque and difficult to understand for most people. Most patients don't know about the choices available in clinical trials or which programs are most promising. The opportunity to have an open, transparent dataset about trial probabilities is extremely promising and empowering for people. I am excited about what Kalshi is building and the opportunities to empower people while still adhering to ethical guidelines.

— Anne Wojcicki, Founder, 23&Me

What a prediction market is not, and must not be allowed to become, is a guide to personal medical decisions. A price is an aggregate expectation about a regulatory or trial outcome, not a judgment about whether a given treatment is right for a given patient. Claiming that someone might choose a course of treatment, or decide whether to seek a particular drug, on the strength of a market probability is exactly the wrong use, and the report does not endorse it. Those decisions belong with patients and their clinicians, grounded in professional medical advice and individual circumstances. Any explainer language accompanying these markets should be developed together with patient groups rather than imposed on them, and the effect of these markets on trial enrollment should be studied rather than assumed. Prediction markets are not a substitute for medical advice and should never be positioned as one.

A separate and still-unresolved question is whether patients who are not enrolled in a trial but are following a drug they hope to access should be permitted to use these markets to hedge the economic risk of a program's failure. The financial logic is coherent; the ethical concerns are real. This is a question for patient advocacy communities, bioethicists, and regulators to address, not market designers alone.

Pharma and Biotech Strategy Teams

Inside pharmaceutical companies, internal PoS estimates are produced by people close to the program, inside a culture that rewards optimism, and they run systematically higher than realized success rates. An external market price is a corrective: produced by participants with no institutional stake and real money at risk. When a market prices a company's Phase 3 program at 25% while the internal team is at 65%, that divergence warrants an explanation before the Phase 3 budget is approved, and prices on competitor programs give a continuous external read that periodic analyst reports do not.

A further possibility is that these markets do not merely observe drug development but also inform better capital allocation. A continuously updated external probability, priced by participants with capital at risk and no institutional attachment to a program, constitutes a different signal than the internal estimates that guide development spending, which tend to exceed realized outcomes. A sharp divergence between the external price and the internal estimate is itself information, indicating a program that warrants reexamination before additional capital is committed. In aggregate, and conditional on the prices proving accurate, the effect would be capital reallocating away from failing programs earlier and concentrating on stronger candidates sooner. It is the mechanism by which these markets could contribute to more efficient development and, ultimately, to treatments reaching patients sooner.

Outside pricing from knowledgeable participants can provide a useful corrective to internal overconfidence.

— Ragu Bharadwaj, creator of Pharmer's Market

One application prediction markets enable is the one Lilly piloted but did not carry to production scale: a market run inside the company, on the company's own pipeline. The structural barriers that make internal pharma markets hard to sustain are not fully solved by external public markets, but the external design has produced a template internal efforts can now adopt: anonymous trading, cash incentives, transparent resolution criteria, and governance outside the chain of command.

I wonder if any pharma companies would want to run this internally. Have your top hundred scientists in the company, give them each some amount of money to bet each year, and have them manage a portfolio of bets.

— Eli Weinberg, healthcare and life sciences partner, Bain & Company

A second, underappreciated application is manufacturing demand planning. Before a drug is approved, manufacturers begin building inventory against demand assumptions that can vary by a factor of ten, and a major launch with poor demand forecasting can leave more than \$100 million in excess safety stock. Prices on approval outcomes, updating continuously as trial data accumulates and FDA review progresses, provide a better planning input than the episodic PoS estimates manufacturing planners currently rely on.

Physicians and Clinical Researchers

For physicians the value is narrow but real: an incentive-aligned, continuously updated probability that adds a dimension to clinical judgment without replacing it, since a trial priced at 70% is a different prospect than one at 20% even when the sponsor sounds equally confident about both. For researchers, prices are a live external benchmark, and the forecasting literature is consistent that experts who regularly confront estimates differing from their own grow better calibrated over time.

This type of probability data could generate perspectives on how current and future clinical trial resources could be allocated.

— Dr. Alexander Drilon, Clinical Director, Early Drug Development Service, Memorial Sloan Kettering Cancer Center

Some practitioners see this developing into a structural shift in how medical evidence is socialized. Luca Dezzani, Vice President of US Medical Affairs at BioNTech, has argued publicly that the field is moving from “science as theater,” where data is presented to a passive audience at medical congresses, toward “science as a market,” in which the broader medical community stakes positions on what a study will show before it reads out.³² In that vision, the market price of a scientific hypothesis becomes a trusted metric alongside the p-value, and the most influential clinician is not the loudest voice on the podium but the one with the strongest record of correct forecasts. It is a projection rather than a description of the present, and it rests on the same unsettled assumption this report has flagged, that aggregated forecasting accuracy will hold in open markets. But it captures why some inside the industry see prediction markets as more than a trading venue.

Journalists

Biopharma journalism has a vocabulary problem. When a trial produces results that are statistically significant but clinically modest, or when a sponsor characterizes a mixed read-out as positive, or when FDA extends a PDUFA date for the third time, journalists have few tools for conveying the uncertainty accurately and accessibly.

Polling gave political journalism a shared vocabulary for uncertainty: a candidate at 68% in the polling average is a different story than one at 52%, and readers learned to interpret those numbers. The equivalent does not exist in health journalism, where a Phase 3 drug is described as having “promise” or “potential,” language that conveys no probability information. A drug priced at 27% is a different story than one priced at 78%. The number is not an oracle; it is an aggregate of participant judgment from a thin market at current liquidity. But, used with that caveat, it can be a more honest single-number summary of expert opinion on a binary question than the “promise” and “potential” language journalism currently falls back on.

The most direct application is reporting the divergence. The sponsor-on-track-while-the-price-collapses gap described earlier under accountability is, for a journalist, a reportable fact: it documents that participants with money at stake disagree with the public characterization. The resolution record, which programs were priced high and succeeded, which

were priced high and failed, which surprised the market, is an emerging dataset for investigative reporting. Programs that failed after being priced above 70% are candidates for retrospective investigation; programs that succeeded after being priced below 30% raise different questions. The prediction market creates a time-stamped archive of how expert opinion evolved over each program's development, an archive that does not currently exist for any other information source in biopharma.

Regulators

For regulators at the CFTC, the FDA, and in the legislative branch, biopharma prediction markets present two questions: what governance frameworks are needed for the markets to function with integrity, and what information the markets generate that is useful to the regulatory process itself.

On governance, the most pressing gap is explicit trading prohibitions for defined categories of pre-decisional participants, FDA staff, trial investigators, sponsor employees, DSMB members, and subcontracted CRO personnel, barred from contracts tied to data or decisions they hold nonpublic. Kalshi's Source Agency Prohibition is an important step, but like any attestation-based rule it needs an institutional enforcement layer behind it, especially in the drug-approval context where it has not yet been tested. A template emerged in March 2026, when Major League Baseball and the CFTC signed a memorandum of understanding, described as the first between a sports league and a federal agency, to establish an integrity framework for event contracts on the league's games.³³ A parallel CFTC–FDA memorandum would give rules like the Source Agency Prohibition the institutional backing that makes them enforceable rather than self-attested.

On the informational side, prices on specific regulatory outcomes provide a more granular read on how public information is being interpreted than biotech stock prices alone. A market on a specific NDA approval that drops from 85% to 40% the day after an advisory committee briefing document is published is signaling something specific about how that document is being read, a signal available to anyone who checks the price, including the regulators whose decisions the market is pricing.

Part VII: Resolution

A prediction market is only as trustworthy as the process that settles it. Price discovery, information aggregation, and accountability for sponsor claims are not realized if the market resolves incorrectly, resolves late, or resolves in a way participants can reasonably dispute. This is the part of the infrastructure that has not yet been built.

How Markets Currently Settle

Kalshi uses an in-house Markets Team that checks each contract's Source Agency against the Payout Criterion. A Market Outcome Review process allows participants to formally challenge a proposed resolution before it is finalized. The Source Agency Prohibition bars members with material nonpublic information from trading related contracts. A different resolution model operates on the decentralized venues: Polymarket settles through UMA's optimistic oracle, in which a proposed outcome stands unless it is challenged within a set window, with disputes escalating to a token-holder vote, a design whose dispute rate is low but whose vulnerability to governance-level manipulation surfaced in a March 2025 incident.³⁴

Five Archetypes of Ambiguity

Five ambiguity archetypes recur often enough to treat as the baseline operating condition, not the exception.

Composite endpoints where only one component reaches significance. The press release says “the trial met its primary endpoint.” The registered protocol specifies a composite, and only one component cleared the alpha threshold. Whether meeting one component constitutes meeting the composite depends on the trial's prespecified analysis plan, co-primary, composite, and hierarchically tested endpoints are governed by different rules, and most press releases do not include enough statistical detail to determine which applies from public disclosure alone.

Mixed primary and secondary outcomes. The primary endpoint succeeded; two key secondary endpoints failed. The contract referenced “primary endpoint,” but participants priced it as though it covered the overall program result. Resolution against the literal contract language is technically defensible but practically misleading.

Mid-trial protocol amendments. A Phase 3 trial reports positive results, but the alpha boundary was moved via protocol amendment during the trial, and the ClinicalTrials.gov registration reflects the amended threshold. Whether resolution follows the original or amended endpoint is not addressed by standard contract language. Dr. Luiz Diaz of Memorial Sloan Kettering put the integrity problem plainly: “Changing endpoints after the trial is running, it's like changing weight classes for boxers.”

The defense is to freeze the endpoint at the moment of listing: the contract resolves against the prespecified primary endpoint as it stood when the market opened, regardless of any subsequent mid-trial revision. Without that rule, there is a specific exploit. An

insider who knows a trial will miss its original endpoint could take a YES position, then push the sponsor to revise toward a second endpoint they privately expect to hit, laundering nonpublic information into a market outcome. Freezing removes the payoff from that maneuver. Equally important is that the freeze rule be disclosed plainly in the market's terms, because the deeper risk is contestation: a participant holding a large position that resolves against them may lobby to have the contract judged against the revised endpoint instead. As Daniel Taylor, the Arthur Andersen Chaired Professor at The Wharton School and director of the Wharton Forensic Analytics Lab, framed the ambiguity: "Somebody would see a revision in the endpoint, and they would think the contract is for the revised endpoint, as opposed to the original endpoint, and so they'd invest a ton of money." Stating the original-endpoint rule up front, rather than adjudicating it after a dispute, is what denies sophisticated actors the ambiguity to exploit. A harder edge case remains where a revision leaves the original endpoint not merely contested but unevaluable, the trial never generates the data to answer it, in which case resolution is structurally impossible rather than simply difficult, and the contract terms have to specify in advance how that scenario is handled.

Conditional approval framed as full approval. FDA issues accelerated approval; the press cycle describes it as "FDA approval." The contract specifies full standard-pathway approval and resolves NO against the narrative. This is the correct resolution. It will not feel correct to participants who followed the press coverage.

Interim DSMB disclosures before topline readouts. An unblinded interim surfaces publicly before the full topline. The market has effectively resolved in everyone's mind, but the contract's qualifying disclosure has not yet occurred. Whether the interim constitutes a qualifying disclosure is not addressed by most existing contracts.

What Reliable Resolution Requires

Source discipline. Only qualifying institutional disclosures count as evidence; analyst commentary, social media, and secondary press rewrites are excluded. The sources that count: registered data on ClinicalTrials.gov, FDA decisions at Drugs@FDA or in approval letters or CRLs, SEC 8-K filings, peer-reviewed publications, and primary conference disclosures at major medical meetings.

Scoping discipline. The most valuable resolution decision is whether a contract should be listed at all. Contracts that resolve against a named public document are sound; contracts that require interpretation of sponsor language are vulnerable. This distinction should be made before listing, not after a dispute arises.

Independence and resolver conflicts. A resolver knows things the market does not: which way an ambiguous case will be called, and often the timing of a determination before it is public. That makes resolution itself a source of material non-public information. The controls are therefore explicit: at AppliedXL, employees are prohibited from trading any market the company resolves, and the timing and internal processes of a pending resolution are kept confidential. Reviewers cannot hold positions in contracts they adjudicate, and any affiliation with an entity whose outcome is being resolved must be disclosed. A resolver that resolved ambiguous cases to favor an exchange would trade its only durable asset, credibility, for a one-time gain; independence from the venue's commercial interest is the core of the function, not a nicety.

Methodology and human adjudication. Reliable resolution borrows its discipline from journalism. AppliedXL's standard replicates a newsroom's fact-checking process, accelerated with technology but with a human in the loop for adjudication on both the AppliedXL and Kalshi sides, so no contract settles on a machine reading alone. The people overseeing that adjudication have backgrounds in biology and biopharma, so the judgment calls, what a disclosure qualifies as, whether an endpoint was genuinely met, are made by people equipped to read the science. Resolution relies only on publicly disclosed information, FDA actions, trial registries, scientific publications, and company statements; no patient-level or otherwise non-public clinical data is used or disclosed.

Source-correction handling. Public sources are not infallible: a ClinicalTrials.gov posting can be corrected, an 8-K amended, a press release superseded by the authoritative filing. The resolution standard must specify which source governs when they conflict (the authoritative regulatory document over the sponsor's characterization) and how a resolution is treated if the governing source is itself later corrected, so that the rule is fixed in advance rather than improvised under dispute.

Auditability. Every resolution must produce a complete, time-stamped audit trail: which sources were consulted, what the assessment recommended, what the reviewer decided, and why any deviation from the recommendation was made.

A default-to-NO rule. When a qualifying disclosure is ambiguous, incomplete, or inconsistent with the registered protocol, the default resolution is NO. This is a procedural commitment to the principle that a market resolves YES only when the qualifying event has unambiguously occurred.

The Ecosystem Gap

The category does not yet have a credible, independent resolution-data provider, one that checks the authoritative public sources against the registered protocol using a transparent, inspectable method.

Existing services track when decisions are expected; they do not verify what happened against a registered protocol. Kalshi's in-house Markets Team resolves Kalshi's contracts capably, but an exchange resolving its own markets is structurally different from an independent provider whose only product is credible resolution. That limitation isn't a criticism of the platform; it simply reflects that the independent layer is a different function from the one it was built to perform.

This gap is the most consequential unmet need in the category's current infrastructure. The exchanges have been built. The oracle infrastructure has been built. The independent resolution layer that makes prices trustworthy enough to serve institutional, clinical, and journalistic audiences has not.

The deeper point is that the market structure itself is a trust mechanism, but only if it settles credibly. As Ragu Bharadwaj, who built one of the first pharma prediction markets, put it: "The anonymous market structure, with money around it, is a good substitute for the trust." The structure substitutes for trust on the trading side. Resolution is where that substitution either holds or fails.

Conclusion

Eli Lilly's 2003 experiment worked. The market correctly identified the most promising drug candidates and revealed shades of opinion that formal processes could not surface. It was never run at production scale, not because it failed, but because a continuously published probability sits uneasily inside a hierarchy, an incompatibility of structure, not a failing of the company that had the foresight to try it.

Twenty years later, the infrastructure that did not exist in 2003 has been built. The exchanges are regulated. The legal framework for event contracts is taking shape, even as the appellate courts work through its boundaries. The anonymity that internal markets could not provide is a design feature of every major platform. The barriers that killed corporate prediction markets do not apply to external public exchanges.

What has not been built is the resolution infrastructure that decides whether those prices can be trusted. The harder problems this report has named, the ambiguity archetypes, the subcontracted-CRO blind spot, the unproven transfer of accuracy from vetted pools to open ones, the values objection the clinical community has not conceded, are not rhetorical hedges. They are the conditions the category has to satisfy before its prices deserve the weight this report argues they could carry.

The evidence does support three things. Drug development has a structural information problem. Internal mechanisms to surface dispersed expert knowledge have proven hard to sustain for twenty years. And external public prediction markets sidestep the specific structural barriers that constrained every internal attempt: the hierarchy that discourages dissent, the absence of anonymity, and the awkward fit between a live probability and decisions already set through normal channels. Whether they clear those barriers accurately enough, at sufficient scale, with adequate governance, is what the next several years will decide.

The Lilly experiment demonstrated the concept. The current commercial category is the first serious attempt to operationalize it at scale. The gap between demonstration and operationalization is where the work is, and the resolution layer is where that gap is widest.

Glossary

Accelerated approval. An FDA pathway that allows a drug to be approved on the basis of a surrogate endpoint reasonably likely to predict clinical benefit, subject to confirmatory post-approval trials. Distinct from full, standard-pathway approval.

Advisory committee (AdCom). An external panel of experts the FDA convenes to review a drug's data and vote on questions such as efficacy and benefit-risk. The vote is non-binding but historically a strong leading indicator of the agency's decision.

BLA / NDA. Biologics License Application / New Drug Application, the formal submissions through which a sponsor seeks FDA approval to market a biologic or a small-molecule drug, respectively.

CAPS-5. The Clinician-Administered PTSD Scale, a standardized severity measure used as the primary endpoint in the MDMA PTSD trials discussed in the report.

ClinicalTrials.gov. The U.S. public registry where trial protocols and prespecified endpoints are recorded before a trial begins and where results are later posted; a primary source for resolving trial-readout contracts.

Complete Response Letter (CRL). The FDA's formal notice that it will not approve an application in its current form. A CRL resolves an approval contract NO.

Composite endpoint. A single trial endpoint combining several outcomes (for example, a count of any of several adverse events). Meeting one component does not necessarily mean the composite was met.

Designated Contract Market (DCM). A CFTC-regulated exchange authorized to list event contracts (Kalshi is one). DCM status brings KYC, position-limit, and reporting obligations.

DSMB (Data and Safety Monitoring Board). An independent committee that reviews unblinded interim trial data for safety and efficacy and can recommend stopping or continuing a trial.

Endpoint (primary / secondary). The prespecified outcome a trial is designed to measure. The primary endpoint is the basis on which the trial is powered and judged; secondary endpoints are supporting measures.

Equipoise (clinical). Genuine uncertainty in the expert community about whether a treatment is better than the alternative, the ethical precondition for randomizing patients, including to a placebo arm.

Event contract. A financial contract that pays out based on whether a specified real-world event occurs; the instrument traded on prediction-market exchanges. Held by the Third Circuit (for sports contracts) to be a "swap" under the Commodity Exchange Act.

Misappropriation theory. The legal doctrine, established in securities law (O'Hagan) and applied by the CFTC under Rule 180.1, that trading on confidential information in breach of a duty owed to the source of that information is illegal insider trading.

Optimistic oracle. The dispute-based resolution mechanism (used by UMA, which settles Polymarket contracts) in which a proposed outcome stands unless challenged within a window; challenges escalate to a token-holder vote.

PDUFA date. The target date by which the FDA, under the Prescription Drug User Fee Act, aims to complete its review of an application, the most-watched single catalyst in the category.

Phase 1 / 2 / 3. The sequential stages of human clinical testing: Phase 1 (safety, small), Phase 2 (efficacy signal and dosing), Phase 3 (large, confirmatory, the basis for approval). The report's initial scope covers Phase 3 readouts.

Position limit. A cap on how large a position a single participant may hold in a given contract, used to limit manipulation in thin markets.

Price discovery. What a market does while it is open: continuously aggregating participant judgment into a live probability.

Probability of success (PoS). An estimate of the likelihood a program reaches a defined milestone (endpoint success, approval). Produced privately today by banks, expert networks, and internal pharma models.

Resolution. How a market settles when it closes: determining the binary outcome and paying out. A public probability is only as trustworthy as the resolution that settles it, the distinction that runs through this report to its closing argument.

Source Agency Prohibition. Kalshi's rule (an implementation of the "Eddie Murphy Rule") barring members who hold material nonpublic information from a relevant source agency from trading the affected contracts.

Special Protocol Assessment (SPA). A binding agreement between a sponsor and the FDA on a Phase 3 trial's design and endpoints, reached before the trial begins.

Surrogate endpoint. A laboratory or imaging measure used as a stand-in for a direct clinical outcome (for example, tumor shrinkage in place of survival), often the basis for accelerated approval.

Sources and Notes

1. Hamel, G. (2009, September 24). The end of management. *Time*. The figures — roughly 50 employees, six drug candidates, and the market correctly forecasting the three most successful — are documented in *Prediction markets for corporate governance* (Working paper), University of Chicago Law & Economics, <https://chicagounbound.uchicago.edu>, which cites Pethokoukis, J. M. (2004, August 30). All seeing, all knowing. *U.S. News & World Report*. The experiment was also reported in *Nature* (2005). Quotations from Alpheus Bingham are from an interview conducted for this report.
2. Hanson, R. (n.d.). Prediction markets need trial & error. *Overcoming Bias*. <http://www.overcomingbias.com>. See also Cowgill, B., & Zitzewitz, E. (2015). Corporate prediction markets: Evidence from Google, Ford, and Firm X. *The Review of Economic Studies*, 82(4), 1309–1341.
3. McKinsey & Company. (2025, February 11). Pharma's Rx for R&D [The Week in Charts]. <https://www.mckinsey.com/featured-insights/week-in-charts>. Average cost to bring a single drug to market, approximately \$2.3 billion.
4. U.S. Food and Drug Administration. (2026, April 13). FDA reminds more than 2,200 sponsors and researchers to disclose trial results [News release]. <https://www.fda.gov>. The agency's analysis found that 29.6% of studies highly likely to be subject to mandatory reporting had submitted no results; Commissioner Marty Makary is quoted therein.
5. FDAAA TrialsTracker [Live informatics tool]. (2018). University of Oxford, Bennett Institute for Applied Data Science (EBM DataLab). <https://fdaaa.trialstracker.net>. On methods, see DeVito, N. J., Bacon, S., & Goldacre, B. (2018). FDAAA TrialsTracker: A live informatics tool to monitor compliance with FDA requirements to report clinical trial results. *bioRxiv*. <https://doi.org/10.1101/266452>. The penalty figure is a theoretical ceiling: the per-day maximum civil penalty under 21 U.S.C. § 333(f)(3)(B) (\$11,569 in the current schedule, adjusted annually for inflation) applied across all delinquent trials for each day past the reporting deadline, assuming notice issued the day after the missed deadline. Actual penalties levied are far lower; the figure illustrates the scale of non-compliance, not fines owed. The FDA Amendments Act of 2007 (Pub. L. No. 110-85), Title VIII, establishes the ClinicalTrials.gov results-reporting requirement; the twelve-month deadline is set out in the 2016 Final Rule (81 Fed. Reg. 64982).
6. 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://academic.oup.com/biostatistics>.
7. Precedence Research. (n.d.). Clinical trials market size. Clinical trials market size, 2024, approximately \$83.75 billion. This is the size of the clinical-trials services market, distinct from total industry R&D spend.
8. Statista. (n.d.). Global pharmaceutical industry — Statistics & facts. <https://www.statista.com>. Global pharmaceutical market approximately \$1.7 trillion in 2024; industry forecasts from multiple research firms (e.g., Precedence Research, Fortune Business Insights) place the market between roughly \$1.8 trillion and \$2.1 trillion in 2026.
9. Clifford Chance. (n.d.). The rise of the Chinese biotech sector. <https://www.cliffordchance.com>; and Center for International Relations and Sustainable Development. (n.d.). A biopharmaceutical superpower: China's rise. <https://www.cirs.org>. China accounted for roughly 30.5% of the global innovative-drug pipeline in 2025 (versus 33% for the U.S.), up from 23% in

- 2023, and generated record license-out transaction value (figures ranging from approximately \$50 billion to \$135 billion depending on methodology and period).
10. U.S. Food and Drug Administration. (n.d.). Novel drug approvals for 2025. <https://www.fda.gov>; and GoodRx. (n.d.). FDA approval stats. <https://www.goodrx.com>. FDA novel-drug approvals totaled 46 in 2025 (CDER), alongside new expedited-review mechanisms such as the Commissioner's National Priority Voucher program, which the agency describes as compressing review timelines from the standard 10–12 months toward 1–2 months. On AI's effect on discovery timelines and pipeline expansion, see Precedence Research / BioSpace. (n.d.). Pharmaceutical market transformation: AI driving innovation. <https://www.biospace.com>. Note that annual novel approvals have not risen monotonically (55 in 2023, 50 in 2024, 46 in 2025); the acceleration is in pipeline size, discovery tooling, and review pathways rather than in raw approval counts.
 11. Anderson, M. L., Chiswell, K., Peterson, E. D., Tasneem, A., Topping, J., & Califf, R. M. (2015). Compliance with results reporting at ClinicalTrials.gov. *New England Journal of Medicine*, 372, 1031–1039. <https://www.nejm.org>; and DeVito, N. J., Bacon, S., & Goldacre, B. (2020). Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: A cohort study. *The Lancet*, 395, 361–369.
 12. Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *Journal of the American Medical Association*, 303(20), 2058–2064. <https://pubmed.ncbi.nlm.nih.gov/20501928>. Of 72 eligible trials with nonsignificant primary outcomes, spin appeared in the abstract conclusions of 42 (58.3%), with more than 40% of reports showing spin in at least two main-text sections.
 13. Polgreen, P. M., Nelson, F. D., & Neumann, G. R. (2007). Use of prediction markets to forecast infectious disease activity. *Clinical Infectious Diseases*, 44(2), 272–279. <https://academic.oup.com/cid>. Accuracy figures (71% by end of target week, approximately 50% one week ahead, approximately 36% historical-average baseline) as reported by CIDRAP. (n.d.). Iowa group bets on market to predict pandemics. <https://www.cidrap.umn.edu>.
 14. Cowgill, B., & Zitzewitz, E. (2015). Corporate prediction markets: Evidence from Google, Ford, and Firm X. *The Review of Economic Studies*, 82(4), 1309–1341; and Chen, K.-Y., & Plott, C. R. (2002). Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem (Social Science Working Paper No. 1131). California Institute of Technology. On the geographic-proximity effect, see the Google internal-markets findings reported in Cowgill & Zitzewitz.
 15. The DARPA Policy Analysis Market (part of the FutureMAP program) was announced and cancelled in late July 2003 following criticism led by Senators Ron Wyden and Byron Dorgan. See Congressional Research Service / Federation of American Scientists records, <https://sgp.fas.org/congress>; and Policy Analysis Market. (n.d.). In *Wikipedia*.
 16. Details of Pharmer's Market — the participants, the platform, and the slate of drug contracts — are drawn from an interview with its creator, Ragu Bharadwaj, conducted for this report, and from contemporaneous accounts of the project's 2009 launch on the Crowdcast platform.
 17. U.S. Securities and Exchange Commission. (2012). SEC charges hedge fund firm CR Intrinsic and two others in \$276 million insider trading scheme involving Alzheimer's drug (Press Release No. 2012-237). <https://www.sec.gov>. The tip to Mathew Martoma came from Dr. Sidney Gilman, chair of the safety monitoring committee for the Phase II bapineuzumab trial.

18. U.S. Securities and Exchange Commission. (2011). Litigation Release No. LR-21928 and Press Release No. 2011-91. <https://www.sec.gov>. See also Hedge fund manager charged with insider trading in Human Genome Sciences case. (2011, April 13). *The Washington Post*. Dr. Yves Benhamou served on the steering committee for the Phase 3 Albuferon (albinterferon alfa-2b) hepatitis-C trial; Joseph “Chip” Skowron’s funds avoided roughly \$30 million in losses.
19. U.S. Securities and Exchange Commission. (2023). Press Release No. 2023-123. <https://www.sec.gov>. The release states that Amit Dagar’s trading generated approximately \$214,395 in illicit profits; the parallel DOJ/SDNY criminal case cited a figure of more than \$270,000. Dagar was senior statistical program lead on Pfizer’s Paxlovid trial. See also Fierce Pharma, <https://www.fiercepharma.com>.
20. U.S. Securities and Exchange Commission. (2021). Litigation Release No. LR-25813 and Press Release No. 2021-264. <https://www.sec.gov>. Dr. Daniel V. T. Catenacci, lead investigator on the Phase II FIGHT trial of bemarituzumab (Five Prime Therapeutics), realized \$134,142. See also BioSpace, <https://www.biospace.com>. A related case, SEC v. Holly Hand (Neuralstem), involved a clinical-trial project manager and is documented in SEC Press Release No. 2021-94.
21. Kalshi’s “Source Agency Prohibition” (an implementation of the so-called “Eddie Murphy Rule”) is defined in the KalshiEX LLC rulebook (CFTC org-rules filings). <https://www.cftc.gov>.
22. Jontay Porter received an NBA lifetime ban (April 2024) and pleaded guilty to conspiracy to commit wire fraud. See ESPN, <https://www.espn.com>; and CBS Sports.
23. Kalshi’s prohibited-trader definitions, surveillance and onboarding-screening practices, and disciplinary process are described in Insider trading prohibitions, Kalshi Market Integrity Hub. <https://kalshi.com/market-integrity/insider-trading>. The federal provisions referenced (7 U.S.C. § 6(c)(1) and CFTC Regulation 180.1) and the Section 4c(a)(4) “Eddie Murphy” rule are summarized there and in the CFTC enforcement advisory cited at note 24.
24. U.S. Commodity Futures Trading Commission, Division of Enforcement. (2026, February 25). Advisory on enforcement authority over event contracts (CFTC Release No. 9185-26). <https://www.cftc.gov>. The advisory documents a Kalshi platform disciplinary action: \$5,397.58 disgorgement plus a \$15,000 penalty (\$20,397.58 total) and a two-year suspension of a member who traded on advance knowledge of unreleased content.
25. U.S. Commodity Futures Trading Commission. (2026). CFTC charges U.S. service member with insider trading in Nicolás Maduro-related event contracts (Release No. 9217-26). <https://www.cftc.gov>; and U.S. Department of Justice. (2026, April 23). U.S. soldier charged with using classified information to profit from prediction market bets. <https://www.justice.gov/opa>. Defendant Gannon Ken Van Dyke; profits of roughly \$404,000 (CFTC) to approximately \$409,881 (DOJ). The CFTC characterized this as its first insider-trading case involving event contracts.
26. Biotechnology stock prices before public announcements: Evidence of insider trading? (2000). PubMed PMID 10736971. Across 98 products in Phase 3 trials, 1990–1998, the divergence between winners (+27%) and losers (-4%) over the 120 days before announcement was significant at $p = 0.0007$.
27. KalshiEX LLC v. Flaherty, No. 25-1922 (3d Cir. Apr. 6, 2026), 2-1 (Porter, J., joined by Chagares, C.J.; Roth, J., dissenting). See Holland & Knight. (n.d.). Federal appeals court: CFTC jurisdiction over sports event contracts likely exclusive. <https://www.hklaw.com>. The Ninth Circuit heard consolidated argument in parallel cases on April 16, 2026.

-
28. Kalshi imposes position limits and accountability rules under its rulebook (provisions historically numbered Rule 5.14; numbering varies by rulebook version), per CFTC org-rules filings. <https://www.cftc.gov>.
 29. U.S. Food and Drug Administration. (n.d.). FY 2025 PDUFA performance report. <https://www.fda.gov/media>. The agency met or exceeded the large majority of its review-performance goals.
 30. Mitchell, J. M., Bogenschutz, M., Lilienstein, A., Harrison, C., Kleiman, S., Parker-Guilbert, K., ... Doblin, R. (2021). MDMA-assisted therapy for severe PTSD: A randomized, double-blind, placebo-controlled Phase 3 study [MAPP1]. *Nature Medicine*, 27, 1025–1033; and the confirmatory MAPP2 trial, *Nature Medicine* (2023). See MAPS PBC announcements, <https://www.mapsbcorp.com>. Both trials met the primary (CAPS-5) and key secondary (Sheehan Disability Scale) endpoints.
 31. FDA Psychopharmacologic Drugs Advisory Committee vote, June 4, 2024 (2-for/9-against on effectiveness; 1-for/10-against on benefit–risk), per *Pharmaceutical Executive*, <https://www.pharmexec.com>. Complete Response Letter issued August 9, 2024, per *Fierce Biotech*, <https://www.fiercebiotech.com>; and BioSpace.
 32. Dezzani, L. (n.d.). [Public post on the future of medical affairs and scientific prediction markets]. LinkedIn. <https://www.linkedin.com/in/lucadezzani>. Luca Dezzani is Vice President of US Medical Affairs, BioNTech. The remarks are his own forward-looking views, expressed publicly rather than in an interview for this report, and are reproduced here in paraphrase with short quoted phrases.
 33. MLB–CFTC memorandum of understanding announced March 19, 2026. See MLB.com; Sports Video Group. (n.d.). MLB names Polymarket exclusive prediction market exchange partner and signs agreement with CFTC to establish integrity framework. <https://www.sportsvideo.org>; and DeFi Rate, <https://www.defirate.com>. Sportradar serves as MLB's data distributor.
 34. UMA. (n.d.). Improving oracle efficiency with managed proposers. <https://blog.uma.xyz>. The post states an optimistic-oracle dispute rate of approximately 1.3%; the \$750 USDC proposer bond is documented in Polymarket's developer and help-center documentation. On the March 2025 governance incident, see Orochi Network. (n.d.). Oracle manipulation in Polymarket 2025. <https://www.orochi.network>.
-