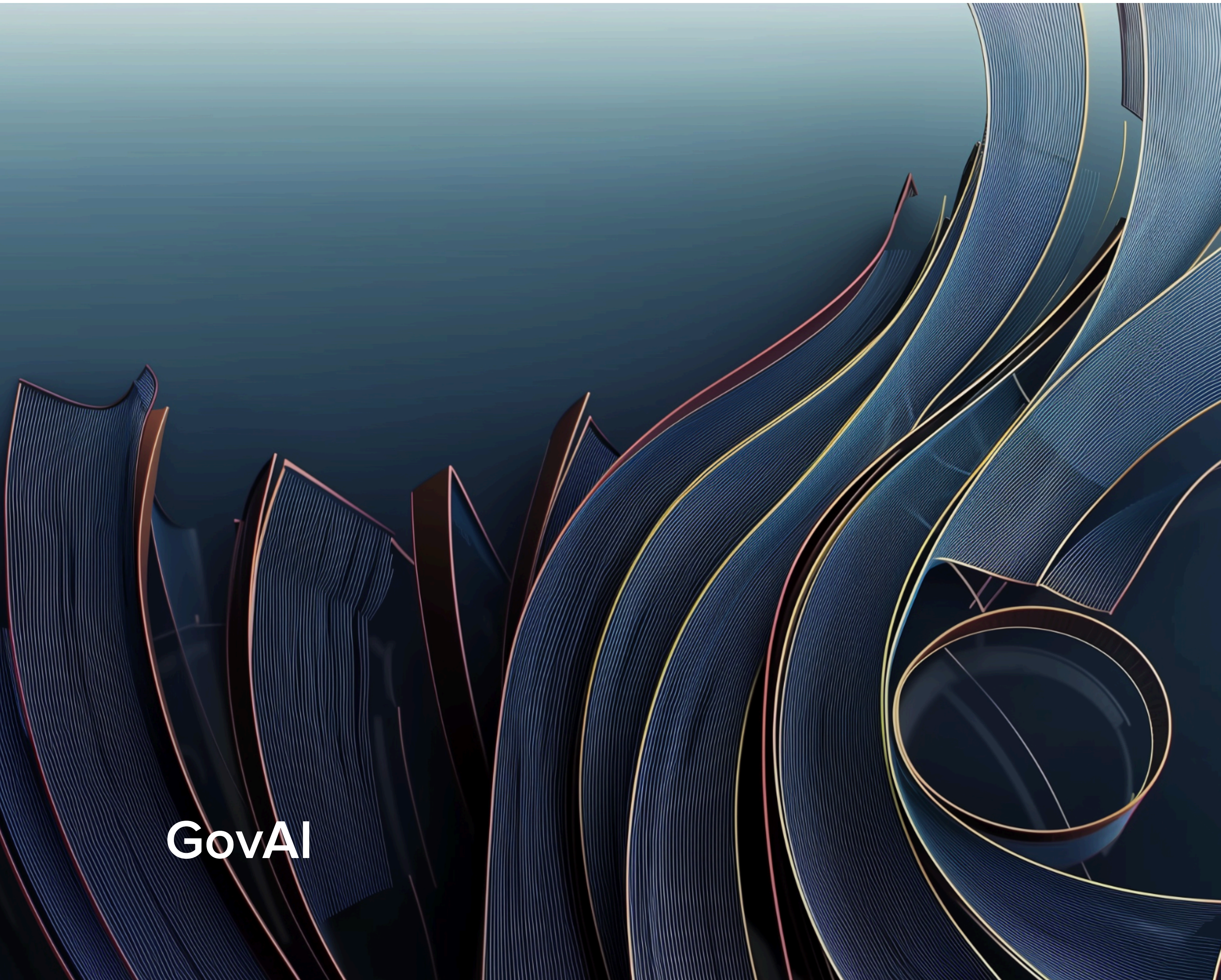
Dual-Use AI Capabilities and the Risk of Bioterrorism

Converting Capability Evaluations to Risk Assessments

Luca Righetti



GovAI

Authors and affiliations

Luca Righetti

GovAI

Contact

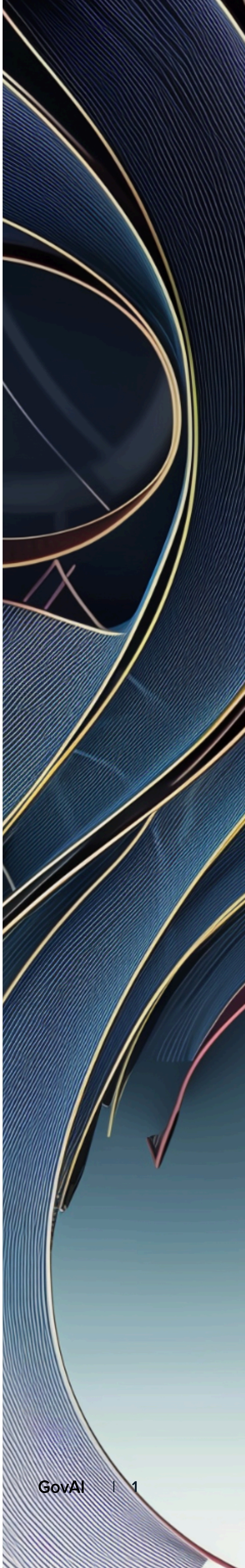
Corresponding author: Luca Righetti (luca.righetti@governance.ai).

Please cite as

Righetti, L. (2025). *Dual-Use AI Capabilities and the Risk of Bioterrorism: Converting Capability Evaluations to Risk Assessments*. GovAI.

GovAI Reports, including this one, go through a double-blind peer review process. Any remaining errors are the authors.

This work represents the views of its authors, rather than the views of the organisation, nor does it constitute legal advice.



Acknowledgements

This paper benefited greatly from a large number of discussions and interviews over the last year. My thanks includes to Gary Ackerman, Alex Anwyl-Irvine, Tessa Alexanian, Chris Bakerlee, Damon Binder, Sonia Ben Ouagrham-Gormley, James Black, Roger Brent, Christina Brunelle, Rocco Casagrande, Diarmuid Cassidy, Seth Donoughe, Kevin Esvelt, Anjali Gopal, Ella Guest, Ezra Karger, Holden Karnofsky, Alex Kleinman, Jens H. Kuhn, Joe Torres, David Manheim, Greg Lewis, Alyssa Mathiowetz, Joshua Monrad, David Moore, Richard Moulange, Cassidy Nelson, Jassi Pannu, Claire Qureshi, Jonas Sandbrink, Yo Shavit, Benjamin Stewart, Anna Wang.

I would like to especially thank Rose Hadshar and Bhuvana Sudarshan for some of the analysis that this report draws on, to Greg Lewis for in-depth comments, to Bridget Williams and the Forecasting Research Institute for helping to run a forecasting survey, to Ozzie Gooen and the the Quantified Uncertainty Research Institute for developing a web application, as well as to participants of the surveys themselves. I thank Markus Anderljung, Stephen Clare, Connor Aidan Stewart, John Halstead, and Matthew van der Merwe at GovAI for edits.

I would also like to thank Jassi Pannu and Coleman Breen for their very helpful comments on the paper during the double-blind peer review process.

Providing review and advice does not imply endorsement of the paper or its findings. All remaining errors are my own.

About GovAI

GovAI is a 501c(3) non-profit organisation. Our mission is to help decision-makers navigate the transition to a world with advanced AI, by producing rigorous research and fostering talent. Researchers at GovAI work on a wide range of topics, with a particular emphasis on the security implications of frontier AI.

Abstract

Several frontier AI companies test their AI systems for dual-use biological capabilities that might be misused by threat actors. But what do these test results imply about the overall risk of bioterrorist attacks? There is much expert debate about how seriously to view such threats, especially from lone wolf actors. This report creates a framework for how to convert capability evaluations into risk assessments, using a simple model that draws on historical case studies, expert elicitation, and reference class forecasting. I conclude that if AI systems were to increase the number of STEM Bachelors able to synthesise pathogens as complex as influenza by 10 percentage points and also enable them to design concerning operational attack plans, then the annual probability of an epidemic caused by a lone wolf attack might increase from 0.15% to 1.0%. This is equivalent to 12,000 additional expected deaths per year, or ~$100B. Risk scenarios where AI or other tools also help discover novel viruses reach higher damages, whereas risk can also be significantly lowered if mitigations are put in place. A review of this report by six subject-matter experts and five superforecasters found similar medians, though all forecasts had high uncertainty. This work demonstrates a methodological approach for converting capability evaluations into risk assessments, whilst highlighting the continued need for better underlying evidence and expert discussion to refine assumptions.

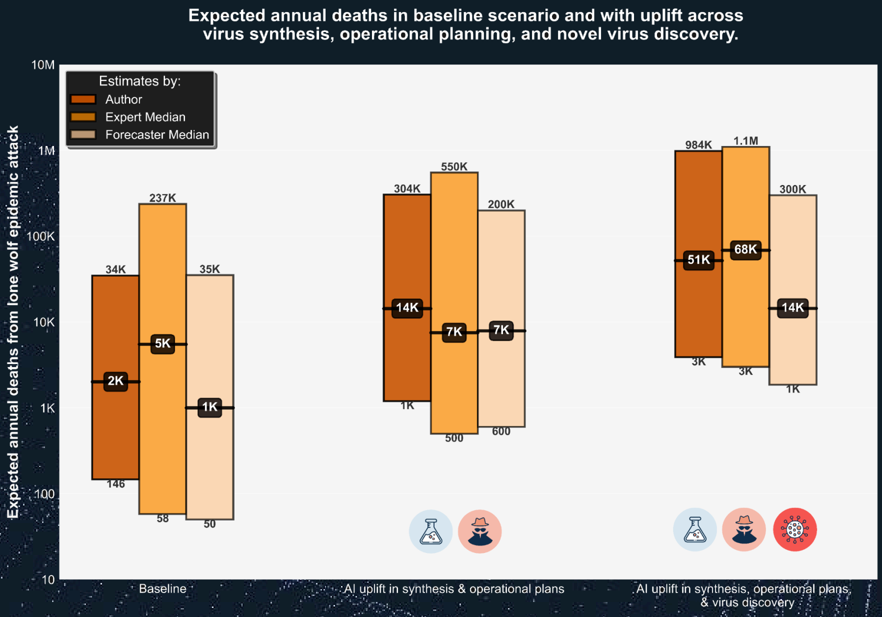


Figure ES I Author and other estimates of the ex-ante annual deaths caused by lone wolf epidemic attacks. Reviewers incl. 6 subject-matter experts (E) and 5 superforecasters (F). For the author the box shows the 5th, 50th, and 95th percentiles. For reviewers, it shows the median response to the 5th, 50th, and 95th percentiles.

Executive summary

- Some experts warn that certain dual-use AI capabilities could assist terrorists in carrying out biological attacks, prompting several AI companies to test their models on biological tasks of concern. AI performance has been found to be improving markedly across various such biological benchmarks. But what these results mean for actual societal risk remains unclear.

- To inform high-stakes policy decisions, such capability evaluation results need to be converted into risk assessments. A qualitative approach can add details and ground evidence in national security expertise. A quantitative approach can help better reflect the overall level of uncertainty and highlight core premises where different experts disagree. Together, these approaches can help decision-makers decide if capability evaluations results provide cause for concern – and better weigh the costs against the benefits of implementing further safeguards to their AI systems before they are commercially deployed.

- This report addresses this problem in the context of a specific risk scenario: "lone wolf epidemic terrorism", referring to the threat of individual actors or very small groups engineering viruses to cause an epidemic. This is just one narrow AI biological misuse pathway and other threat models, such as anthrax and state biological misuse, should also be considered that require separate analysis.

- This report identifies three key technical barriers that currently make such lone wolf attacks very unlikely, and proposes concrete AI capability thresholds that could indicate if future AI systems were to erode these barriers:

    - **Virus Discovery:** Future AI systems may help identify epidemic-potential pathogens, either by discovering novel dual-use information or proliferating existing sensitive information;

    - **AI Lab Coach:** Future AI systems may teach specialised skills needed to synthesise viruses, including detailed scientific protocols and troubleshooting laboratory experiments;

    - **AI Ops Coach:** Future AI systems may help in designing and executing complex operational attack plans, including circumventing current defences like DNA synthesis screening or avoiding detection by law enforcement.

- This report's methodology combines multiple sources of evidence to estimate how these AI capability thresholds might affect societal risk. This includes analysing bioterrorism case studies, constructing reference classes in other domains for events that lack historical precedence, analysing data from a forecasting survey, and conducting interviews with a range of biology and biosecurity experts.

- This evidence is integrated into a six-parameter model that breaks bioterrorism into discrete stages: from the number of relevant actors who have the resources and intent, through technical and operational success rates, to the likelihood of escalating into an epidemic and how many might die from such an event (Table ES).

- I estimate that the current evidence suggests that if future AI systems cross the 'AI Lab & Ops Coach' thresholds[1], then, absent further mitigations, the likelihood of an epidemic lone wolf attack increases from 0.15%/yr [0.02% – 1.4%, 90th percentile range] to 1.0%/yr [0.15% – 13%]. This corresponds to going from 2,000 expected deaths/yr [146– 35,000] to 14,000 [1,200 – 305,000]. Including Virus Discovery could further increase this to 50,000 deaths [3,000 – 984,000]

- Any single assessment relies heavily on subjective judgement, so this analysis was reviewed by six experts and five superforecasters, who provided their own estimates for the model parameters and overall judgements. The survey's median and 90% credible interval mostly matched the author's, if somewhat lower for some outputs. The likelihood of an epidemic lone wolf attack increases from 0.1%/yr [0.01% – 1%] to 0.6%/yr [0.1% – 3.3%] – or 2,000 expected deaths/yr [100 – 68,000] to 8,000 [550 – 250,000] (Figure ES).

- A copy of this simple model is available at https://biocalc.vercel.app/ where readers can explore their own assumptions and see the impact on the results.

- These findings suggest that the 'AI Lab & Ops Coach' would meet the definition of 'severe harm' in OpenAI's Preparedness Framework of $100B, as well as move lone wolf epidemic risk into a more concerning category per frameworks like the UK's National Risk Register. Crossing such a capability threshold would likely necessitate some targeted mitigations – and more damaging scenarios additional mitigations.

[1] These thresholds are operationalised as randomised control trial finding that (1) an additional 10% of STEM Bachelor are able to synthesise influenza pathogens, and (2) AI systems help to create "satisfactory" bioterrorist attack plans, as judged by an expert panel similar to Mouton et al. (2023).

- Overall, this report shows both the potential value of converting AI capability evaluations into risk assessments and the inherent limitations that any such analysis must wrestle with, including estimating low-probability events and assessing unprecedented risks. High-stakes decisions about AI development will likely need to be made under irreducible uncertainty. But this report proposes one method that can be used to better navigate these challenges.

	MODEL 'BASELINE' ESTIMATES			**'AI LAB & OPS COACH' ESTIMATES**		
	Type Of Individuals					
PARAMETERS	WLB PhDs	STEM BSc.s	Other	WLB PhDs	STEM BSc.s	Other
A: Number of individuals of this type	~150K	~20M	~200M	~150K	~20M	~200M
L: % could synthesise virus in lab-setting	~20%	~1%	~0.1%	~40%	~11%	~1%
O: % would still be caught or stopped	~67%	~33%	~33%	~67%	~33%	~33%
	Number of individuals who could engineer viruses					
	89K actors [23K – 405K]			**981K** [206K – 4.87M]		
R: % would try to make a virus that year	~0.3/1M	~0.03/1M	~0.01/M	~0.3/1M	~0.06/1M	~0.02/M
E: % attack "takes off" into an epidemic	~20% [10%–40%]			~20% [10%–40%]		
	Likelihood of an epidemic from lone wolf attack					
	0.15%/yr [0.02%-1.40%]			**1.05%/yr** [0.15%-12.75%]		
D: Potential deaths if a "take off" occurs	2.5M [0.1M–10M]			2.5M [0.1M–10M]		
	Ex ante annual damages from lone wolf attack					
	2K deaths/yr [146–35K]			**14K deaths / yr** [1K – 305K]		

Table ES | Author parameter estimate of the lone wolf epidemic terrorism threat model. WLB PhDs refers to wet lab biology PhDs and STEM BSc.s to people with a bachelor's degree in Science, Technology, Engineering, or Mathematics. The table shows the 'baseline' values and how these change if the 'AI Lab & Ops Coach' thresholds are crossed. The key parameters changed in the latter scenario are highlighted in orange, and calculations are in blue. Brackets show the 90% credible intervals from Monte Carlo simulations.

Contents

Introduction

Motivation

Historically, bioterrorism attacks have been very rare (Tin et al., 2022) and the current level of risk is still debated (Koblentz & Kiesel, 2021). Some experts think that future AI systems might make such attacks more likely by allowing more terrorists to create biological weapons, whilst others see such discussion as speculative (Rose et al., 2024; Peppin et al., 2024). Such debates have become more urgent as , AI performance on benchmarks that measure competency in biological science is improving quickly (IAISR, 2025 [p81]; Justen, 2025; Figure 1.1a), and several companies have warned that their "models are on the cusp of being able to meaningfully help novices create known biological threats" (OpenAI, 2025a).

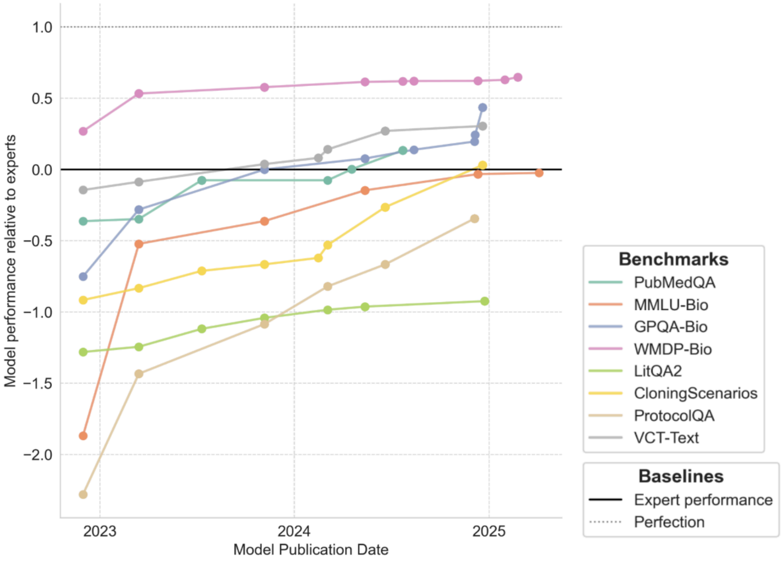


Figure 1.1a | AI system performance on eight biology benchmarks normalised to expert baseline.
Justen, 2025

National security concerns have prompted many AI developers to commit to evaluating the biological capabilities of their models (The White House, 2023, 2025). This includes testing whether AI systems can help plan bioterrorist attacks (Mouton et al., 2023) and help people perform virology tasks more than existing tools like the internet (FMF, 2024; Anthropic, 2025 [p24]; Google DeepMind, 2025 [p10]; OpenAI, 2025b [p17]).

However, evaluating these capabilities is only part of the challenge. We also want to know what the results of these tests imply about the risks an AI system ultimately poses. *If* a study finds that an AI system enables 10% more STEM Bachelors to synthesise an influenza virus, *then* how much more likely is a major epidemic outbreak to occur? Answering this question requires understanding current barriers to bioterrorism and how they might be lowered by AI systems.

To date, there has only been limited attempts at mapping AI capability results onto risk estimates (OpenAI, 2024a; Paris AI Action Summit, 2025; Murray et al., 2025). This report seeks to better inform how future AI capabilities might affect biological risk, characterise how uncertain experts are about these risks, and outline what assumptions might drive their overall disagreements. This can then help inform high-stakes decisions, something this report does not directly do. See Figure 1.1b.

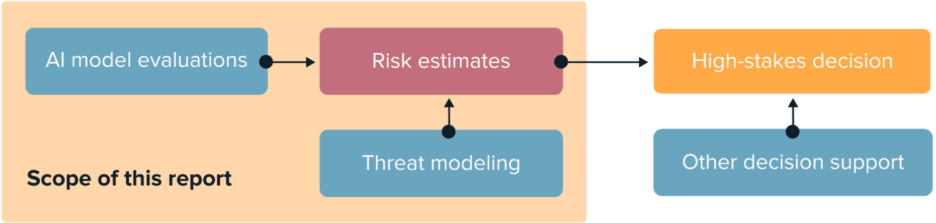


Figure 1.1b | Risk estimates as one potential tool for translating AI model evaluations into high-stakes decisions

This report is structured as follows:

- Section 2 outlines the "lone wolf epidemic terrorism" threat scenario and identifies three AI capabilities that could reduce current barriers to carrying out such attacks.
- Section 3 develops a simple risk model and then synthesises different evidence to estimate baseline risk and that from potential future AI-enhanced scenarios.
- Section 4 describes how these estimates have been validated through expert review by biosecurity specialists and highly-credentialed forecasters.
- Section 5 discusses policy implications and future research directions.

Prior Work

Capability evaluations can be mapped onto risk estimates using quantitative risk assessments informed by judgemental forecasting. In applying these methods to a domain as uncertain and complex as bioterrorism, it is important to understand prior work and methodological limitations inherent in the field.

Quantitative Risk Assessments

Quantitative risk assessments assign numerical values to the impact and likelihood of a threat scenario. Whilst developed for engineered systems, such assessments have been applied to other fields, including high-stakes AI development decisions (NIST, 2025a [p31]; Koessler et al., 2024), combatting terrorism (NASEM, 2008a [p11]; Ezell et al., 2010), and risk management more broadly (Apostolakis. 2004).

Relevant applications of quantitative risk assessments in the context of terrorism or biosecurity include passenger aircraft bombs (Stewart & Mueller, 2018), improvised explosive devices (Grant & Stewart, 2012), nuclear terrorism (NASEM, 2023; Bunn, 2006; Mueller, 2009; Baum et al., 2018), gain-of-function research (Gryphon Scientific, 2015; Lipsitch & Ingelsby, 2014), and avian influenza outbreaks (Fischhoff et al., 2006).

Experts have extensively discussed whether it is appropriate to apply quantitative risk assessments to terrorism risks (NASEM, 2008a; Aven & Renn, 2009; Ezell et al., 2010; Brown & Cox Jr; 2011; JASON, 2009). The literature highlights that such assessments can be improved by including qualitative discussion of the threat, keeping models simple to improve transparency, and ensuring estimates reflect a representative range of subject-matter expert opinions. Because terrorist attacks are relatively rare, many note that risk analyses in this domain should be “semi-quantitative” and incorporate qualitative considerations (Aven & Renn, 2009).

Judgemental Forecasting

Judgemental forecasts involve deriving probability estimates from subjective opinions and holistic assessments of different evidence (Goodwin & Wright, 2009). These can be a major input into quantitative risk assessments. They have shown promise in domains lacking objective reference classes (Tetlock et al., 2017). Since data on terrorism is relatively sparse, expert judgements have been described as “the main source for reducing epistemic uncertainty” (Aven & Renn, 2009).

Several past exercises have asked subject-matter experts to predict the likelihood of chemical, biological, radiological, and nuclear (CBRN) terrorist attacks. Baxter et al. (2024) provides a literature review, starting with the Lugar survey in which experts estimated a 50% likelihood of a CBRN attack over the next five years (Lugar, 2005). Some forecasting markets also contain related predictions about bioterrorism (Metaculus, 2016; Good Judgement, 2024). A notable benefit of forecasts is that they are clearer, when much of the terrorism literature has been described as “rather vague and imprecise predictions” (Bakker, 2012).

However, the accuracy of estimates depends heavily on the quality of forecasts. Notably, experts can reach vastly different conclusions, even when using similar models: see contrasting takes by Bunn (2006) and Mueller (2009) on nuclear terrorism; or Fouchier (2015) and Klotz (2015) on lab accidents. Koblentz (2011) notes how forecasts of CBRN terrorism can be affected by cognitive biases. JASON (2007) notes how predictions of rare events like bioterrorism do not gain much signal from being validated against the past, creating limitations by which to assess accuracy

The literature highlights that such assessments can be improved by accompanying forecasts with a clear explanation of the process used (Narayanan & Kapoor, 2024), regularly updating estimates when new evidence emerges, and reflecting a range of subject-matter expert opinions on any such questions.

Scenarios: Lone Wolf Epidemic Terrorism

Scoping Threat Actors And Misuse Vector

There are many possible threat scenarios for AI biological misuse. This report focuses on one scenario in particular: "lone wolf epidemic terrorism". These terms are defined in this report as per Table 2.1a. The report's limited scope is purposeful to allow for more depth and rigor (Kapoor et al., 2024).

Key Definitions

Lone Wolf Terrorist – A "lone wolf" in this report is defined as an individual or very small group of up to five people who aim to cause harm and are not acting under the direction of a terrorist organisation or a state (although they may share their ideology). Examples include Bruce Ivins (suspected of the Amerithrax attacks), Ted Kaczynski (the 'Unabomber') and Omar Mateen (who was motivated by Jihadi terrorism, but was not *directly instructed* and had no formal links with ISIS). Examples would not include Aum Shinrikyo, who had more than five members working to build weapons of mass destruction.

Epidemic – An "epidemic" is an outbreak of a disease that spreads across a large geographic area and affects a significant proportion of the population. For simplicity, this report defines it as causing >10,000 deaths in excess mortality within a 3-year period via a transmissible pathogen. Examples would include COVID-19 and 2009 Swine Flu. It would *not* include 2001 Anthrax Attacks.

Table 2.1a | Key definitions of the risk scenario in this report.

Table 2.1b contextualises this risk scenario with a larger taxonomy of these scenarios, drawing on the criteria in NIST (2025b). Other AI biological misuse scenarios, such as how AI systems might affect rogue states' bioweapon programmes or how well-resourced groups might try to weaponise anthrax, also warrant investigation and consideration by decision-makers (Lentzos et al., 2024).

This report prioritises epidemic terrorism in part because epidemic pathogens could cause catastrophic outcomes, such as over 100,000 deaths, than most other possible threats. This report prioritizes lone wolves in part because they are currently amongst the least able actors to succeed today and thus there is a larger space for which future AI can have a counterfactual impact. See Appendix 1.1 for details.

Threat Actor	Biological Agent	Method of Acquisition	Route To Harm
Non-expert individual(s) – 1-5 people, none with more than an undergraduate degree and a basic at-home setup. Budget of $10K–$1M.	**Epidemic pathogens** – e.g. Esvelt's (2022) worry about smallpox, 1918 influenza, or potentially novel candidates	**Self-manufacture pathogens** – e.g. doing a reverse genetics protocol (WHO, 2015;) OR taking a wild-type strain and inducing mutations (Lipsitch, 2018)	**Deliberate misuse** – e.g. Rhodesia killed hundreds of African nationalists in the 1970s via its CBW programme; Bruce Ivins is suspected of sending anthrax letters in 2001 (Carus, 2017 [p41-43])
Highly skilled individual(s) – 1-5 people with PhDs in a relevant domain and potential access to a university facility. Budget of $10K–$1M.	**Weaponised bacteria**[2] – e.g. Johari's (2002) worry of a plane dispersing 100kg of aerosolised anthrax	**Sourced from existing suppliers** – e.g. tricking, bribing, or coercing actors that supply such pathogens without needing to apply further changes (Soice et al., 2023)	**Accidental release** – e.g. Sverdlovsk 1979 anthrax leak (Meselson et al., 1994); suspected UK 2007 foot & mouth (DEFRA, 2008); potentially the 1977 Russian Flu (Rozo & Gronvall, 2015)
Somewhat capable group –up to tens of relevant PhDs and a moderately sophisticated facility. Budget of$1M-–$100M.[3]	**Non-WMD bacteria, virus or toxins** – e.g. Jihadi terrorist produced ricin to put into a small explosive (Flade, 2018)		
Moderately capable group – more than tens of relevant PhDs with purpose-built facilities, plausibly including state affiliation. Budget of $10M–$1B.[4]	**Anti-crop agents** – e.g. Agent Orange or stockpiled rice blast to destroy or hinder plant growth (Christopher et al., 1997)	**Taken from natural sources** – e.g. sampling novel virus strains from high-risk wildlife populations like bats without needing to apply further specific changes (Piper, 2022)	**Coercive threat** – e.g. a Chechen separatist leader threatened to acquire biological weapons from a Soviet lab unless Russia released political prisoners (Carus, 2001 [p107])
Highly capable group – A world-class team from industry and academia with state-of-the-art facilities, likely including state affiliation. Budget of >$100M–$10B.	**Novel Global Catastrophic Risk** – e.g. "mirror life" (Adamala et al., 2024)	**Theft** – e.g. stealing a pathogen from a laboratory that is known to store it (Defense Science Board, 2009)	**Other** – More generally, using possession of WMDs to enact harmful pressure without deploying them

Table 2.1b | Taxonomy of AI-biological threat models. Orange is in scope for this report.

[2] 'Weaponised' refers to doing additional steps in the manufacturing process to make it more lethal or infectious.
[3] Aum Shinrikyo is estimated to have spent ~$60M on weapons of mass destruction out of its $0.5–$2B net-worth (Simons, 2006) [inflation adjusted]. Al-Qaeda is estimated to have an annual income of $30M and spent $900K on the 9/11 plot (Roth et al., 2004 [p3-4]) [inflation adjusted].
[4] South Africa and Iraq are estimated to have spent $30M and $80M on their bioweapons programmes. The USSR and USA are estimated to have spent $35B and $700M (Ouagrham-Gormley, 2014).

Given the report's focus on less-resourced actors, much of the analysis focuses on the proliferation of known skills (existing capabilities accessible to larger numbers of individuals) rather than novel risk (increased ceiling of capabilities of the most skilled actors) (Sandbrink, 2023) (Figure 2.1). Thus, any resulting estimate might be better seen as a lower-bound for the total biological risk posed by future AI systems that reach a certain capability threshold.

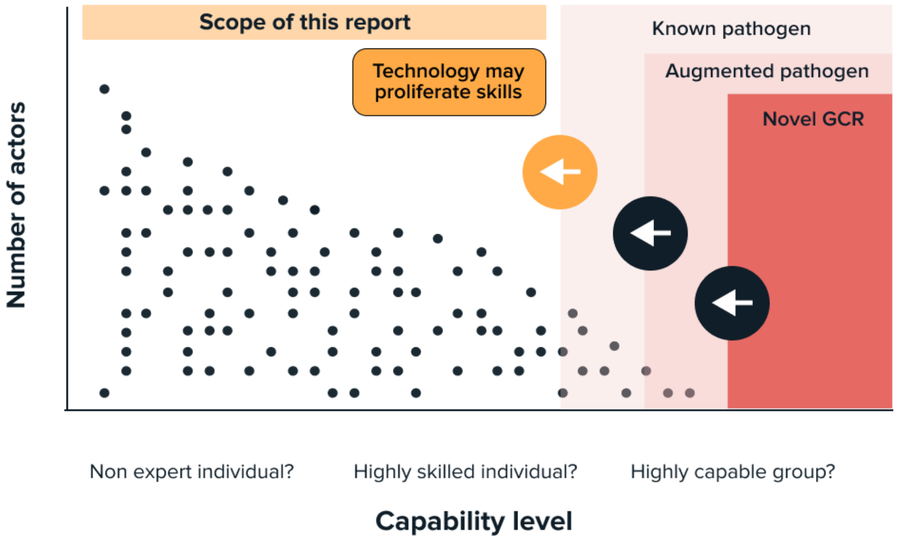


Figure 2.1 | Example illustration of threat actor capability. Orange is in scope. Adapted from SecureBio. Quantities and numbers are purely illustrative.

Potential Technical "Hard Steps" For Bioterrorism

The biosecurity field has discussed the concern of epidemic terrorism long before recent AI progress (Koblentz & Kiesel, 2021). The most common explanation for why such terrorism has not occurred is that it is technically difficult (Revill & Jefferson, 2014) and, thus in part, rarely attempted (Ackerman & Pinson, 2013).

To understand why many believe this, it helps to describe how viruses are currently engineered. The process of building a known pathogen using reverse genetics technology is typically divided into four stages (WHO, 2015). A fifth stage can be added to make viruses more dangerous. As per Figure 2.2a:

1. **Genome Sequence:** know what "code" the desired virus strain is made of.
2. **Acquiring Materials:** needs specific items, e.g. small DNA fragments of that code.
3. **Assembling DNA:** needs to "stitch" fragments together into a correct construct.
4. **Introducing to Cell:** A person needs to "boot up" the construct into a replicating virus.
5. **[Opt.] Mutating:** E.g. infect across hosts, letting natural selection create desired traits.

Anyone trying to build a biological weapon themselves may have to do multiple "design-build-test" loops of these stages (NASEM, 2018). There is additional difficulty when trying to make a dangerous pathogen that risks attracting the attention of law enforcement and requires obfuscation. Creating entirely novel designs creates further challenges still (Montague, 2023).

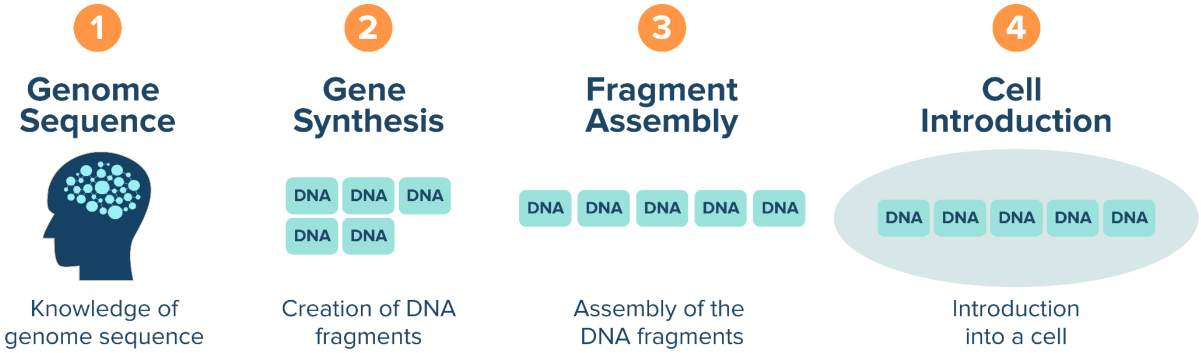


Figure 2.2a | Illustration of the process to engineer a known virus. Adapted from WHO (2015) [p21]. An optional fifth step of Mutating is not shown in this source.

Based on these stages, this report identifies three "hard steps" per Table 2.2a and Table 2.2b, which we can then consider how future AI systems may affect.

AI-bioterrorism uplift	Description of the "Hard Step"	Bio Stages
Virus Discovery	Identifying a pandemic potential pathogen with a high reliability of causing an outbreak is hard. Discovering a novel strain would require major scientific work. Some worry that future AI could help people identify such candidates, either by bringing together sensitive results in existing papers or discovering a novel mutation, such as by using biological design tools (see Thadani et al., 2023).	1 (and potentially some of 5)
AI Ops Coach	Obtaining many materials and operating for several months without detection is non-trivial – especially needing to get DNA materials which some companies screen for. Some worry that future AI could describe a detailed operational plan, such as technical assistance to 'camouflage' DNA orders (see Mouton et al., 2024).	2
AI Lab Coach	Building pathogens requires many virus-specific skills often described as "tacit knowledge" acquired via hands-on-experience and mentorship. Some worry that future AI could provide detailed and sensitive protocols, troubleshoot lab experiments, and otherwise help acquire wet lab skills (see Götting et al., 2025).	3,4 (and potentially some of 5)

Table 2.2a | Summary overview "hard steps" for lone wolf epidemic terrorism. "Bio stages" refers to the stages discussed on page 13 as well as Table 2.2b.

	Technical "Hard Step"				
Stage	**1. Genome Sequence** (and partly stage 5)	**2. Acquire Materials**	**3. Assemble DNA**	**4. Intro to Cell**	**5. Mutating** (optionale)
Current Difficulty	"Hard" Step [Medium Agreement; Limited Evidence]	"Medium" Step [High Agreement; Medium Evidence]	"Hard" Step [Medium Agreement; Limited Evidence]		
Explanation	Experts disagree on how likely pathogens in the current gain-of-function literature are to cause an epidemic if released (Gryphon, 2016 [p215]). A threat actor could "roll the dice" and find out. But it's possible such viruses might not work (e.g. people may have immunity to 1918 flu). In such cases, threat actors would either [i] have to find a new virus strain – which even experts would struggle at – or [ii] wait for an external discovery.	Many gene synthesis companies try to detect suspicious orders. However, there is no screening solution for small fragments, and some companies don't screen at all (IGSC, 2024). So, this barrier seems feasible to overcome. Other equipment is less strictly monitored. Setting up a "garage lab" might encounter some challenges, but it seems feasible for a non-expert.	Performing the relevant protocols requires several wet lab skills (NASEM, 2018 [p38]) and is only regularly performed by a small set of experts, most often by people with extensive PhD training. The skills appear hard but not impossible for STEM Bachelors to acquire (NASEM, 2018 [p38]). Sources disagree on exactly how hard it is (e.g. WHO, 2015 [p8] versus Ougrahm-Gorley, 2014). Advances in tech might have made it easier, but many experts still see it as difficult.		
Example Challenges **[Non-Exhaustive]**	**Identify relevant dual-use gain-of-function studies** – e.g. ilt is not obvious which published results are the most concerning, and public reviews try to redact such information (Gryphon, 2016 [p231]). **Find the correct sequence** – Even if a threat actor chooses a specific virus, public sequence databases can have errors, missing information, or not report the full "quasispecies" needed (NASEM, 2018 [p41]). **Spread Testing** – If an existing candidate does not work they may have to do animal-to-animal testing [see right column] and most likely go through multiple "design loops" (NASEM, 2018)	**Bypass screening** – It is not obvious which companies screen outside of IGCS or how to camouflage orders(Edison et al., 2024 [p2]). A non-expert might also have a less plausible cover. **Avoiding detection** – Illegal labs can be detected (Reedley Report, 2023), although this often seems sporadic and smaller "lone wolf" operations may be even harder to notice. **Setting up garage lab** – Creating sterile conditions, like in a professional lab, might be a challenge (DeFrancesco, 2021) – although such precedence does exist (Ledford, 2010).	**Adding details to biological protocols** – Whilst many detailed instructions for pathogens are now publicly available (Pannu et al., 2021), some might not contain *all* the details or could be scattered across different papers (Revill & Jefferson, 2014 [p605]). A non-expert with less context might get confused or stuck as a result. **Help "troubleshooting" lab work** – Even with clear instructions, protocols often need adjustment from the original author's settings. This requires trial and error and method adaptation in response to failure (DeBenedicts, 2023). This can be hard to achieve, especially for non-expert (Ougrahm-Gorley, 2014). **Teach "fine motor" skills to execute protocol** – Several techniques need practice to execute correctly, such as micro pipetting (Mettler-Toledo, 2013), sterile technique (Cell Signalling, 2022), and avoiding infecting oneself. **[Opt.] Conduct animal testing** – A threat actor could optionally have a "ferret-to-ferret" tunnel whereby the virus mutates to have sufficiently increased transmissibility in mammals (Lipsitch, 2018). This can be a tedious process that might not generalise to the real world or human-to-human transmission (Montague, 2023).		
AI-bioterrorism uplift	Virus Discovery	AI Ops Coach	AI Lab Coach		

Table 2.2b | Detailed overview of "hard steps" for lone wolf epidemic terrorism. Light red indicates what is assumed to be a "hard" difficulty step and red "medium". The easier a step, the more concerning it is that a threat actor may succeed at it, hence the darker color.

Across interviews, many experts emphasised that there isn't a single technical "hard step" that would currently block *all* attempts at epidemic terrorism – or that, if they disappeared, would make all attempts certain. Instead, many cited the idea of a "swiss cheese" model of defence.[5] Thus, **this report is interested in AI-bioterrorism uplift scenarios that "stack" capability thresholds on top of each other – and trying to estimate their relative contributions in a more systematic way** (Figure 2.2b).

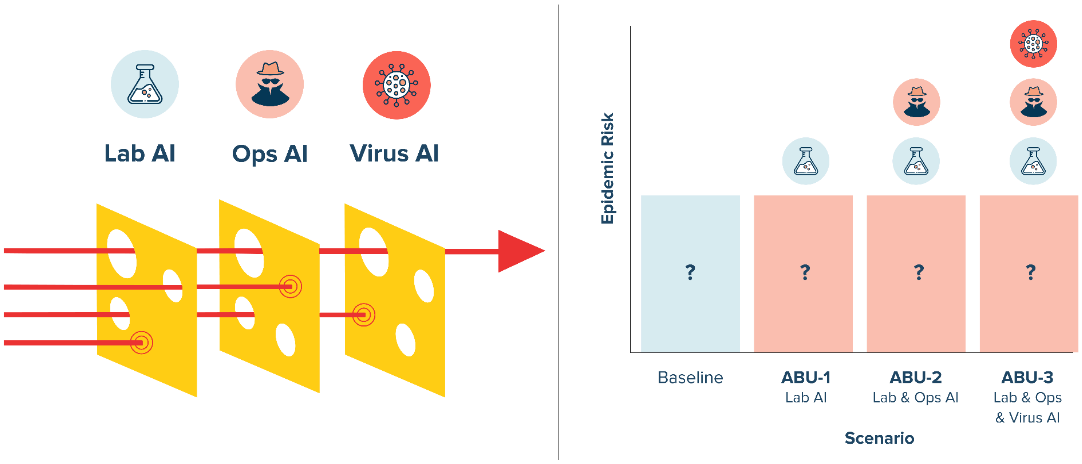


Figure 2.2b | Illustration of "capability thresholds" – and how this corresponds to different AI-bioterrorism uplift scenarios. ABU refers to "AI-bioterrorism uplift"

[5] Removing a single layer can still increase risk; removing a single layer has less of an increase if other defence layers are stronger; removing more than one barrier can have a bigger, multiplicative effect.

Potential AI Capability Thresholds Of Concern

Having identified these technical “hard steps”, **we can now more concretely outline what AI evaluation results might indicate a future AI system could meaningfully assist a threat actor.** This report defines such thresholds via example capability evaluations in Table 2.3 – drawing on expert interviews and reviewing the current literature (FMF, 2024; Pannu et al., 2025).[6]

For simplicity, scenario thresholds here are defined in terms of a single AI capability evaluation, whilst in practice an ensemble of benchmarks, red-teaming exercises, and more can be used to construct scenarios.

	“Capability Threshold” (Describing The Potential Near-Future)			
Stage	**1. Genome Sequence**	**2. Acquiring Materials**	**3. Assembling DNA**	**4. Intro To Cell**
Current Difficulty	“Hard” Step [Medium Agreement; Limited Evidence]	“Medium” Step [High Agreement; Medium Evidence]	“Hard” Step [Medium Agreement; Limited Evidence]	
	if... ⬇	**if... ⬇**	**if... ⬇**	
Capability Threshold	**Virus Discovery** An expert red-team produces a sequence design that the **majority of a panel of judges appointed by the US national security community thinks it is likely to compare to the COVID-19 outbreak** [e.g. >33% chance of R_0>1].	**AI Ops Coach** An RCT finds that frontier AI systems **enables STEM Bachelors to design an epidemic terrorist plan** in two weeks that a panel of judges appointed by the US national security community assigns an **average score of 7/9,** per Mouton et al. (2023)	**AI Lab Coach** An RCT finds that frontier AI systems **enable 10 percentage points more of STEM Bachelors to successfully complete an end-to-end virus rescue protocol** for an influenza strain in a laboratory setting over the course of three months – similar to OpenAI (2024b) and FMF (2025).	
	then... ⬇	**then... ⬇**	**then... ⬇**	
New Scenario Difficulty	“Medium” Step	“Easy” Step	“Medium” Step	

Table 2.3 | “Capability thresholds” for an AI system assisting a lone wolf epidemic terrorist. Light red indicates what is assumed to be a “hard” difficulty step, red “medium”, and dark red “easy”. The easier a step, the more concerning it is that a threat actor may succeed at it, hence the darker color. Note that the previously discussed optional Stage 5 is not included.

[6] See also Justen, 2024; Dev et al., 2025; and International AI Safety Report, 2025 for empirical overviews.

Example Risk Scenario

There are many different ways by which we can imagine the above AI capability thresholds to be met by future AI systems. Importantly, we don't need to prescribe how exactly an AI system might achieve this threshold so as to usefully inform a risk assessment, only that it does have this effect and plays a major counterfactual role relative to other tools.

For example, we could consider an AI assisting with 'virus discovery' either directly, or by being integrated into a Biological Design Tool, or by walking a human through the necessary steps. Similarly, we might imagine an AI assisting as a 'Lab Coach' via very detailed interactive Augmented Reality headsets, or detailed but still text-based protocols.

However, to make an AI capability threshold more concrete and credible to envision, we can speculate about how current AI trends might further develop to hit these thresholds. For example, Table 2.4 describes how near future AI systems might act as a 'Lab Coach' and provide technical assistance on how to build pathogens.

Illustrative Example Of Current And Potential Near-Future AI System Capabilities		
Current "AI Lab Coach" Assistance		
Help "troubleshoot" lab work: Given an image from a laboratory experiment, AIs can often assess the results and determine what went wrong (Götting et al., 2025).	**Adding *simple* details to protocols:** Provided with a protocol, an AI can explain things to laypeople, warn of common failure points, and suggest alternative routes (Gopal et al., 2023).	**Adding *hard* details to protocols:** AIs can often answer questions that are normally believed to "require tracking down authors of relevant papers" (OpenAI, 2024 [p21]).
Near-Future "AI Lab Coach" Assistance		
Giving live feedback on videos: Future AIs could be given a live camera feed of users handling equipment and provide real time advice, similar to prototypes for spotting issues at construction sites (Mollick, 2024).	**Teach "fine motor" skills:** Future AIs could help people by suggesting easier exercises to practice, then tailoring feedback and increasing the difficulty in a tailored manner.	**Agent tooling to automate tasks:** Future AIs could automate some relevant research (Boiko et al., 2023), such as working out what DNA fragments and enzymes are needed for certain experiments (FutureHouse, 2024).

Table 2.4 | Examples of present and potential future AIs teaching biology skills

Simple Model

Methodology

This section develops a simple quantitative risk model to estimate lone wolf bioterrorism risk, building upon frameworks such as JASON (2009), NASEM (2018), and Sandberg and Nelson (2020). As noted in Section 1.2, there are strengths and inherent limitations to explicit quantification. The goal of this report is to complement the qualitative literature, which this work heavily draws on, by systematising it in a manner that is simple, transparent, and amenable to reflecting a diverse range of viewpoints.

The approach employed combines multiple types of evidence, namely:

- **Literature Review:** Analysing academic papers, case studies of attempted bioterrorism, and constructing reference classes from other analogous events (Tetlock, 2005);
- **Structured Survey:** This report surveyed 46 subject-matter experts and 22 credentialed 'superforecasters' on questions closely related to each factor (see Williams et al., 2025);
- **Subject-Matter Expert Engagement:** This report draws on in-depth interviews with experts in virology, biosafety, and national security – many of whom commented on iterative drafts.

To systematise this evidence, I construct a six-parameter "ALORED" model of bioterrorism:

Number of **A**ctors × **L**aboratory Success Rate × **O**perational Success Rate × **R**adicalization Rate × **E**scalation Probability × Potential **D**amages = [Baseline] Annual Expected Deaths

Table 3.1 describes the parameters included in the model, how much evidence is currently available, and how the previously defined AI-bioterrorism uplift thresholds map onto this.

An interactive version of the resulting model, built by the Quantified Uncertainty Research Institute, is available at https://biocalc.vercel.app/.

Model Parameter	Description Of Parameter	Evidence Available	AI-bioterrorism uplift
Number Of Actors Who Could Engineer A Virus (A * L* O)			
A: **Number of Actors**	The number of actors who have both the relevant educational degrees (wet lab biology PhD; STEM Bachelor; Other Actors) *and* who have access to the financial resources (~$10K–$100K) to purchase all necessary equipment.	**High** – There is direct data and other proxies	n/a – But could consider AI lowering the financial barrier
P: **Perseverance** *[indirectly in model by informing L]*	The probability that an actor is willing to put in a certain amount of effort (i.e. try again and again even if they fail). The more effort an actor is willing to put in, the more likely they are to succeed in the two parameters: laboratory success and operational success.	**Medium** – No direct data but many proxy approaches	n/a – But could consider AI increasing motivation
L: **Laboratory Success Rate**	The probability that, for the level of effort, an actor would succeed in synthesising a *known* virus if they possess necessary equipment and are not obstructed by law enforcement. This requires overcoming challenges such as troubleshooting, protocols, and fine motor skills.	**Low** – Experts disagree. Uplift trials may help	AI Lab Coach
O: **Operational Success Rate**	The probability that, even though they could succeed in ideal circumstances that might be studied in an RCT , they might not in the real world as this requires bypassing DNA synthesis screening, setting up a functioning lab, avoiding detection, and dispersal.	**Medium** – No direct data. Uplift trials may help	AI Ops Coach
Annual Likelihood Of an Epidemic From Lone Wolves (A * L * O * R * E)			
R: **Radicalization To Pursue Epidemic Harm**	The probability the actor actually has the goal of weaponising a synthesised virus (which is roughly defined as putting in at least one month of effort). Only a very small fraction would pursue such a goal, which we distinguish here.	**Low** – Experts disagree. NatSec may have private data	n/a – But could consider AI making bioterrorism more salient
E: **Escalation Into An Epidemic**	The probability that, conditional on the actor having successfully released their final pathogen(s), the attack causes at least 10,000 deaths. This requires considering how likely a genome sequence is to have epidemic potential and that the actor correctly selects this genome sequence.	**Low** – No public est. but can bound extremes	Virus Discovery
Annual Expected Deaths (A * L * O * R * E * D)			
D: **Potential Damages**	The expected deaths and economic damages from a successful release that "takes off". Whilst a deliberate human release is unprecedented, there is a large literature on the severity of pandemics that can be drawn from.	**High** – No direct info. but a lot on natural pandemics	

Table 3.1 | Overview of the simple model's input and intermediate outputs

Parameters

The following subsections estimate each parameter in the six-parameter model, combining evidence from multiple sources while acknowledging the substantial uncertainties involved. Further details can be found in the Appendix 2, including a summary of the assumed distributions and further qualitative discussion. Throughout this report, square brackets ("[ ]") represent the 90% credible intervals of a given parameter or calculation.

Number Of Actors (A)

The first parameter in the model is the number of actors. Since aptitude varies significantly between actors of different educational backgrounds and experience, this report identifies and separately estimates the global populations of three different relevant groups: wet lab biology PhDs, STEM Bachelors, and Other Actors.[7] Additionally, this first parameter accounts for what fraction of each group have the potential financial resources necessary to attempt bioterrorism as lone wolves.

To estimate the global number of actors in each group, this report first estimates the number of actors for a region where data quality is high (i.e. for the US), and then attempts to extrapolate this quantity to the globe. Table 3.2.1 shows a summary of these calculations and the author's resulting overall assessment. See Appendix 2.2 for details.

This method estimates that there are approximately 10 times more STEM Bachelors than wet lab biology PhDs with at least $10,000 available in financial resources, and 100 times more STEM Bachelors than wet lab biology PhDs overall. This suggests that future risks can be significantly affected by the degree to which the capabilities of non-PhD actors are uplifted by AI, since there are far more potential actors in these groups.

[7] Importantly, education is only one method to classify actors. Alternatives include dividing by 'agenticness' or 'wealth'. This report focused on education given it is more relevant to the extensive discussion of expertise in the literature (e.g. Revill & Jefferson, 2014) and available data. However, other approaches could be useful too.

Sub-Parameter	Description	Wet Lab Biology PhDs	STEM Bachelors	Other Actors
US People With Appropriate Experience Per Cohort	There is high-quality US data on how many people each year receive wet lab biology PhDs or STEM Bachelors (NSF, 2022; NCES, 2024). For Other Actors I simply assume to roughly double the number of all Bachelors, which is ~2M/yr.	2.6K [1K–5K]	435K [350K–0.5M]	4M [3M–5M]
US Effective Number Of Such Cohorts Today	Assuming that the potential threat actor population is approximately equivalent to the working age population (18-64 years olds), there are up to 45 cohorts. I also adjust for wet lab biology PhDs and STEM degrees that were less prevalent in the past or may have taught less relevant skills (e.g. Asimov, 2024).	16X [8X–25X]	25X [20X–30X]	40X [30–45X]
US Fraction With Sufficient Disposable Wealth	Different experts estimate that the relevant equipment needed to build a pathogen may cost between $10K – $100K (DeFrancesco, 2021). Only a fraction of people can afford this (SCF, 2022). I also adjust for the fact that [i] more educated people tend to be wealthier (IPUMS, 2023) and [ii] terrorists tend to be younger and less wealthy (Williams et al., 2018).[8]	0.6X [0.2X–0.85X]		0.3X [0.1X–0.7X]
Generalising To The Rest Of The World	To appropriately scale the US estimates to the globe, I consider multiple references, such as the US share of high-income people (World Inequality, 2022), total R&D (UNESCO, 2021), and number of universities (Förster, 2022).	3.5X [3X–5X]		4X [3-6X]
All Number Of Individuals	*[Calculation: Multiplying all of the values in each column together]*	~100K [20K–200K]	~20M [5M–50M]	~200M [50M–500M]
Overall Estimate	I cross check these estimates with private sources and holistically adjust accordingly.[9]	**~150K** [40K–400K]	**~20M** [10M–40]	**~200M** [100M–400M]

Table 3.2.1 | Overview of "Number of Individuals" estimation methods and author assumptions

Perseverance Rate, P

The perseverance rate parameter *indirectly* informs the model by contributing to the model's second parameter: the laboratory success rates. Assessing how much effort actors are willing

[8] Note that if we already narrowed down the reference population to younger and wealthier individuals, then this affects how likely they are to commit bioterrorism – which may be different from the entire population.
[9] These estimates were later compared against private work by a biosecurity consulting firm, which used alternative data sources, such as the number of published journal articles in different countries.

to dedicate before stopping helps estimate the probability of achieving laboratory success, since an actor who is willing to try more often is more likely to succeed.[10]

To model this dynamic, the report considers a decay function that describes what fraction of threat actors are willing to persevere for different durations of time. For example, it asks "Out of the ~150K wet lab biology PhDs actors – assuming someone does have bioterrorism intent – what fraction are willing to spend at least 1 month of effort? What fraction are willing to spend at least 2 months? And so on."[11]

To estimate this function, the report draws on a mixture of methods to estimate data points. These methods include multiple reference classes and a survey of subject-matter experts and superforecasters. They are summarised in Table 3.2.2a, Figures 3.2.2. See Appendix 2.3 for details.

Different Potential Estimates Of The Perseverance Rate	Value (m = months)
Key Case Studies: I examine a small number of terrorist case studies of bomb plots (e.g. Hamm & Spaaj 2015), as well as bioterrorism and biocriminal activities (e.g. Carrus, 1998). Rose Hadshar constructed a database where she subjectively interpreted how many months it appears that each case study took to plan relying on public reporting. See Figure 3.2.2 – blue, red, and yellow dots.	3m: 46-58% 6m: 37-64% 12m: 5-45% 24m: 5–25%
Terrorist Planning Behaviour: There is a literature on how much preparation previous terrorist attempts underwent, largely based on a dataset by Gill et al. (2014) that encodes behaviours such as what fraction underwent firearms training specifically for the attack. I subjectively interpret how much 'effort' each of the lone wolf behaviours might represent using an AI system, and then draw an approximate "decay function" for how this drops off. Note this data draws overwhelmingly on non-*biological* incidences like mass shootings and explosives. See Figure 3.2.2 – black line.	~50% activity like firearm training ~25% activity like acquiring finances
Judgemental Forecasting Survey: I show the results of my subject-matter expert in biosecurity and superforecasters participants in Williams et al. (2025). We can see that the medians of both samples are broadly similar to the other methods used but do think lone wolf biological terrorists would be somewhat less perseverant, perhaps suggesting that the specific challenge of creating an epidemic weapon could dissuade actors. See Figure 3.2.2.	3m: ~35% 6m: ~15% 12m: ~5% 24m: ~1%
[Sanity Check] Non-Terrorist Datasets: Given that the above methods rely a lot on subjective judgements, I also briefly consider more general reference classes about how much humans persevere at normal activities. There are several academic studies on clinical dieting (Landers & Landers, 2004), new years' resolutions (Oscarsson et al. (2020), and gym membership (Sperandei et al, 2016). These estimates have an obvious external validity issue, but can nonetheless act as a helpful "sanity check" given the more available data. See Figure 3.2.2 – green, orange, and cyan dots.	n/a

Table 3.2.2a | Overview of "Perseverance Rate" estimation methods

[10] Intuitively, the likelihood that a lone wolf succeeds at synthesising a virus differs a lot based on whether they are willing to try for 3 months versus 12 months. For example, Aum Shinrikyo had several failed attempts to create biological agents like anthrax before giving up and making sarin gas instead (Danzig, 2012).

[11] Two years is a stopping point for this analysis, since after two years, the educational category distinctions blur. For example, a STEM Bachelor spending >2 years pursuing wet lab biology skills can just obtain a PhD.

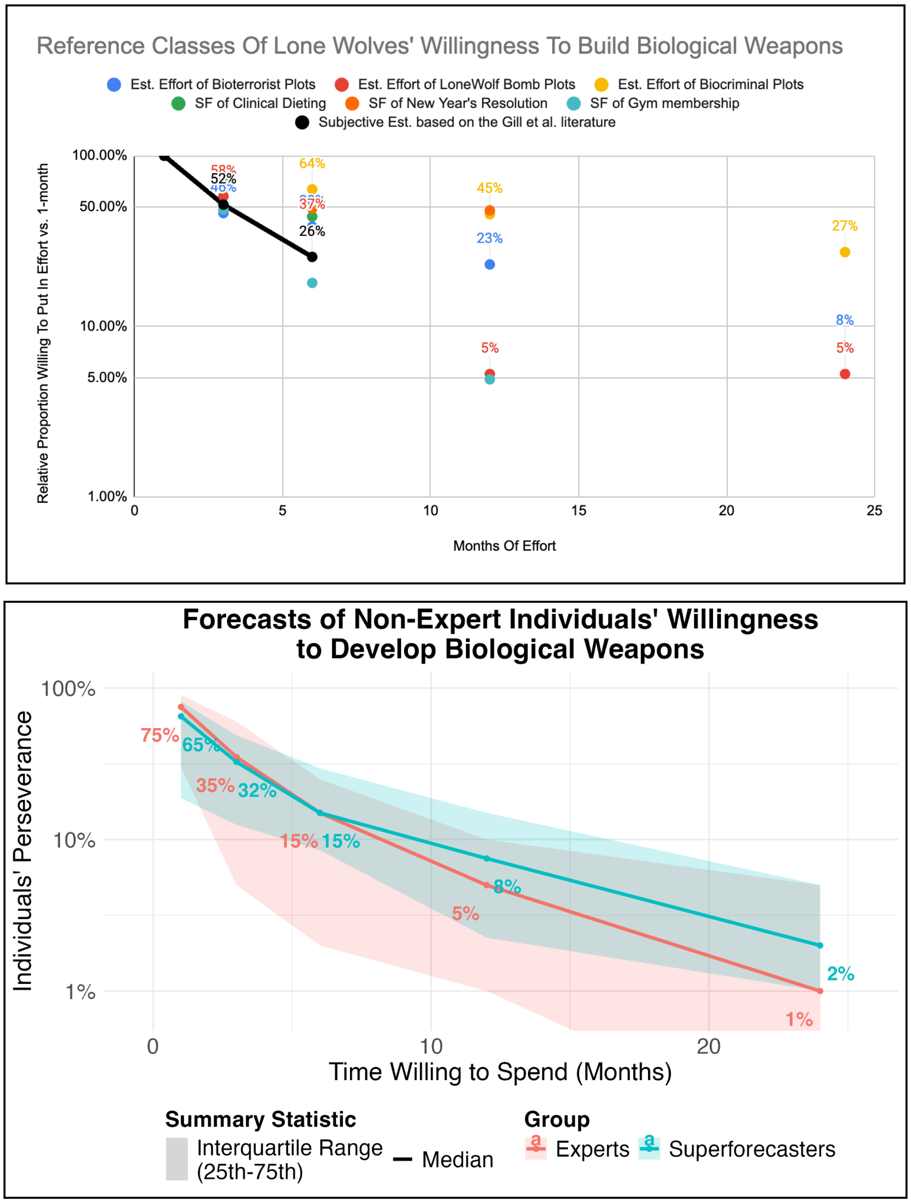


Figure 3.2.2 | Decay Functions of "Perseverance Rate" estimation methods

Table 3.2.2b shows how I combined these different methods into an overall assessment. The table shows that several methods appear broadly consistent with an approximation that roughly half of actors persevere for each doubling in time of effort. I.e., if an actor is willing to put in 1.5 months of effort, there is approximately a 32% chance they persevere for 3 months, approximately a 16% chance of 6 months, approximately 8% chance of 12 months, and approximately 4% chance of 24 months.[12] Still, there is a considerable range of uncertainty across this estimate, especially as we move to a longer time horizon.

Although this perseverance rate is not directly included in the simple model's equation, we use it next to help estimate and interpret the laboratory success rate. A key insight for this application is that it suggests that many potential threat actors are unwilling to spend long periods attempting to develop biological weapons – plausibly because easier attack vectors exist that they could pursue instead (e.g. conventional explosives).

Thus, as will be discussed in the next section, it could be that for future AI to create a large amount of additional "uplift", they may need to create a very large uplift. I.e., to raise, for example, the overall success rate of STEM Bachelors, an AI system would likely need to assist not only highly persistent actors (>12 months) but also those who would otherwise stop after 3–6 months, given that this is where most actual novices would fall.

Parameter	Fraction Willing To Spend At Least X Time (Normalised at 1 month)				
	1 month	**3 months**	**6 months**	**12 months**	**24 months**
Perseverance (All)	1.00 [1.0 – 1.0]	0.32 [0.20–0.60]	0.16 [0.07–0.40]	0.08 [0.02–0.30]	0.04 [0.01–0.20]

Table 3.2.2b | Overview of "Perseverance Rate" author assumptions

Laboratory Success Rate, L

The second parameter of the model is the percentage of actors in each group who, if they put in a given amount of effort, could successfully synthesise a virus. Even under ideal conditions where an actor operates out of a functional laboratory and is not at risk of intervention by law enforcement, synthesis of a virus is difficult and success is unclear – especially for novices. This variable captures the ability to reconstruct a known virus strain that is viable and infectious, but regardless of whether that specific strain goes on to be pandemic credible. The latter consideration will be discussed in Section 3.2.6.

Different virus strains vary greatly in difficulty to synthesise (NASEM, 2018, [p40]). To set a concrete threshold for this parameter, the estimation methods for this variable focuses on influenza. Subject-matter experts note that influenza is easier to assemble than smallpox but still contains several strains of concern, such as 1918 influenza and human-to-human transmissible H5N1 (Lancet, 2024).

[12] Currently this report assumes that all types of actors have the same level 'persistence' (i.e. wet lab biology PhDs, STEM Bachelors, Other Actors all behave similarly). This is somewhat supported by the Williams et al. (2025) showing only minor differences in subject-matter experts' estimates for experts and non-experts. Still, future work may want to investigate differentiating estimates here. For example, it is plausible that wet lab biology PhDs who think that they are more likely to succeed in the first place are willing to try more attempts.

Reviewing the literature, Revill and Jefferson (2014) note “the ease through which individuals can synthesise life remains contested”. Similarly, the experts that I interviewed disagreed strongly on this topic, with beliefs ranging from the position that even a wet lab biology PhD would struggle to replicate known pathogens outside of their specific expertise to the position that most novices could succeed using online resources and courses.

Table 3.2.3a summarises many of the qualitative arguments raised in this debate. Several efforts to summarise these debates concluded that recreating known viruses is difficult but not insurmountable for non-experts. The National Academies (2018) concludes that production “would be achievable by an individual with relatively common cell culture and virus purification skills”. Thus, this report will also follow that overall judgement.

Qualitative Discussion Of Laboratory Success Rate	
Technical And Niche Knowledge	
Case *For* Low Success Rate: Reverse genetics is characterised by considerable trial and error to adapt known experiments to new lab environments (DeBenedicts, 2023). This requires virus-specific technical knowledge such as troubleshooting experiments and knowledge that may not be written down or shared across teams. Experts often cited their own experiences here.	**Case *Against* Low Success Rate:** Viruses like influenza are well documented online, including video tutorials (e.g. JOVE). When virologists were asked to write questions they think are “obscure to anyone not working in the field” or “without hands-on experience”, AI systems already answer these questions well (OpenAI, 2024), suggesting that such information is not so obscure to begin with.
Tacit Skills	
Case *For* Low Success Rate: Reverse genetics can require a lot of skills that necessitate hands-on practice. Examples include developing an intuitive sense for how to “crush the cells with just the right amount of pressure” (Ouagrham-Gormley 2014) or needing “weeks to get the proper [pipetting] technique” (Revill & Jefferson, 2014).	**Case *Against* Low Success Rate:** Many of the tacit knowledge claims use 1980s/90s case studies, but synthetic biology has advanced significantly. Protocols have become simpler (Neumann, 2021), for example e-gel devices purify DNA in “as little as 10 minutes”. Additionally, far more resources now exist to acquire such skills (e.g. NEB; DIY-Bio).
Anecdotal Evidence	
Case *For* Low Success Rate: Some subject-matter expert interviewees emphasised new students need months to execute protocols, even with supervision being provided (Jefferson et al., 2014).	**Case *Against* Low Success Rate:** Some subject-matter expert interviewees emphasised significant variation in talent and skill. So while some non-experts might indeed need years, others may be able to learn quite fast.

Table 3.2.3a | Overview of arguments relevant to the “Laboratory Success Rate”

Turning this overall conclusion into a quantitative parameter is difficult due to the extent of expert disagreement and lack of direct data. Thus, the report draws on a mixture of methods, including previous literature estimates, reference classes, and a survey of subject-matter experts and superforecasters. These are summarised in Table 3.2.3b. See Appendix 2.4 for details.

Different Potential Estimates Of The Laboratory Success Rate	Value
Literature Review: To the best of my knowledge, there has been no public empirical study of how well non-experts perform at wet lab tasks. Rose et al. (2024) [p45] was the only explicit quantitative estimate found, relying on the assessment of 12 independent experts who concluded that novices have a remote probability of 0-5% and highly skilled individuals between 55-75%.	Wet lab biology PhD: 55-75% STEM Bachelors: 0-5%
Judgemental Forecasting Survey: Additionally, I also show the results of my subject-matter expert in biosecurity and superforecaster participants in Williams et al. (2025). See Figure 3.2.3a and b. Though both expert and superforecaster estimates span multiple orders of magnitude, many responses implied wet lab biology PhDs were greater than 10 times more likely to succeed, especially for less than six-months of effort. The parameter is estimated by combining the expert median responses of the perseverance rate and conditional laboratory success rate I.e. summing the fractions of actors via {% fraction willing to spend X months} * {% success if spent X months} = {% success rate of that fraction}	Wet lab biology PhD: ~11% STEM Bachelors: ~1%
[Sanity Check] Non-Biological Datasets: Given that the above methods rely a lot on subjective judgements, I also briefly consider more general reference classes about the error rates in other technical procedures. For example, if there are 14 hard steps in an influenza protocol (Martínez-Sobrido & García-Sastre, 2010) and we *assume* the chance of making a mistake is as high as the 3% seen in medical surgeries (Fabri & Zayas-Castro, 2008), then a naive calculation suggests an overall likelihood of success is 65%.[13] I construct five estimates for both experts and non-experts using such data. These estimates have an obvious external validity issue, but can nonetheless act as a helpful "sanity check" and use more available data. See Appendix 2.4 for details.	Wet lab biology PhD: ~13% but highly uncertain STEM Bachelors: ~0.02% but highly uncertain

Table 3.2.3b | Overview of "Laboratory Success Rate" estimation methods

The Judgemental Forecasting Survey is especially noteworthy. We can see in Figure 3.2.3a that the survey reflects the large amount of disagreement in the field, with even the 75th percentile range of responses spanning several orders of magnitude. It also lets us break down success

[13] I.e. (1-3%)^14 =65%. Note, however, that treating each step as "independent" biases the results down. If a surgeon succeeds at 10 surgeries this suggests they are more capable, so their real success rate for the 11th should be higher.

according to the previously discussed perseverance rate, which we can then combine together into an overall rate of laboratory success. This is shown in Figure 3.2.3b.[14]

Doing so implies that for STEM Bachelors, the current chances of success are heavily concentrated in highly persistent actors. Thus, for AI to have a large counterfactual impact, it would likely need to substantially accelerate novices within a 3–6 month window – a comparatively high capability threshold.

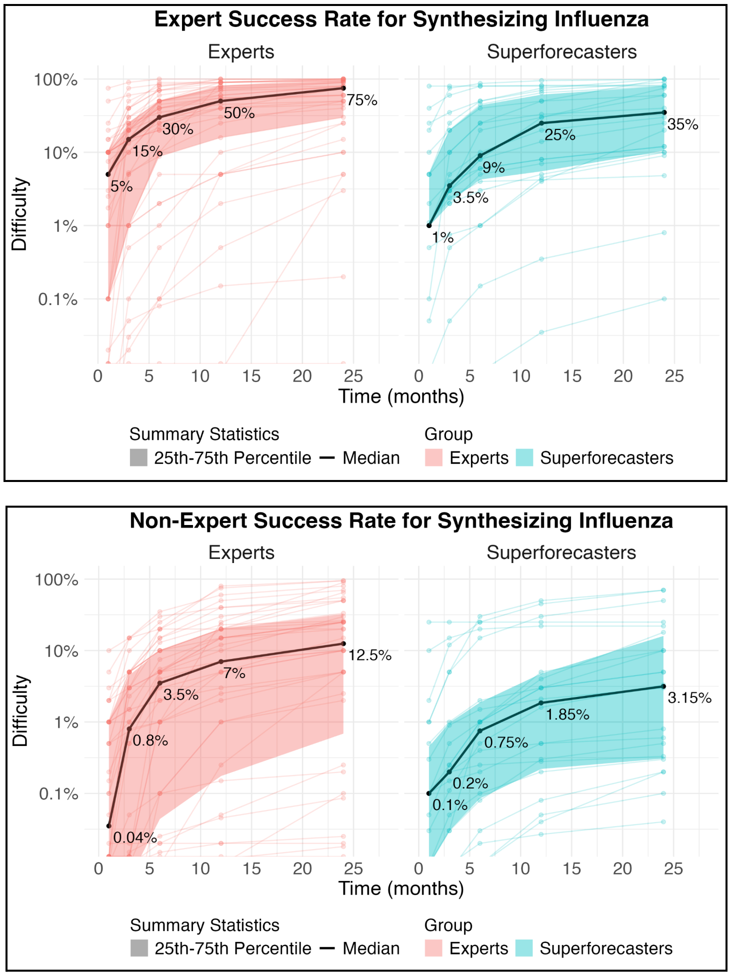


Figure 3.2.3a | Overview of "Conditional Laboratory Success Rate" forecasts

[14] Figure 3.2.3a shows that the median respondent thinks the success rate for STEM Bachelors climbs sharply from 0.8% at 3 months to 3.5% at 6 months to 7% after 12 months. Given the number of actors who persevere through these durations, most of predicted successes accrue to actors investing at least 12 months of effort (63% of the success rate), and the overwhelming share of chances of success is available to actors investing at least 6 months of effort (85% of the success rate).

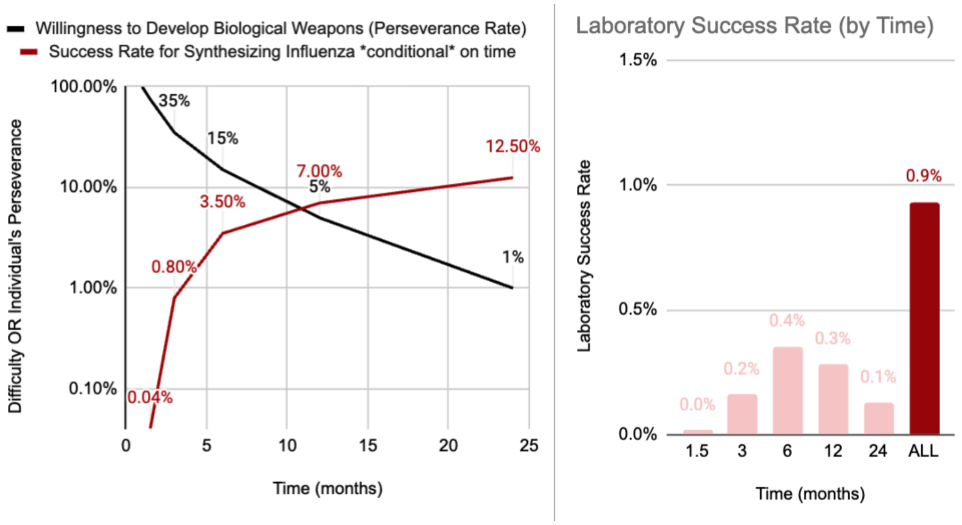


Figure 3.2.3b | "Perseverance Rate" and "Conditional Laboratory Success Rate"combining to an overall "Laboratory Success Rate"

Table 3.2.3c shows how I combined these methods into an overall assessment. It reflects that uncertainty remains high but that a plausible best guess might be that a wet lab biology PhDs has an approximately 20% success rate and STEM Bachelors approximately 1%.

These estimates suggest that despite STEM Bachelors outnumbering wet lab biology PhDs by approximately 100 times, there are only approximately 5 times more STEM Bachelors of concern after accounting for laboratory success. This implies that current baseline risk is still made up to a significant degree by risks from expert threat actors – but that there is a large counterfactual effect that could happen if future emerging technology becomes powerful enough to uplift novices.

Such conclusions could be subject to change. For example, more objective information may emerge through human-uplift studies that test to see how well novices do at various wet lab biology tasks in the laboratory (Paskov et al., 2025). Similarly, the actual difficulty of tasks might itself change with advances in synthetic biology and new automating tools separate from AI.

Variable	Wet Lab Bio PhDs	STEM Bachelors	Other Actors
Number Of Individuals [see prev.]	~150K [40K–400K]	~20M [10M–40M]	~200M [100M–400M]
Laboratory Success Rate (%)	20% [5%–50%]	1% [0.1%–4%]	0.05% [0.005%–0.2%]
Number Who Succeed At Biology	33K [4K–100K]	250K [20K–1M]	130K [8K-500K]

Table 3.2.3c | Overview of "Laboratory Success Rate" author assumptions

Operational Success Rate, O

The third parameter of the model is the percentage of actors who, having the financial resources and capability to successfully synthesise a virus in ideal laboratory conditions , would overcome real-world challenges and succeed operationally. These real-world challenges can include bypassing DNA synthesis screening, setting up a functioning laboratory, and avoiding detection by law enforcement. These are elaborated on in Table 3.2.4a.

Qualitative Discussion Of Operational Success Rate	Author Est.
Acquiring hazardous material (i.e. specific DNA). A common obstacle experts raised is that engineering a dangerous virus requires ordering dual-use DNA that many companies try to screen for (OSTP, 2024). However, experts also emphasised that such defences are currently limited. Governments have issued guidelines, but these are not legally required and implementation can vary a lot in practice (Kane & Parker, 2024). Edison et al. (2024) describes how orders can be 'camouflaged' to avoid detection – and whilst some contested its specific methods, even such responses admit "there remains no screening solution" when someone orders many small pieces from multiple providers (IGSC, 2024). There have also been private red-teamings studies of DNA synthesis screening conducted.	Medium Difficulty
Acquiring other materials for a sterile lab. A threat actor would need to also obtain non-DNA materials and equipment. Because such materials are less directly hazardous, most of these are not tightly controlled: Micro-pipetters can be ordered on Amazon, tissue culture hoods on Alibaba, and there are second-hand markets for other equipment. Still, experts noted that setting these all up in a sterile environment – including reliably powered refrigerators and incubators – is non-trivial. Some cited that Aum Shinrikyo created 'fermentors' to produce C. botulinum for a terrorist attack but plausibly failed due to non-sterile conditions (Danzig et al. 2012), and handling viruses would be even more delicate.	Low Difficulty
Avoiding detection by law enforcement. Some experts noted that illegal biology laboratories have been detected and shut down by law enforcement. However, others cautioned that in practice there is limited oversight (Greene et al., 2023). For example, the illegal Reedley Biolab was discovered due to building regulation vioolation – not because authorities knew that it was a biological laboratory specifically (Reedley Report, 2023).	Low to Medium Difficulty
Dispersal. Some experts noted that whilst an infectious virus will be self-spreading, causing initial infections can still be a non-trivial challenge. Steps include packaging whilst keeping viruses functional and releasing them without denaturing. For example, Aum Shinrikyo's attempts at spraying C. Botulism failed (Danzig et al. 2012) – and viruses will again be more delicate. That said, other experts strongly pushed back that delivery would be much of an additional barrier, and threat actors could pursue several attempts once they have a stock. Specific discussion on strategies is omitted since it is sensitive.	Low to Medium Difficulty

Table 3.2.4a | Overview of arguments relevant to the "Operational Success Rate"

This parameter is included in the simple model to prevent over-estimating the number of actors who might succeed in synthesising a virus in a controlled environment but would still fail at challenges presented in a real world environment. This matters especially if we are to interpret results from human-uplift studies that don't include such operational obstacles (Paskov et al., 2025). Additionally, by estimating the operational and laboratory success rates as separate parameters, we allow for the simple model to reflect that future AIs may have differential effects on laboratory and operational risks (Mouton et al., 2023).

Table 3.2.4a identifies and discusses a non-exhaustive list of operational challenges. Many interviewees agreed that operational challenges provide significantly lower difficulty compared to laboratory ones – and that, one way or another, "a determined adversary cannot be prevented from obtaining very dangerous biological materials intended for nefarious purposes" (Defense Science Board, 2009). However, they also emphasised that the literature on lone wolf terrorism has found that a large fraction of actors are incompetent and prone to operational mistakes (Keynon et al., 2023).

To turn this discussion into a quantitative parameter, this report uses several methods, including a simple sub-model of the operational steps outlined above, compares it to the success rates from CBRN and non-CBRN terrorism plots, and a structured survey of subject-matter experts and superforecasters. These are summarised in Table 3.2.4b. See Appendix 2.5 for details.

An important qualification – highlighted by several experts and superforecasters reviewing an earlier draft of this report – is that the laboratory and operational success rates may exhibit a strong correlation that we need to account for. If we think that individuals who succeeded at the difficult laboratory tasks did so because they are unusually capable or willing to put in the effort, then one's conditional operational success rate should be higher. In other words, we don't want to estimate the operational success rate per se, but instead the operational success rate conditional on having succeeded at the laboratory success rate already.[15]

Thus, there is a compelling argument to take the method estimates from Table 3.2.4b that are for the general population and to adjust these up to the relevant population – especially for the more novice actors. For the purposes of keeping the simple model transparent and easy to amend, I do not introduce a formal correlation parameter and instead make a holistic adjustment (though this can be done such as per the steps in Burgess, 2022).

[15] More formally, instead of estimating "p(laboratory success) * p(operational success)", we are estimating "p(laboratory success) * p(operational success | laboratory success)", and where there is strong reason to think that "p(operational success | laboratory success) > p(operational success)". Note that if we think that p(laboratory success) is mostly due to luck rather than skill, then adding this conditional will have little to no effect.

Different Potential Estimates Of The Operational Success Rate	Value
A Simple “Sub” Model of Operational Steps	
Taking the four operational challenges in Table 3.2.4a, we could simply assume that each represents an independent complex step. Whilst we do not have direct data on these, the Human Error Assessment and Reduction Technique establishes baseline error probabilities for different generic tasks and assigns novel, complex problem-solving tasks the highest failure rate of ~30% (EPD, 2014). Thus, a naive calculation might imply an overall operational success rate of ~25% [=(1-30%)^4].	~25%
CBRN-Based Reference Classes	
The Violent Non-State Actor CBRN Database (2024) records 581 attempts by terrorists to obtain a CBRN weapon, of which 263 were successful at doing so: 45%. We can further filter for individual actors (21/82 = 26%), biological agents (26/130 = 20%), and both (10/48 = 21%). I note that such databases are likely to be biased towards including successes versus attempts (since the former is easier to observe), but also that acquiring general biological equipment may be easier than many of the attempts described that may have involved more monitored substances, such as sarin gas.	~20%
Carus (2001) [p8] similarly notes that in the 20th Century, there were 14 cases of terrorists interested in a biological weapon (though not viruses specifically) and eight of whom successfully acquired these: 57%. I note that the same adjustments as mentioned above likely apply, but expect this to overall be an overestimate of the true success rate for this report.	<57%
Terrorism/Criminal-Based Reference Classes	**Value**
A large literature on lone wolf terrorism has found many actors to leak information, which suggests poor operational security. Schuurman et al. (2018) suggested 26% of lone wolf actors divulge specific intentions about their attacks; Ellis et al. (2016) found “21% shared some details of the planned attack with others”; and Hamm and Spaaij (2015) found 76% engaged in some kind of leakage behaviour. If we assume that specific leakaging of information is a good proxy for people failing at the operational steps necessary, then this gives us an upper bound estimate of ~74-80%.	<75%
Smith et al. (2015) observe that the average lone wolf “survives” 1,900 days from the date of their first preparatory act to their arrest, with 80% lasting more than a year. (For terrorist groups it’s 370 days, with 50% lasting more than a year.) This suggests avoiding detection may not be a high barrier, although it is also unclear how much illicit activity they actually engaged in during that time. Thus, I take this to be more of an upper-bound estimate.	<50%–80%
The Global Terrorism Database (2020) notes that in Europe, only 25% of terrorist attacks cause any deaths (5.5K/21.5K); in the US 11% do (300/3K). If we assume that most attacks	>10%–30%

intentionally try to cause deaths, this suggests that we should expect fairly poor operational planning. However, I note it could well be easier for law enforcement to detect (say) someone acquiring explosive materials than widespread biological equipment and some of the failures may be confounded due to technical (not operational) steps, so I take this overall terrorist failure rate as somewhat of a lower bound.	
Wikipedia (2024) notes that post-WWII, there were 32 plots to assassinate the US president, and only two succeeded in at least injuring the target. I imagine there could be many undocumented cases. Additionally, given the very high security around the president, it's not clear epidemic-scale weapons are harder operations wise. Thus, I take it as a lower bound.	>6%
Lafleur et al. (2014) examine 23 sophisticated and high-value heists. Of these, only five failed. I imagine there is self-selection in the case studies, and actors here are better resourced than CBRN lone wolves. Thus, I take this as an upper bound.	<78%
Judgemental Forecasting Survey	
Additionally, I also show the results of my survey of subject-matter experts in biosecurity and superforecaster participants in Williams et al. (2025), which asked how likely it is that STEM Bachelors could acquire the plasmids for 1918 Influenza DNA. Figure 3.2.4 shows that the median expert thinks such operational challenges could be significant, with the median expert thinking only 10% might succeed even in a no-mandatory-gene-synthesis regime. However note that the survey asked for *any* STEM Bachelor – not those *conditional* on succeeding at laboratory steps. There was also a large cluster of experts who put their estimate at far higher, some of whom cited private red-teaming studies. Thus, I interpret this survey to be a lower bound of the relevant parameter estimate.	>1-10% for STEM Bachelors

Table 3.2.4b | Overview of "Operational Success Rate" methods

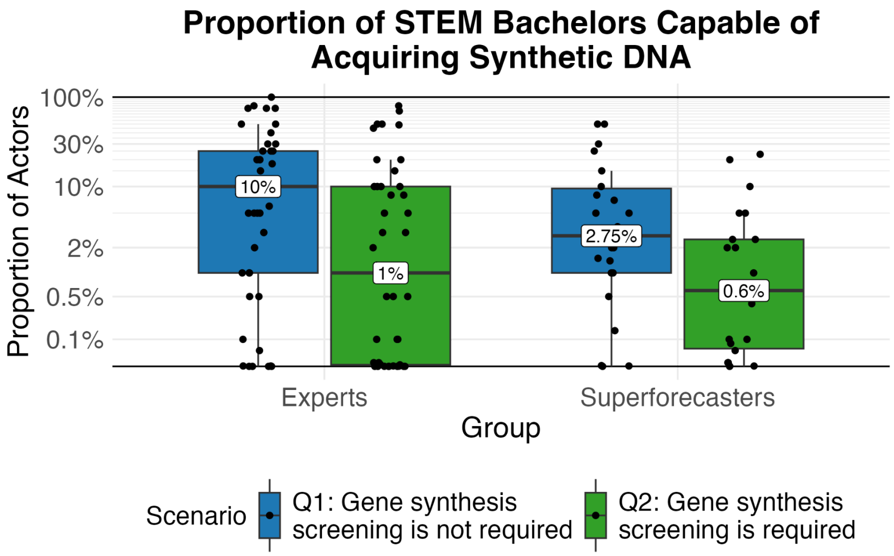


Figure 3.2.4 | Overview of "Operational Success Rate" forecasts

Overall assessments are presented in Table 3.2.4c. Whilst the methods suggest the success rate for the *general* novice population might be circa 20% [10-60%], I holistically adjust upwards to 50% for the *relevant* population. This estimate reflects that operational challenges appear non-trivial but by no means insurmountable – especially for those who would also succeed at laboratory tasks. Moreover, I assign somewhat higher estimates to wet lab biology PhDs as they seem more able to overcome challenges like synthesis screening since they can provide more legitimate seeming pretenses for their operations.

Variable	Wet Lab Bio PhDs	STEM Bachelors	Other Actors
Num. Who Succeeds At Lab [prev.]	33K [4K–100K]	250K [20K–1M]	130K [8K-500K]
***Conditional* Ops. Success Rate (%)**	65% [40%–100%]	26% [10%–70%]	26% [10%–70%]
Number Who "Could" (Lab & Ops)	22K [2K–66K]	100K [4K–310K]	49K [2K–150K]
All	89K actors [23K – 405K]		

Table 3.2.4c | Overview of "Operational Success Rate" author assumptions

Note that any estimate is subject to change, such as if improved and mandated synthesis screening were widely adopted and changed the difficulty of these operational tasks. This was especially reflected in the forecasting survey. Real-world red-teaming work could help create better estimates, such as by empirically investigating how effective DNA synthesis screening is in-practice (Esvelt, 2024; IGSC, 2024).

Radicalization To Pursue Epidemic Harm, R

The fourth parameter of the model is the probability that an actor has the goal of weaponising a synthesised virus. That is, having estimated the number of actors who *could* synthesise a pathogen, this parameter looks at how many actors *would* in fact put in at least 1 month of effort in a given year.

It appears widely accepted that few people have the intention to commit mass violence. Annual homicides in the US occur at 5-8 per 100,000 people (Herre & Spooner, 2023). Wanting to indiscriminately kill hundreds of thousands of people is an even rarer subset still, let alone to do so specifically via engineering a virus.

Exactly how rare the motivations for epidemic terrorism might be is strongly debated (Ingelsby & Relman, 2015). On the one hand, the historical rate of bioterrorism is low (Tin et al., 2022), with few public instances of terrorists pursuing bioweapons to commit mass murder, and even then focusing on non-epidemic pathogens (Carrus, 1998). On the other hand, there are distinct ideologies that may motivate epidemic terrorism if the capability becomes available (Torres, 2018) – and several arguments for not focussing too much on historical data as a predictor of future events, such as ‘availability bias’ (Koblenz, 2017).

Table 3.2.5a shows the main arguments in the literature and from expert interviews. Almost all experts agreed that epidemic terrorism is currently an exceptionally rare motivation to have. However, several still expressed that they think such attempts could be plausible “once-in-a-generation” events and, given the potential large effect, still warrants attention.

Qualitative Discussion Of Intention To Cause Epidemic Harm	
Motivation To Kill As Many People As Possible	
Case *For* Low Intention Rate: Most terrorists do not try to kill as many people as possible. They might be more motivated to attract media attention to their cause or something else. Other methods, such as bombings or political assassinations may be better suited for such goals and involve much less difficulty than pursuing an epidemic pathogen. Many sources also note terrorists are often not very strategic in their operations (Gwern, 2017).	**Case *Against* Low Intention Rate:** There are some extremist ideologies that may motivate epidemic terrorism, such as doomsday cults, anti-natalism, and eco-terrorism – some of whom we have seen experiment with AI (Righetti, 2025). Historically, Aum Shinrikyo and Ted Kaczynski were cited as examples of actors who may have pursued viral pathogens if modern biology had been more accessible then.
Willingness To Kill Indiscriminately	
Case *For* Low Intention Rate: Pursuing a pathogen might be unattractive because infectiousness risks killing people considered ‘in group’. al Qaeda and ISIS directed their scientists to non-transmissible CBRN weapons like chemical weapons and anthrax (Salama & Hansell; UNITAD, 2023), in contrast to, say, Ebola (Hummel, 2016)	**Case *Against* Low Intention Rate:** By virtue of acting alone, lone wolves may be more willing to kill indiscriminately since they have fewer allies and people considered ‘in group’ (Simon, 2013). Torres (2018) shows evidence of some online communities explicitly discussing using pathogens and accepting the collateral damage.
Willingness To Try Something Hard and Unprecedented	
Case *For* Low Intention Rate: There have been no historic instances of terrorists successfully using infectious pathogens (Tin et al., 2022), with many people explicitly citing this is due to its difficulty (Hummel, 2016). Ackerman and Pinson (2013) find lone actors typically engage in simpler plots. Terrorists might be risk averse to trying a novel method, such as how al Qaeda was initially skeptical of using multiple aircrafts as weapons in 9/11 due to “its scale and complexity” (9/11 Commission, 2004).	**Case *Against* Low Intention Rate:** Someone may be motivated to pursue an epidemic weapon precisely because it is novel and thus be especially shocking and get media attention. It could be that some external events – such as extensive media coverage of a new dangerous virus being discovered – might galvanise several bioterrorist attempts at once. Also, the example of 9/11 shows that just because an attack vector is unprecedented does not mean that we can rule it out.

Table 3.2.5a | Overview of arguments relevant to the “Radicalization Rate”

To turn this discussion into a quantitative parameter, this report again uses several methods including reference classes from CBRN terrorist events, other forms of violence such as presidential assassinations, and a survey of forecasts. These approaches are compared in terms of attempts per million people per year (attempts per 1M PYs) and summarised in Table 3.2.5b. See Appendix 2.6 for details.

Description	Attempts per 1M people-years (PYs)
CBRN-Based Reference Classes: There are several detailed case studies of attempts at CBRN and bioterrorism (e.g. Gryphon Scientific, 2016; Carrus, 2017). For these, we can go through and make a subjective judgement as to "Would this threat actor have plausibly pursued an epidemic weapon if they could have?". CBRN Database (2024) notes 13 instances of people pursuing biological agents 1993–2024, of which half appear to have a STEM degree. Rose Hadshar constructed a database and concluded ~2 analogous cases of STEM Bachelors that might have switched to epidemic agents. On the one hand, there may be more undocumented cases; on the other, engineering a virus is very specific. Overall, I take this to roughly cancel and thus treat epidemic terrorism events as occurring twice per 30 years, or ~2 per 200M PYs.	~ 0.01 per 1M people-years (for STEM)
Violence-Based Reference Classes: We can compare epidemic terrorism to other violent events on which there is more available data on and then ask whether we subjectively assess it to be more or less likely. Reference classes include US presidential assassination (Wikipedia), pilot suicides (Kenedi et al., 2016), mass shootings (Duwe et al., 2023), and serial killers (Aamodt, 2016). I think that epidemic terrorism is likely >10X rarer than all of these. But I also worry that the databases underreport the true number of attempts by a large amount. Overall, I assume that *all* epidemic terrorism attempts globally may be approximately as frequent as the rate of *publicly listed* US presidential assassination attempts by Americans, of which there have been ~32 per 400M people-years.	~ 0.1 per 1M people-years
Belief-Based Reference Classes: We can look at how many people hold beliefs that may induce epidemic terrorism (Torres, 2018) and multiply this by how many people in general act violently on their extreme beliefs in other domains (e.g. Westwood et al., 2022). Ideology is only one driver of intent and I only look at three relevant ideologies, so I overall view it as a lower bound.	> 0.00002-0.01 per 1M people-years
Judgemental Forecasting Survey: Lastly, I show the results of my survey of subject-matter experts in biosecurity and superforecasters. We can see both surveyed groups think that wet lab biology PhDs are orders of magnitude more likely to do this than other actors. See Figure 3.2.5. I note that some people's forecasts seem unreasonably high to me, which suggests they may have misinterpreted the question. So I overall view this as an upper bound.	< 1-100 per 1M (for wet lab biology PhDs) < 0.01-100 per 1M (for STEM Bachelors)

Table 3.2.5b | Overview of "Radicalization Rate" reference classes

This parameter appears especially difficult to interpret and subjective judgement is required. Several methodological limitations are highlighted in Table 3.2.5c. Future investigations are likely to find reducing uncertainty in this parameter relatively more difficult than other parameters. However, government agencies with access to private information may have more options available for reducing uncertainty.

Issues With A Quant. Estimate	Description
Public record missing attempts	Not every attempted terrorist attack will be recorded in datasets. If a terrorist plot was foiled or given up on it might not be included in databases like GTD (START, 2020), even though this seems relevant for our purposes. Notably, there is "no mandatory incident reporting requirement [...] to report criminal activity that appears to be ideologically motivated and is mitigated at the SLTT level" (FBI, 2022 [p6]) – and law enforcement may choose to not publicise certain cases. If forecasts just use these official incidences, we might under-estimate the risk. For example, when looking at ricin, there have only been ~8 public incidents in the US in the 21st Century, implying a base rate of 0.3/yr (Wikipedia). But speaking to subject-matter experts, some felt strongly that the true "intention rate" for ricin in the US is >10/yr, or >30X the official figures.
Terrorism as a "wave" phenomena	Terrorists' motivations and weapons of choice can vary a lot year-to-year and are shaped by certain geopolitical events. For example, Jihadi attacks in the West increased from <5 per year to >15 per year in the 2010s (ICCT, 2023) – where half were 'inspired' by individuals with no group connections. This should also make us more uncertain in our extrapolations from historical data. For example, it could be that today few threat actors consider epidemic terrorism, but this might become more salient if there is more media coverage about the topic or an event that inspires "copy cats" – similar to how some studies suggest there are contagion effects due to media reporting of suicides (NIH, 2019) or public shootings (Pew et al., 2019). It is notable that it appears that events like COVID-19 did not seem to increase the salience of biological weapons to ISIS and al-Qaeda (Parachini & Gunratana, 2022) – though of course lone wolf behaviour and future events may be different.
Lining the "Radicalization Rate" and "Laboratory Success Rate"	Throughout expert interviews, several people noted that different types of actors might have different "radicalization rates" than others. For example, epidemic terrorism might be more salient to wet lab biology PhD students because they are more aware that this exists as an option. Bruce Ivins, an anthrax researcher, is suspected of the 2001 Anthrax Attacks (FBI, 2010). Additionally, because they are more likely to succeed they might also be more likely to try it. Intuitively, if someone thinks that building a bioweapon will take a year and only has a 1% chance of succeeding, it seems less likely they'd spend effort compared to someone who thinks it takes 3 months and has a 20% chance of success. All of this pushes in favor of giving groups with more expertise a higher radicalization rate. Importantly, future work could explore linking laboratory success rate and radicalization rate in a formal sub model.
Assessing low probability events	Accurately distinguishing between very low probabilities is difficult, especially when there isn't direct observational data over a sufficiently long time horizon (Koblenz, 2017). I.e., assessing whether an event might occur 1-in-100 years, 1-in-1,000, 1-in-10,000, etc. If this ultimately requires subjective judgement, then we should also be aware of relevant psychological biases.

Table 3.2.5c | Methodological issues with estimating the "Radicalization Rate"

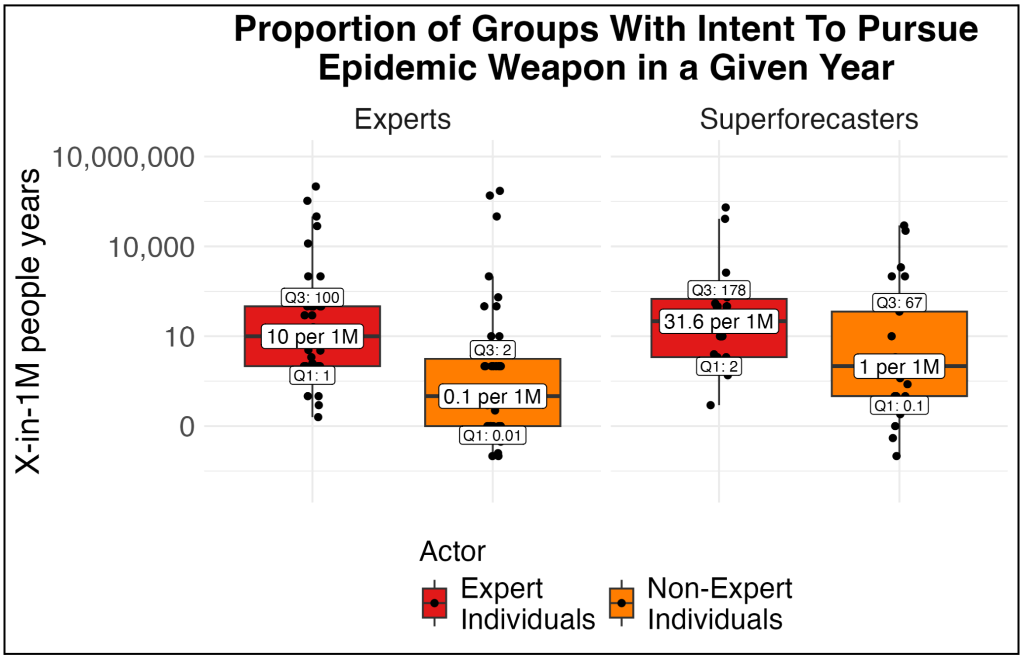


Figure 3.2.5 | Overview of “Radicalization Rate” forecasts

Table 3.2.5d shows overall assessments for this parameter. Credible intervals span multiple orders of magnitude, reflecting significant uncertainty. As noted in the previous tables, the reference class estimates are given more weight than the forecasting survey – and wet lab biology PhDs are given a higher intention rate than others.

Variable	Wet Lab Bio PhDs	STEM Bachelors	Other Actors
Number Who “Could” [see prev.]	22K [2K–66K]	100K [4K–310K]	49K [2K–150K]
Radicalization Rate	0.3 per 1M people yrs [0.03–3]	0.03 per 1M people yrs [0.003 – 0.3]	0.01 per 1M people yrs [0.003 – 0.3]
Attack Likelihood	0.35%/yr [0.02%-6%]	0.1%/yr [0.00%-2%]	0.05%/yr [0.00%-1%]
Attack Likelihood [merge above]	**1%/yr** [0.1%-8%]		

Table 3.2.5d | Overview of “Radicalization Rate” estimates

Escalation Into An Epidemic, E

The fifth parameter in the model is the probability at least 10,000 deaths ensue from the release of a successfully synthesised virus. To motivate this, note that even if an actor synthesises and releases a virus, it does not necessarily follow that the virus will result in an epidemic killing 10,000 people. For this outcome to occur, there must [i] exist knowable epidemic strains (Adalja et al., 2018), [ii] the actor must correctly choose one of these epidemic strains, and [iii] the released virus must spread despite any countermeasures that might be put in place, such as quarantines.

Regarding the existence of knowable epidemic strains, Gryphon Scientific (2016) summarises: "Although several potentially dual-use studies have already been published, translating animal studies of transmissibility to empirically predict an exact R_0 in a human outbreak is currently impossible; therefore, we cannot determine if the studies already published could be used to create strains of influenza that could cause a global pandemic". Learning with certainty whether a sequence is in fact epidemic credible requires extensive "spread-testing" that cannot be done without high risk of detection and for which laboratory settings might be poor proxies (Montague, 2023).

Speculating which specific viruses might have epidemic potential is information-hazardous (Lewis, 2019). This article does not therefore discuss specifics. However, at a high-level, from the perspective of an actor pursuing bioterrorism, there seem to be four salient strategies:

1. Attempting to build smallpox, which is widely believed to be able to cause an epidemic but also significantly harder to build than other viruses like influenza;
2. "Roll the dice" that a strain in the literature turns out to be epidemic credible;
3. Waiting for an external scientific discovery to then misuse.
4. A lone wolf novice actor discovers a novel strain by themselves (though this appears more unlikely unless there is a lot of additional scientific support).

See Table 3.2.6a and Appendix 2.7 for further discussion.

The remaining obstacles appear simpler to overcome. Regarding acquiring the correct strain, some experts note that lists of epidemic potential pathogens are often discussed in the literature (Neumann & Kawaoka, 2023) – and a lot of underlying sequence data can be readily identified in public-databases such as GenBank. Regarding dispersal, this seems very difficult to contain if 5-10 humans are infected, and terrorists might have several attempts at this (Lipsitch & Inglesby, 2014). Some note governments might be able to contain outbreaks from going global like Ebola; others note COVID-19 shows it remains difficult.

Qualitative Routes For Pursuing a Pandemic Potential Pathogen	
Smallpox	Smallpox is often cited as a pandemic threat (John Hopkins, 2001). It was eradicated in 1980 and no longer being widely vaccinated against, meaning many people will not have immunity to it if re-released (Schoch-Spana et al., 2017). However, smallpox is also considerably harder to engineer than (say) influenza viruses – many experts citing it as “10X harder”. There is also a known vaccine and many countries stockpile it, lowering some risk (NASEM, 2024).
Other ‘known’ candidates	Importantly, not knowing whether a currently known pathogen is $R_0>1$ does not mean we can rule out that such a pathogens might in fact lead to a human outbreak (Gryphon Scientific, 2016 [p215]). A threat actor could pick one or several in and then “roll the dice” that one of these works (e.g. Schoch-Spana et al., 2017; Lancet 2024). Interviewees were divided on how successful such an attempt would be, but agreed it is a “known unknown”. Some noted that the likelihood as to whether 1977 pandemic flu and other potential lab-leaks did in fact occur or not should also inform our estimates here.
Waiting for a novel candidate	It could be that as science progresses, more gain-of-function experiments are done (Burki, 2018) or the ability to predict transmission rate improves. Thus, a threat actor may not have to develop a candidate themselves but instead use someone else's work.
Discovering a novel candidate	Many interviewees thought that it unlikely that an individual – especially a non–expert – would discover a novel pandemic potential pathogen themselves, given that even professional scientific groups currently struggle to do so. Carter et al. (2023) notes: “The challenge of discovering novel variants with altered or enhanced characteristics is underscored by the extensive resources and expertise legitimate researchers require to conduct such research.” Some experts said that if future AI helps with better R_0 predictions or uncovers novel strains that would be major R&D advancement but could change the current status quo.

Table 3.2.6a | Overview of “Epidemic Take Off” arguments

To turn this discussion into a quantitative parameter, this report does not go into as much detail as with other sections due to potential sensitivities.[16] However it acknowledges that there is a fair amount of expert disagreement. For example, it is telling that in the structured survey accompanying this report, both experts and superforecasters appeared to follow a bimodal distribution as to whether an epidemic credible design exists (Figure 3.2.6).[17] As a result, I do not weigh the survey highly.

[16] Additional details may be shared upon request with some decision makers.

[17] Some ambiguities in the question also appear to confuse a few participants. Some participants who said >95% cited past pandemic strains like COVID-19 as “proof of concept” when these strains would not necessarily cause a *new* pandemic today. Many also seemed conflicted on whether RE>1 accounts for real world public health measures or not (which per the intended survey question, it should).

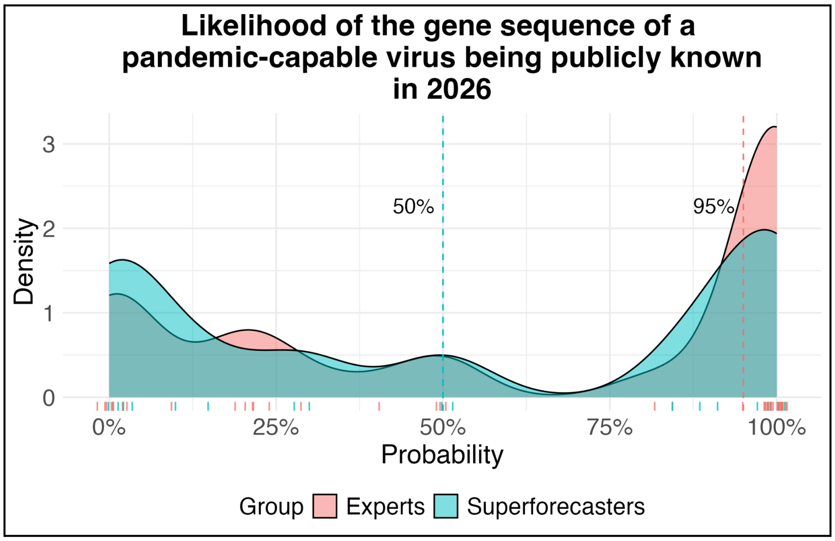


Figure 3.2.6 | Overview of "Epidemic Take Off" forecasts

Table 3.2.6b shows this report's overall assessment that, conditional on a release of a known virus, there is a 20% chance that it results in >10,000 deaths. This estimate draws significantly on private expert conversations, and recognizes the overall wide array of beliefs with a wide credible interval.

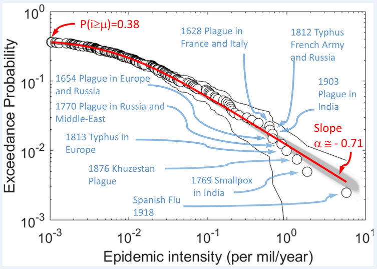
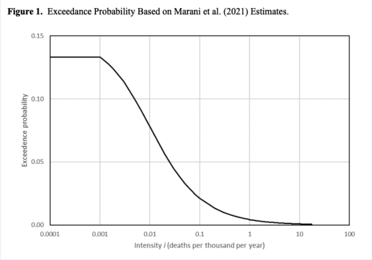
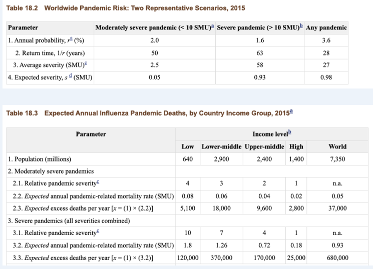
Variable	Description	Value
Likelihood of attack	*See previous section*	1%/yr [0.1%-8%]
Likelihood "takes off"	Based on expert discussion and Gryphon Scientific (2016). Approx.: 40% chance known designs are epidemic credible * 50% chance terrorist identifies this * 80% it takes off	~20% [10%–40%]
Likelihood of epidemic	*[Calculation: Multiplying all of the values in each column together]*	0.15% [0.0%-1.4%]

Table 3.2.6b | Overview of "Take Off" estimates

Potential Damages, D

The sixth and final parameter in the model is the probability that, conditional on having killed at least 10,000 people, how much harm the epidemic virus goes on to cause. A deliberate epidemic terrorism event has not yet occurred (Gryphon Scientific, 2016 [p243]) – and so we do not directly know how many people might die from this However, the severity of natural pandemics appears to be a well suited direct proxy – and there is ample data on this.[18]

Table 3.2.7a summarises three prominent sources that examine the distribution of epidemics over time.[19] Additionally, to validate these estimates, I also asked about this variable in the structured survey conducted by Williams et al. (2025) (Figure 3.2.7a).[20]

Marani et al. (2021)	Glennester et al. (2023)	Fan et al. (2016)
2.2M deaths per year; event lasts 1-3 years	1.5M deaths per year; event lasts 1-3 years	~20M deaths per event
Study epidemics since 1600 and find that follows a Pareto distribution	Adjust Marani, such as by using only modern data, cutting 1B death tail etc.	Review influenza pandemics in the 20th Century and split these into two scenarios
		

10^0
10^-1
10^-2
10^-3
Exceedance Probability
P(i≥μ)=0.38
1628 Plague in France and Italy
1812 Typhus French Army and Russia
1654 Plague in Europe and Russia
1903 Plague in India
1770 Plague in Russia and Middle-East
Slope α ≅ - 0.71
1813 Typhus in Europe
1876 Khuzestan Plague
1769 Smallpox in India
Spanish Flu 1918
10^-3
10^-2
10^-1
10^0
10^1
Epidemic intensity (per mil/year)

Figure 1. Exceedance Probability Based on Marani et al. (2021) Estimates.

0.15
0.10
0.05
0.00
Exceedence probability
0.0001
0.001
0.01
0.1
1
10
100
Intensity i (deaths per thousand per year)

Table 18.2 Worldwide Pandemic Risk: Two Representative Scenarios, 2015

Parameter	Moderately severe pandemic (< 10 SMU)[a]	Severe pandemic (> 10 SMU)[b]	Any pandemic
1. Annual probability, r[a] (%)	2.0	1.6	3.6
2. Return time, $1/r$ (years)	50	63	28
3. Average severity (SMU)[c]	2.5	58	27
4. Expected severity, s [d] (SMU)	0.05	0.93	0.98

Table 18.3 Expected Annual Influenza Pandemic Deaths, by Country Income Group, 2015[a]

Parameter	Income level[b]				
	Low	Lower-middle	Upper-middle	High	World
1. Population (millions)	640	2,900	2,400	1,400	7,350
2. Moderately severe pandemics					
2.1. Relative pandemic severity[c]	4	3	2	1	n.a.
2.2. *Expected* annual pandemic-related mortality rate (SMU)	0.08	0.06	0.04	0.02	0.05
2.3. *Expected* excess deaths per year [x = (1) × (2.2)]	5,100	18,000	9,600	2,800	37,000
3. Severe pandemics (all severities combined)					
3.1. Relative pandemic severity[c]	10	7	4	1	n.a.
3.2. *Expected* annual pandemic-related mortality rate (SMU)	1.8	1.26	0.72	0.18	0.93
3.3. *Expected* excess deaths per year [x = (1) × (3.2)]	120,000	370,000	170,000	25,000	680,000

Table 3.2.7a | Overview of "Potential Deaths" literature

[18] We might think that because threat actors might pursue the "easiest" to build viruses rather than the most severe ones. But threat actors are deliberately trying, which lowers the probability of the least severe ones.

[19] Notably, both the Marani and Glennester paper appear to measure severity in terms of annual deaths not total deaths of an epidemic. For example, COVID is denoted as 2.5M recorded deaths in the first 72 weeks [Marani p3], we know that it caused 20M in excess mortality over three years. Since we are interested in total excess deaths, we need to adjust for that – which I do by assuming "2-years per event". (Not needed for the Fan paper.)

[20] Importantly, survey participants were given, so this should indeed be more seen as a validation rather than an independent form of evidence

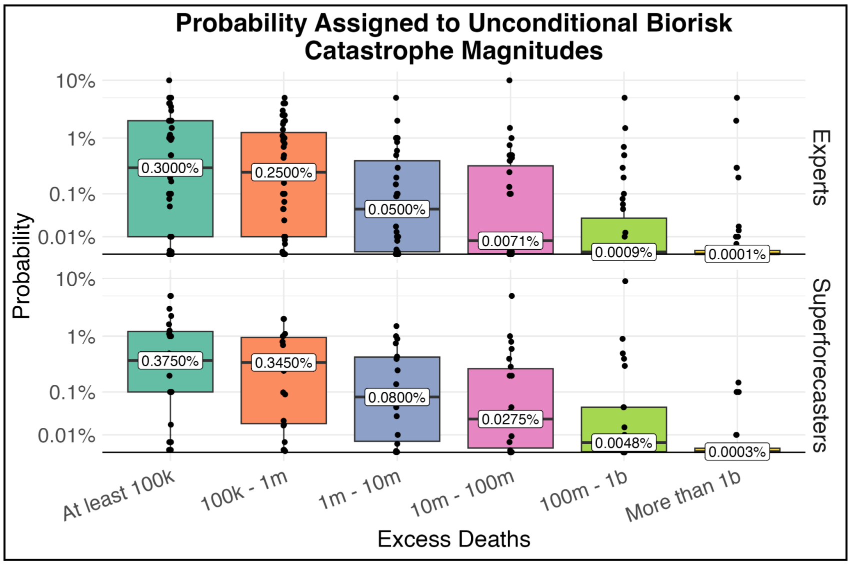
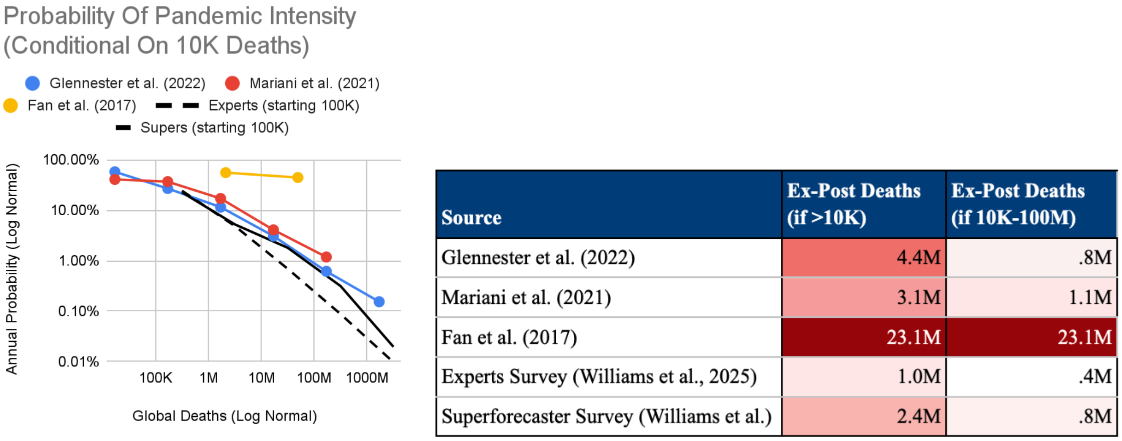


Figure 3.2.7a | Overview of "Potential Deaths" forecasts

To help compare across these different methods, Figure 3.2.7b plots each distribution onto a single graph and estimates the corresponding 'average' size implied. Excluding the Fan et al. (2016) method, we can see most estimates range 1M-4.5M deaths. However, much of this average appears to be driven by a small chance of very large epidemics that kills >100M people. If we exclude such events too we see the range of averages decline to 0.4M-1.1M. Thus, how to weight tail events thus appears especially important.[21]



Source	Ex-Post Deaths (if >10K)	Ex-Post Deaths (if 10K-100M)
Glennester et al. (2022)	4.4M	.8M
Mariani et al. (2021)	3.1M	1.1M
Fan et al. (2017)	23.1M	23.1M
Experts Survey (Williams et al., 2025)	1.0M	.4M
Superforecaster Survey (Williams et al.)	2.4M	.8M

Figure 3.2.7b | Overview of "Potential Deaths" sources

[21] Interestingly, forecasters, especially subject-matter experts, seem to think that very large human-caused epidemics killing >100M people are far less likely than the historical data in Mariani et al. (2021) appears to imply – perhaps suggesting that designing such a virus intentionally is far harder.

Table 3.2.7b shows this report's overall assessment. Future work may want to do more to differentiate this parameter by threat actor type, such as if more sophisticated actors can perhaps build more deadly epidemic viruses.

Variable	Description	Value
Likelihood of epidemic	*See previous section*	0.15% [0.0%-1.4%]
Potential Deaths	Assumes that epidemics follow a Pareto distribution, declining more after ~1M deaths (Marani et al., 2021). To model this, I take a range mean between Ebola (0.1M) and outbreaks slightly smaller than COVID (10M).	~2.5M [0.1M-10M]
Expected Deaths	*[Calculation: Multiplying all of the values in each column together]*	2K/yr [146–35K]

Table 3.2.7b | Overview of "Potential Deaths" sources

For the reports' headline results, the only type of damage considered is mortality. However, epidemics can also cause economic, morbidity, and educational harm (Glennester et al., 2022). Thus, for an additional statistic, Appendix 2.8 also looks at how economic damages typically scale with epidemic size, and then combines these into a single dollar number using a value of statistical life – a concept typical for cost-benefit assessments (Joiner, 2023).[22]

Model Results

Baseline Scenario

Having estimated all six parameters, we can combine these to estimate the baseline risk of an epidemic caused by a lone wolf bioterrorist using simple Monte Carlo simulations. The model estimates that 89,000 actors today might be capable of synthesising a known virus [23,000 – 405,000]. This is a non-trivial number, but far from anyone. Since few are motivated to cause an epidemic, the model implies a 0.15% annual chance of an attack [0.02% – 1.4%], equivalent to 2,000 ex ante deaths per year [146 – 35,000] (Table 3.3.1).

We can see these results have large credible intervals driven by uncertainty about radicalization rates among wet lab biology PhDs and the tail risks of epidemic severity. Nonetheless, the estimates suggest that while the risk of epidemic terrorism is non-zero, it is much below other public health priorities. Note, the credible intervals here are fairly simple estimates from running 100,000 Monte Carlo simulations. Future work could add more complexity to the model to get more sophisticated results.

[22] For decision-makers interested in doing this additional step, see Appendix 2.8 for my suggestion.

Variables	Model 'Baseline' Est.		
Type Of Individuals:	**WLB PhDs**	**STEM BSc.s**	**Other**
A: Number of individuals of this type	~150K	~20M	~200M
L: % could synthesise virus in lab-setting	~20%	~1%	~0.1%
O: % would still be caught or otherwise stopped	~67%	~26%	~26%
Number of individuals could engineer viruses:	**89K actors** [23K – 405K]		
R: % would try to make a virus that year	~0.3/1M	~0.03/1M	~0.01/M
E: % attack "takes off" into an epidemic	~20% [10%–40%]		
Likelihood of an epidemic from lone wolf attack:	**0.15%/yr** [0.02%-1.40%]		
D: Potential deaths if a "take off" occurs	2.5M [0.1M–10M]		
Ex ante annual damages from lone wolf attack:	**2K deaths/yr** [146–35K]		

Table 3.3.1 | Overview of Author's "Baseline" Estimates

Scenario A: "AI Lab & Ops Coach"

Having established a baseline estimate of risk, we can now examine how specific risk scenarios where specific future AI capabilities emerge might change it. The first risk scenario is defined as seeing AI systems exceed two capability thresholds that map onto the Laboratory and Operational parameters:

AI Lab Coach. An RCT finds that frontier AI systems increase the proportion of STEM Bachelors able to successfully complete an end-to-end virus rescue protocol for an influenza strain in a laboratory setting over the course of three months by 10 percentage points. I.e. if previously only 1% succeed, now 11% do. The study is precisely defined in Appendix 3.1.

AI Ops Coach. An RCT finds that frontier AI systems enables STEM Bachelors to design an epidemic terrorist plan in two weeks that a panel of US national security community-appointed judges assigns an average score of 7/9 ("Satisfactory") per the grading rubric in Mouton et al. (2023). The score should cover the specific strain, how to obtain DNA, and disperse the virus. The 2023 study found Internet-only teams of 3-people scored 3.5/9 ("Problematic").

The scenario estimates risk increases before any additional safety mitigations are implemented, since the purpose is to help decision-makers assess the value of such mitigations.

I estimate this risk scenario by changing certain parameters as shown in Table 3.3.2.[23] We can see that doing so implies that the number of actors capable of synthesising a known virus might increase from 89,000 to 981,000 – about a ten-fold increase. As a consequence, the risk of a lone wolf epidemic increases from 0.15%/yr to 1.05%/yr – a smaller seven-fold increase. As a result, the expected damages increase from 2,000 deaths/yr to 14,000 deaths/yr. Using a value of statistical life of ~$8.75M (US FEMA, 2022; see Appendix 2.8), this increase of 12,000 deaths/yr is equivalent to $100B/yr, excluding additional non mortality damages.

Variables	Model 'Baseline' Est.			'AI Lab & Ops Coach' Est. *(Pre-Mitigation)*		
Type Of Individuals:	**WLB PhDs**	**STEM BSc.s**	**Other**	**WLB PhDs**	**STEM BSc.**	**Other People**
A: Number of individuals of this type	~150K	~20M	~200M	~150K	~20M	~200M
L: % could synthesise virus in lab-setting	~20%	~1%	~0.1%	~40%	~11%	~1%
O: % would still be caught or otherwise stopped	~67%	~26%	~26%	~67%	~26%	~26%
Number of individuals could engineer viruses:	**89K actors** [23K – 405K]			**981K** [206K – 4.87M]		
R: % would try to make a virus that year	~0.3/1M	~0.03/1M	~0.01/M	~0.3/1M	~0.06/1M	~0.02/M
E: % attack "takes off" into an epidemic	~20% [10%–40%]			~20% [10%–40%]		
Likelihood of an epidemic from lone wolf attack:	**0.15%/yr [0.02%-1.40%]**			**1.05%/yr** [0.15%-12.75%]		
D: Potential deaths if a "take off" occurs	2.5M [0.1M–10M]			2.5M [0.1M–10M]		
Ex ante annual damages from lone wolf attack:	**2K deaths/yr** [146–35K]			**14K deaths / yr** [1K – 305K]		

Table 3.3.2 | Overview of Author's "AI Lab & Ops Coach" Estimates (Pre-Mitigation)

[23] The most direct change in this model is to raise STEM's laboratory success [L] from 1% to ~11%. I also assume that the increase in operational success rate [O] roughly cancels between the fact that we previously conditioned on people getting to this stage having higher innate ability. I also consider the following indirect effects: if an AI system is powerful enough to help STEM Bachelors synthesise viruses, it might also help wet lab biology PhDs, even if the effect on their capacity is smaller. I raise it by a factor of two to 40%; if people know an AI increases their success rate by a lot, more people might be willing to try it – and I raise it to 0.06/1M.

Scenario B: “AI Lab, Ops & Virology Coach”

The second risk scenario now asks that in addition, AI systems (or other technology) also assists with the third hard step. Specifically, it is defined as follows:

Virus Discovery. An expert red-team produces a sequence design and report in a secure setting that the majority of a panel of judges appointed by the US national security community thinks is likely to be comparable to a COVID-19 outbreak if released to an unexposed population due to vaccine escapes or other traits [i.e. >33% of R_0>1].

I operationalise this risk scenario by further changing certain parameters as shown in Table 3.3.3.[24] We can see that doing so implies that the number of actors capable of synthesising a known virus don't increase any further – remaining at 981,000. However, the increase in epidemic credibility means that the risk of a ‘successful’ outbreak now increases notably from 0.15%/yr to 1.3%/yr. Additionaly, because epidemics are now also deadlier, the expected damages increase further too: from 2,000 deaths/yr to 52,000. This increase of 50,000 deaths is equivalent to ~$440B. Again, this is before any mitigations. In Appendix 3.2, I verify that this result is approximately similar to responses in the large-scale survey in Williams et al. (2025).

Variables	Model ‘Baseline’ Est.			‘AI Lab, Ops & Virology’ Est. *(Pre-Mitigation)*		
Type Of Individuals:	**WLB PhDs**	**STEM BSc.s**	**Other**	**WLB PhDs**	**STEM BSc.**	**Other People**
A: Number of individuals of this type	~150K	~20M	~200M	~150K	~20M	~200M
L: % could synthesise virus in lab-setting	~20%	~1%	~0.1%	~40%	~11%	~1%
O: % would still be caught or otherwise stopped	~67%	~26%	~26%	~67%	~26%	~26%
Number of individuals could engineer viruses:	**89K actors [23K – 405K]**			**981K [206K – 4.87M]**		
R: % would try to make a virus that year	~0.3/1M	~0.03/1M	~0.01/M	~0.3/1M	~0.06/1M	~0.02/M
E: % attack “takes off” into an epidemic	~20% [10%–40%]			~30% [10%–60%]		
Likelihood of an epidemic from lone wolf attack:	**0.15%/yr [0.02%-1.40%]**			**1.3%/yr [0.17%-14.63%]**		
D: Potential deaths if a “take off” occurs	2.5M [0.1M–10M]			10M [0.1M–100M]		
Ex ante annual damages from lone wolf attack:	**2K deaths/yr [146–35K]**			**52K deaths / yr [4K – 985K]**		

Table 3.3.3 | Overview of author’s “AI Lab, Ops & Virus” estimates (pre-mitigation). Cells adjusted compared to the baseline highlighted in orange.

[24] I incorporate these into my judgemental forecast by increasing the expected deaths conditional on an epidemic outbreak being much deadlier (“D”). There is also a case to make the odds of an epidemic happening in the first place also shift (“E”).

Additional Results

In addition to scenarios that change multiple parameters for multiple actors at once, we can also investigate how changing individual parameters affects the models' outputs, forming the basis for future work to construct sensitivity analyses. For example:

Figure 3.3.4a shows how deaths vary when just changing the laboratory success rate for only STEM Bachelors. It shows the median lies at ~2,000 deaths, as well as the 5th and 95th percentile range of our Monte Carlo simulations. We can see that increasing this parameter is associated with higher deaths, but even increasing it to 10% still only reaches <10,000 deaths by itself, since other bottlenecks limit the total effect. We get a similar result when just changing the radicalization rate: even if it reaches 1 person per million we are still below <33,000 deaths.

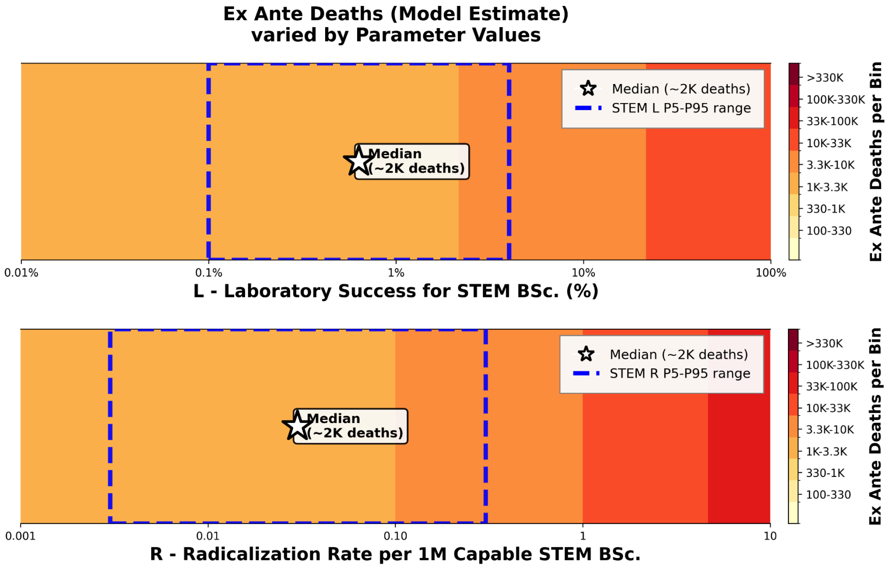


Figure 3.3.4a | Model's est. of ex ante deaths varying a parameter. 100M Monte Carlo simulations divided into bins and color coded according to value.

Figure 3.3.4b shows what happens when we vary both the Laboratory Success and the radicalization rate for STEM Bachelors. This has a multiplicative effect: If actors are both more capable and more willing, then the increase in damages is much higher than either dynamic alone. For example, a ~10% laboratory success rate and a 1 per million radicalization rate are together associated with >100,000 deaths – far greater than either effect individually.

Overall, these results help to highlight the importance of interaction effects: If AI systems make assembling viruses easier, it matters if this also causes bioterrorism to be a more attractive option for threat actors to pursue – a dynamic that seems plausible. This combination is what helped drive the large scenario damage estimates in the previous subsections.

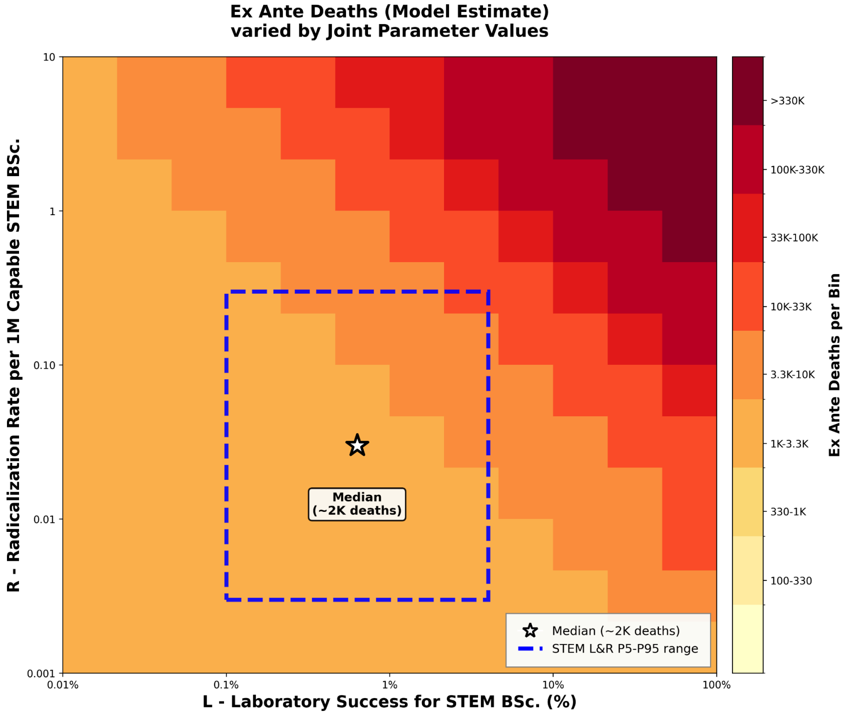


Figure 3.3.4b | Model's est. of ex-ante deaths by varying two parameters. STEM L&R refers to joining variations in "Laboratory Success Rate" and "Radicalization Rate" for only STEM Bachelor actors. 1B Monte Carlo simulations divided into bins and color coded according to value.

Such interaction effects will be hard to measure and capture, but failing to take these into account may result in a large underestimate of the total risk. A simple model can at least help us make such assumptions more explicit and transparent. Future work and risk assessments may want to expand on this dynamic.

Judgemental Forecasting Survey

Methodology

As noted, the estimates in Section 3.2 require a lot of subjective judgements. To review these assumptions and better reflect the views of subject-matter experts, the simple model was then vetted by 11 participants.

This included six subject-matter experts. While these experts were recruited using convenience sampling, I sought to purposefully include individuals with different viewpoints based on their previous published work. It also included five highly credentialed forecasters ("superforecasters") recommended by the Forecasting Research Institute based on having provided high-quality responses in Williams et al. (2025) and without my knowledge of their views.

The individual responses are anonymised, but the group of respondents is listed in Table 4.1. Participants were advised to spend approximately 5 hours constructing their forecasts. To incentivise engagement, experts were paid $1,000, and superforecasters were paid $400. The survey ran between March and April in 2025.

Subject-Matter Expert	Superforcaster
Gary Ackerman is an Associate Professor in Emergency Management at Albany and Founding Director of the Unconventional Weapons & Tech Division at START.	**Nicolò Bagarin**
Sarah Carter is the Principal at Science Policy Consulting, focussing on responsible emerging biotech. She holds a PhD in Neuroscience from the University of California.	**Robert Mahan**
Rocco Casagrande is a Managing Director at Deloitte and owner of Gryphon Scientific, a consultancy in homeland security. He holds a PhD in Biochemistry from MIT	**Dan Mayland**
Forest Crawford is a Senior Statistician at RAND working on reducing risk from biological threats and a previous associate professor of biostatistics at Yale University	**Kjirste Morrell**
Jon Laurent is a Member of Technical Staff at Future House, building AI systems to accelerate science. He holds a PhD in Cell & Molecular Biology from University of Texas.	**Vidur Kapur**
Kathleen Vogel is a Professor in the Future of Innovation at Arizona and wrote "Phantom menace or looming danger?" on biorisk. She holds a PhD in Chemistry from Pinceton	n/a

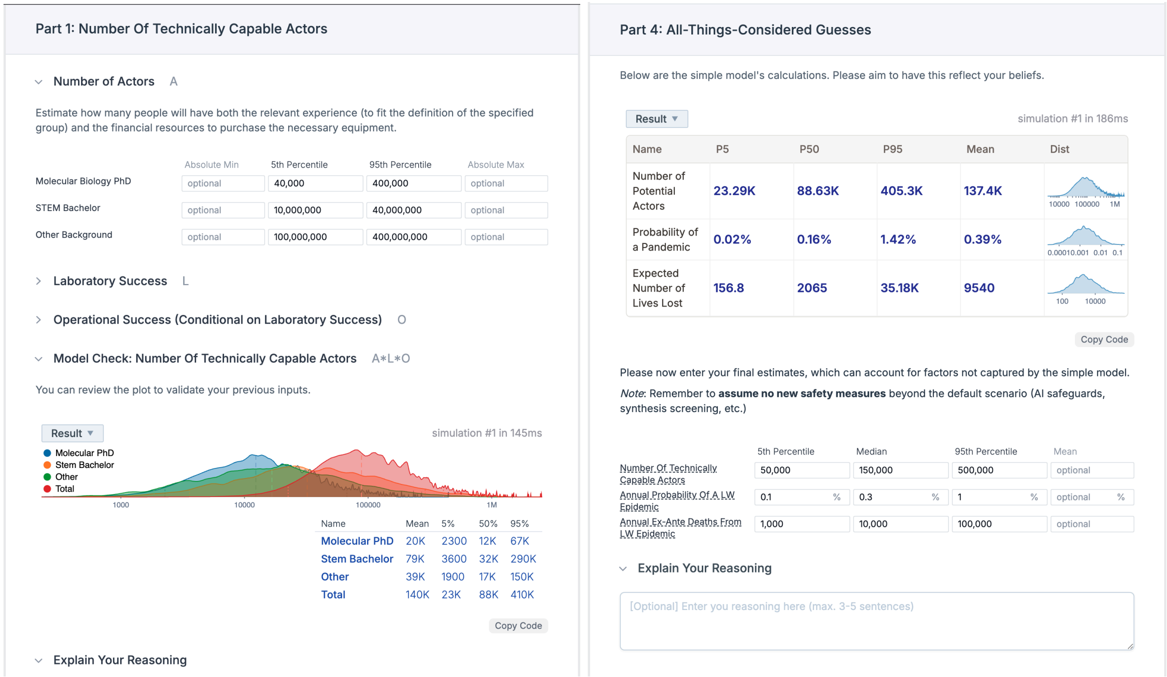
Table 4.1 | Judgemental Forecasting Survey Participants

Participants were provided with an earlier draft of Sections 2 and 3 explaining my reasoning. They were asked to fill out a survey including an interactive version of the simple model built by the Quantified Uncertainty Research Institute (available at https://biocalc.vercel.app/). See Appendix 4.1 for details.

This smaller sample complements the Williams et al. (2025) survey, which saw a large number of respondents produce forecasts without seeing the authors estimates or rationales. Thus, this report can be seen as trading off providing people with more detailed information against anchoring them to that information.

The survey asked participants to first estimate the 5th and 95th percentile of each parameter for each threat actor under each scenario, which was interpolated using a log-normal distribution by default. At the end of each scenario, participants were presented with what the simple model implied about the outcomes of interest and then asked to provide a final "all things considered" estimate of the 5th, 50th, and 95th percentile for each outcome in each scenario. This last step is to allow participants to deviate from the simple model (Figure 4.1).

After participants submitted their responses, the results and rationales were summarised and shared amongst the groups. Participants then had a chance to update any of their estimates. I also alerted those participants who had clear errors or inconsistencies in their responses if they would like to update these.



Part 1: Number Of Technically Capable Actors

Number of Actors A

Estimate how many people will have both the relevant experience (to fit the definition of the specified group) and the financial resources to purchase the necessary equipment.

	Absolute Min	5th Percentile	95th Percentile	Absolute Max
Molecular Biology PhD	optional	40,000	400,000	optional
STEM Bachelor	optional	10,000,000	40,000,000	optional
Other Background	optional	100,000,000	400,000,000	optional

Laboratory Success L

Operational Success (Conditional on Laboratory Success) O

Model Check: Number Of Technically Capable Actors A∗L∗O

You can review the plot to validate your previous inputs.

simulation #1 in 145ms

Name	Mean	5%	50%	95%
Molecular PhD	20K	2300	12K	67K
Stem Bachelor	79K	3600	32K	290K
Other	39K	1900	17K	150K
Total	140K	23K	88K	410K

Explain Your Reasoning

Part 4: All-Things-Considered Guesses

Below are the simple model's calculations. Please aim to have this reflect your beliefs.

simulation #1 in 186ms

Name	P5	P50	P95	Mean	Dist
Number of Potential Actors	23.29K	88.63K	405.3K	137.4K	
Probability of a Pandemic	0.02%	0.16%	1.42%	0.39%	
Expected Number of Lives Lost	156.8	2065	35.18K	9540	

Please now enter your final estimates, which can account for factors not captured by the simple model.

Note: Remember to **assume no new safety measures** beyond the default scenario (AI safeguards, synthesis screening, etc.)

	5th Percentile	Median	95th Percentile	Mean
Number Of Technically Capable Actors	50,000	150,000	500,000	optional
Annual Probability Of A LW Epidemic	0.1 %	0.3 %	1 %	optional %
Annual Ex-Ante Deaths From LW Epidemic	1,000	10,000	100,000	optional

Explain Your Reasoning

[Optional] Enter you reasoning here (max. 3-5 sentences)

Figure 4.1 | Example Screenshots From The Forecasting Survey, Showing Parameter Estimates (E.g. Part 1) And "All Things Considered" Guesses (Part 4). See https://biocalc.vercel.app/

Survey Results

Baseline Scenario

Figure 4.2.1a shows each respondent's 5th and 95th percentile estimate of each parameter in the simple model [light colour bar], resultantly the 5th and 95th median from the both the expert and superforecaster group as a whole [medium colour bar], and lastly how this compares to the author's final results [dark colour bar].[25]

We can see that there was a wide array of individual responses. There are several cases where participants' 5th and 95th percentiles don't overlap – reflecting the real disagreement in the biosecurity debate (Koblentz & Kiesel, 2021). Nonetheless, the medians of both groups appear to largely be consistent with the 5th and 95th percentiles I arrived at. Notably, it seems clear the range of uncertainty within all three groups appears much larger than differences across groups.

A few slight differences between myself, the median expert, and the median superforecaster are listed below. Although, again, none of these appear significant compared to the overall level of uncertainty.

- **Number of Actors:** The median superforecaster thought there were somewhat fewer "Other Actors", with its range of 60M-250M being just over half of my estimate of 100M-400M;
- **Operational Success Rate:** Both the median superforecaster and expert thought "Other Actors" had a lower rate. The expert range of 5.5%-40% was just over half of my estimate of 10%-70%;
- **Radicalization Rate:** The median superforecaster thought there was a somewhat higher rate for actors across the board. STEM Bachelor's 0.01-1 was 3X higher than my estimate of 0.00-0.3;
- **Epidemic Take Off:** Both the median superforecaster and expert thought the takeoff rate varies between wet lab biology PhDs, STEM Bachelors, and Other Actors – whilst I used the same value across all three actor groups.

[25] This percentile aggregation follows Lyon et al. (2015) and others.

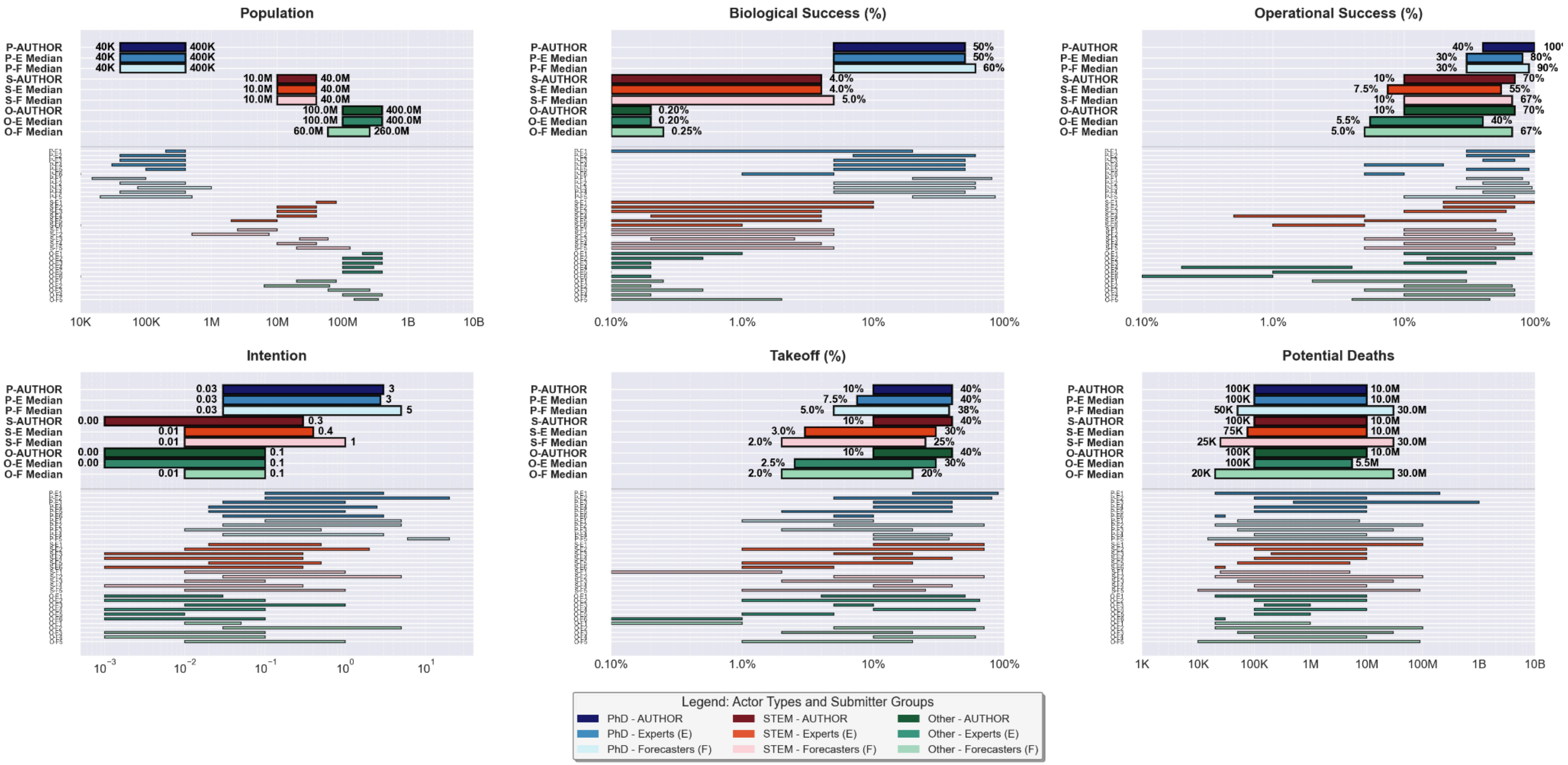


Figure 4.2.1a | Respondent 5th and 95th percentile estimates of the parameters in the Simple Model. Colour indicates the threat actor (Blue = wet lab biology PhD; Red = STEM Bachelor; Green = Other Actor); Colour darkness indicates the group type of respondent (Dark = Author; Middle = Expert; Light = Superforecaster). Wide bars (top of each panel) indicate the group median, whilst narrow bars (bottom of each panel) indicate individual responses

Figure 4.2.1b shows the implied results of the simple model parameters being multiplied together, as well the respondent's ultimate "All Things Considered Guess", which can deviate from this. An overview of the qualitative explanations is presented in Appendix 4.2.

We can again see a similar pattern that there is a large array of individual responses but that once aggregated both groups' range of medians is similar to my overall estimates. Interestingly as well, we can see several participants amended their all-things-considered guesses to be somewhat different from the implied simple model but that overall the model appeared a good approximation.

One notable observation is that we can observe that judgemental forecasts become more uncertain the more real-world assumptions need to be made. That is, taking the median's response to the 5th and 95th percentile estimate for the number of actors spans approximately one order of magnitude; epidemic risk spans two orders of magnitude; and expected deaths three orders of magnitude.

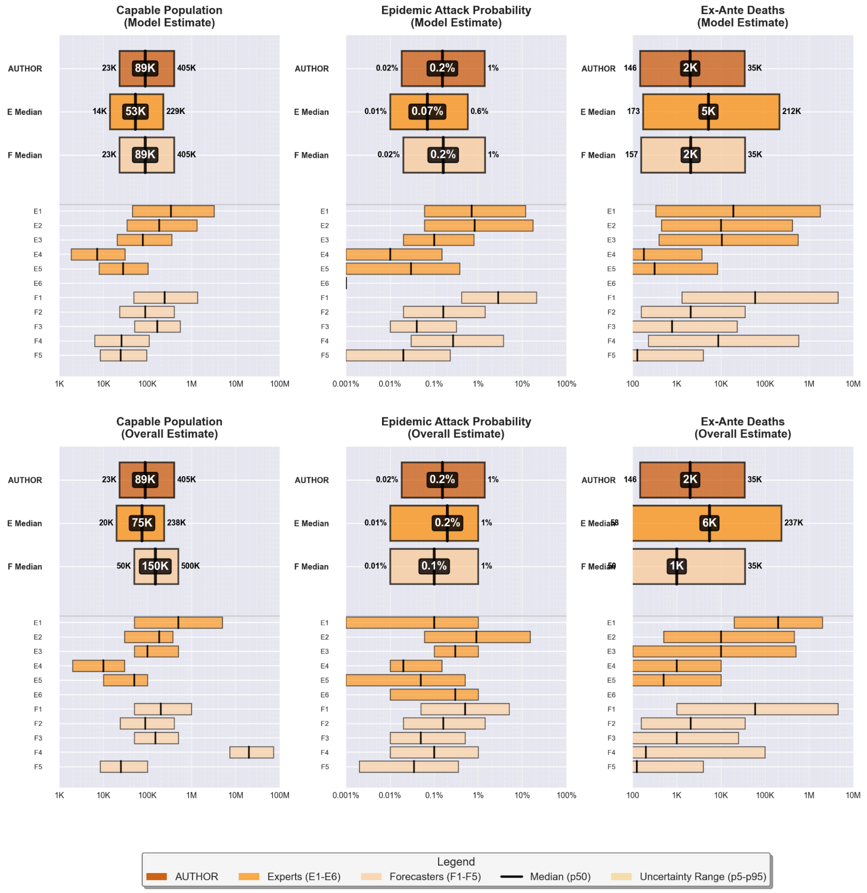


Figure 4.2.1b | Outcomes from the Simple Model and All-Things-Considered Estimates, broken down by individuals (bar = 5th, 50th, and 95th percentiles) and respondent group (bar = 5th, 50th, and 95th median).

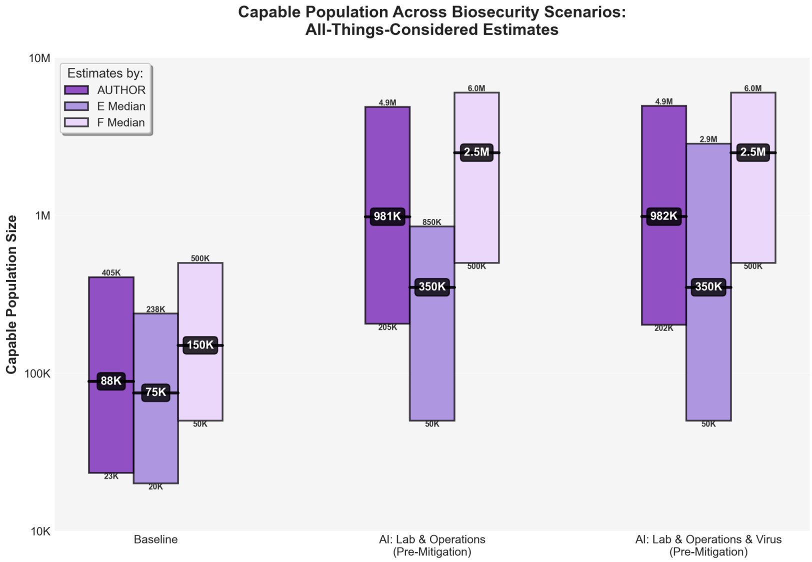
Scenario A & B

Figure 4.2.2 now shows how participants updated their estimates in response to the same two future AI risk scenarios: Scenario A ("AI Lab & Ops Coach") and Scenario B ("AI Lab, Ops & Virology Coach"). For simplicity, this section only reports the all-things-considered estimates and group ranges. Appendix 4.3 contains more detailed figures.

We can again see a large array of opinions but once aggregated both groups' median range is similar to my overall estimates – especially for the ultimate estimate of expected deaths (though some outputs are slightly lower). If we just look at the expert range:

- **Scenario A:** If an AI reaches the threshold of "AI Lab & Ops Coach", then the median subject-matter expert estimate of the number of actors capable of synthesising a known virus might increase from 75K [20K – 238K] to 350K [50K – 850K] (compared to the authors' 89K to 981K). As a consequence, the risk of a lone wolf epidemic increases from 0.2%/yr [0.01% – 1%] to 0.6%/yr [0.08% – 4.3%] (compared to the author's 0.15% to 1.05%). As a result, the expected damages increase from 5K deaths/yr ex ante [58 – 237K] to 7K deaths/yr [500 – 550K] (compared to my estimate of 2K to 14K).
- **Scenario B:** If an AI system also passes the threshold of virus discovery then the number of actors capable remains the same at 350K but with a notably higher range [50K–2.9M] (compared to my 981K). Epidemic risk increases slightly further to 0.8%/yr [0.09% – 7.2%] (compared to my 1.3%). Expected damages increase much further to 68K per year [3K–1.1M] (compared to my 52K).

These results can also be presented as the implied marginal risk from AI systems. See Appendix 4.3.



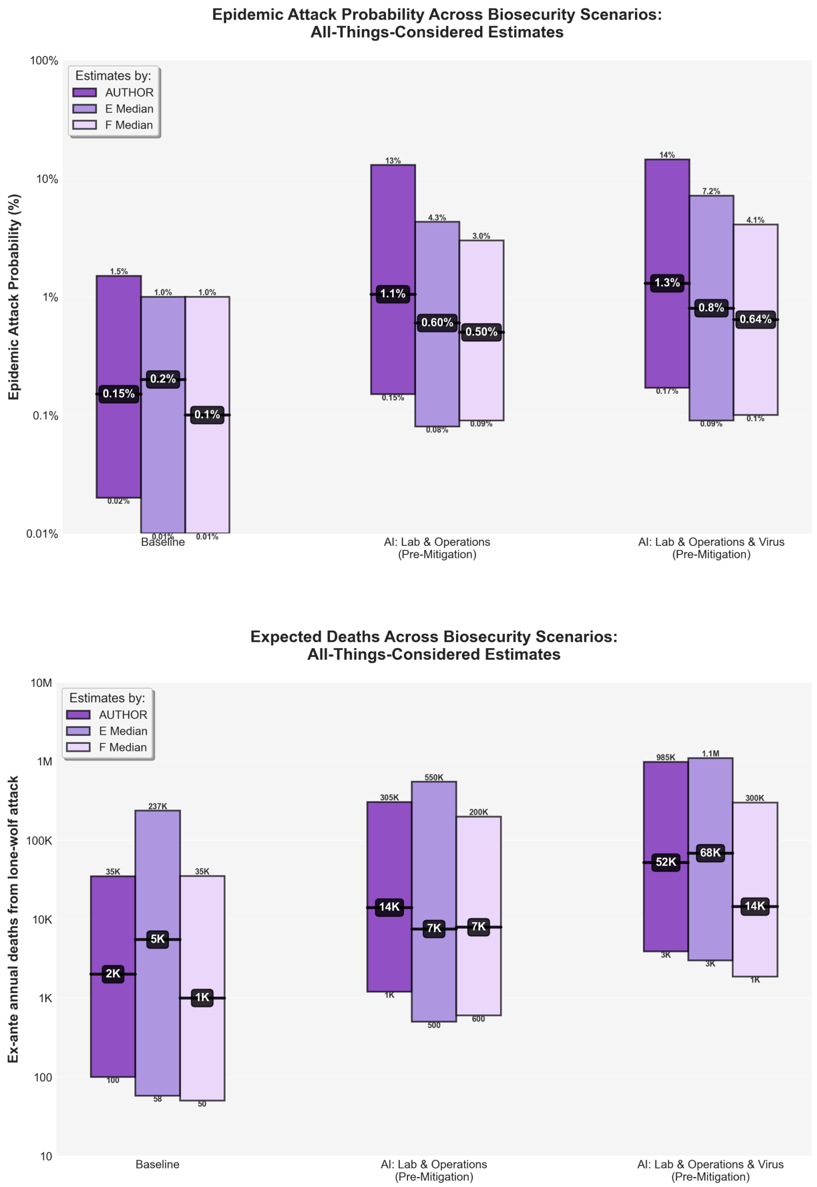


Figure 4.2.2 | Outcomes from all-things-considered estimates, broken down by respondent group (bar = median response to the 5th, 50th, and 95th percentile estimate) and scenario.

Conclusion

This report addresses a critical gap in AI safety assessment: how to translate technical capability evaluations into meaningful risk estimates that can inform high-stakes policy decisions. While frontier AI companies routinely test their models for dual-use biological capabilities, these tests alone cannot tell decision-makers whether – and how seriously – to act without making additional assumptions.

Key Findings

The report develops a single, carefully scoped threat model in more depth: lone wolf epidemic terrorism. It brings together an array of evidence to suggest that future AI capabilities crossing specific thresholds could substantially increase biological risks:

“AI Lab & Ops Coach” (Scenario A) : If AI systems ‘uplift’ ~10% of STEM Bachelors to be able to engineer viruses and also create significant results in an operational risk study, then it is plausible they increase epidemic risk by ~1 percentage points [0.1 – 10] via the ‘lone wolf epidemic terrorism’ threat model, equivalent to 14,000 annual expected deaths [1,000 – 250,000] or ~$100B per year [$10B – $1T].

“AI Lab, Ops & Virus Coach” (Scenario B): If an expert red-team produces a sequence design that a panel appointed by the US national security community thinks is likely to be comparable to a COVID-19 outbreak, then this plausibly signals an additional increase in the risk of an epidemic by ~1 percentage points [0.1 – 10], equivalent to 52,000 annual expected deaths [4,000 – 950,000] or ~$440B per year [$44B–$4T]

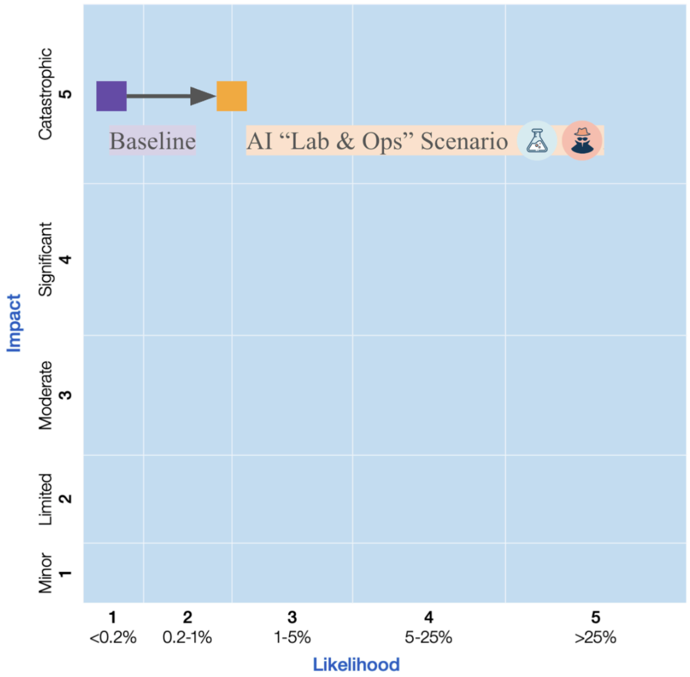
Importantly, these risk scenarios are non-exhaustive. The same AI capabilities could pose additional risks via other misuse vectors, such as chemical weapons or assisting better-resourced groups. They might also be a warning sign of an underlying trend: such as subsequent AI systems being able to design even more catastrophic pathogens corresponding to even greater harms.

Policy Implications

The results suggest that if these capability thresholds are triggered then this would warrant policy attention. Both scenarios meet the definition “severe harm” in OpenAI’s Preparedness Framework, which describes it as “the death or grave injury of thousands of people or hundreds of billions of dollars of economic damage” (OpenAI, 2025c). Similarly, we can visualise the increased likelihood of a deliberate epidemic onto risk assessment matrices such as the UK’s National Risk Register (2025) (Figure 5.1).

At the same time, if we assume that AI capabilities reach but remain within the scale of the two scenarios, then we can also get a more nuanced understanding of appropriate mitigations. The magnitude of risk – while uncertain – demands action, but not overreaction. Tens of thousands of annual deaths, whilst grave, must also be contextualised and triaged against other public health priorities like natural pandemic prevention or global health. This suggests a targeted response: implementing specific safeguards that reduce misuse risk while preserving AI's substantial benefits for science and medicine (e.g. OpenAI, 2025d; King et al., 2025).

Several effective interventions may already be available if these thresholds were to be crossed in the near future. Constitutional AI classifiers (Anthropic, 2025), data filtering (O'Brien et al., 2025), and structured access (Seger et al., 2023) can substantially reduce risk without eliminating beneficial capabilities. Several technical solutions for open source systems are also emerging (Casper et al., 2025). Such a targeted approach would be similar to several historical examples, such as how society addressed automobile safety through mandatory seatbelts rather than banning cars – a proportionate response that preserved utility while reducing harm (Mashaw & Harfst, 1990).



Impact	1	2	3	4	5
Fatalities	1-8	9-40	41-200	201-1,000	>1,000
Casualties	1-18	17-80	81-400	400-2,000	>2,000
Economic cost	Millions of £	Tens of millions £	Hundreds of millions £	Billions of £	Tens of Billions £

Figure 5.1 | Stylised plotting of the author's baseline and scenario risk from a lone wolf epidemic attack onto an impact-likelihood matrix. Adapted from UK National Risk Register (2025) [p14 & p16]

Methodological Contribution

Beyond specific risk estimates, this report also demonstrates how we might better bridge the gap between technical AI evaluations and policy decisions. The report's approach – synthesising historical case studies, expert elicitation, and reference class forecasting through a transparent and simple model – provides a replicable framework for assessing other emerging technology risks. The model's simplicity enables stakeholders to identify and debate key assumptions rather than accepting opaque assessments.

Importantly, the report's methodology also reveals where current evidence is strongest (actor populations, epidemic severity) and weakest (radicalization rates, pathogen viability). This transparency helps prioritise future research and acknowledges that some uncertainty may be irreducible. It also helps to better ground discussion of AI risk into the subject-matter expertise of the fields that it is set to affect, inviting more voices to enter the debate.

Looking Forward

Several key estimates see uncertainty span orders of magnitude and a lot of risk increases are driven by hard-to-measure interaction effects. This underscores a sobering reality: decisions about frontier AI development will have to be made under some irreducible uncertainty. No amount of analysis can eliminate this uncertainty, but structured approaches like that of this report can make it manageable and explicit. In doing so, this report does find reason to take the risks from potential future AI-bioterrorism uplift seriously.

Analysis cannot resolve the deep tensions between innovation and safety, between beneficial applications and misuse potential. What it can do is to help provide a framework for making these trade-offs more transparent, evidence-based, and amenable to expert deliberation. In an era where technological capabilities increasingly outpace our ability to fully understand their implications, such frameworks are essential for navigating the complex landscape of emerging risks while preserving the benefits that make these technologies worth pursuing in the first place.

About the Author

Luca Righetti ✉ 𝕏 in

Senior Research Fellow, GovAI

Luca is a Senior Research Fellow at the GovAI, where he leads a team to investigate national security risks from advanced AI systems. He previously worked at Open Philanthropy' Technical AI Safety team, the University of Oxford's Future of Humanity Institute, and advised the UK Office for AI.

Appendix 1 | Context

1.1 | Prioritising Epidemic-Terrorism By Non-Expert Individuals

Here, I briefly elaborate on why I prioritise [i] pandemic potential pathogens being pursued by [ii] individual actors via [iii] engineering them using reverse genetics to [iv] commit bioterrorism.

Pandemic potential pathogens appear higher fatality than other CBRN scenarios. Marani et al. (2021) note that the average epidemic kills ~300K people [ranging from 10K-100M], with COVID-19 killing 15M (WHO, 2022). It seems reasonable to assume a priori that misuse of a potential pathogen in a pandemic could cause similar harm (Esvelt, 2022). Thus, even a small increase in the likelihood of such an outcome could be highly damaging. For example, an additional 0.5 percentage points in the annual chance of a 300K fatality outbreak is 1,500 annual expected deaths.

By contrast, it seems hard for many other forms of biological misuse to reach such a scale. Tin et al. (2022) notes that between 1970-2019 most terrorist attacks involved anthrax, salmonella, and ricin – and totaled nine deaths overall. The 2001 anthrax letters were the single deadliest instance at five. Of course, larger biological attacks using such agents may be possible – such as if attempts to poison water supplies succeed (Carus, 2001 [p102]). And we might think that such attacks can come in 'waves' (ICCT, 2023; Rapoport, 2022). Nonetheless, to reach >1,000 annual expected deaths, we would need to see >10 biological attacks per year causing >100 deaths each. This seems hard – especially if we imagine that after the first few attacks, there will be a societal defensive response.[26]

One notable exception is mass-weaponised anthrax. The US Office of Technology Assessment (1993) [p54] estimated that 100kg of aerosolised anthrax deployed by an aircraft over Washington DC could kill between 130K-3M. Such an attack has a similar order of magnitude as pandemics. However, such an attack might require producing anthrax at an industrial scale and difficult steps of weaponisation. For context, the anthrax letters contained ~5 grams – or 20,000X less (Broad, 2002). Thus, such a route is plausibly only possible for moderately resourced groups, not lone wolves.

AI uplifting 'non-experts' appears more counterfactual than other threat actors. When doing risk analysis, we care about the additional risk that an AI might pose. If a threat actor already has a high success rate of pursuing a pandemic pathogen, then the counterfactual

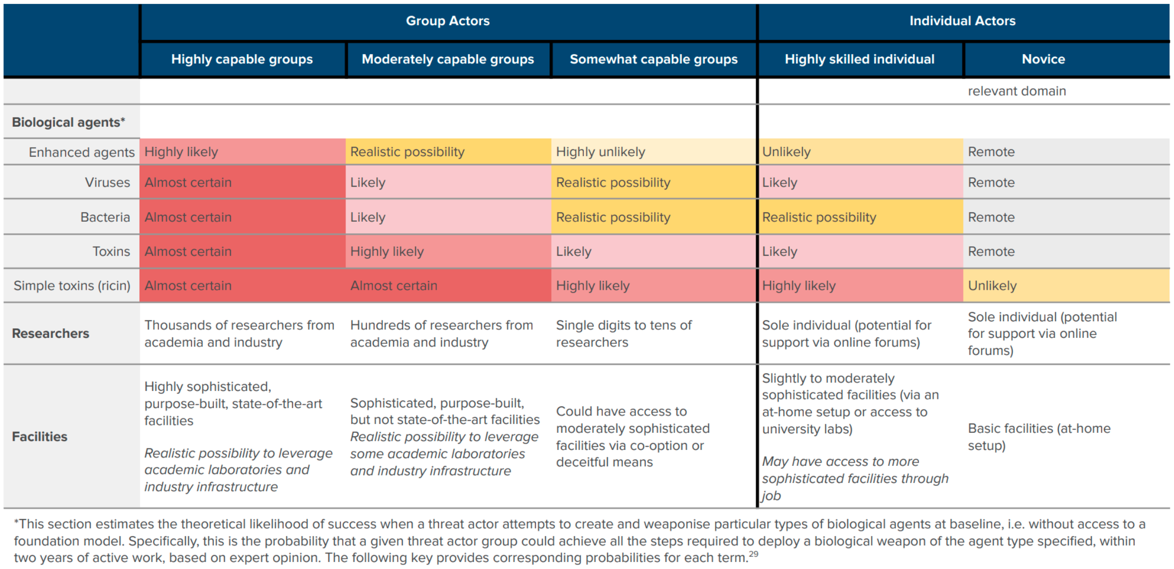
[26] Moreover, if a terrorist wanted to cause >100 deaths there are some non-CBRN pathways they could choose, such as conventional explosives and airplane hijackings (Our World In Data). By contrast, causing ~300K fatalities is plausibly only possible via a weapon-of-mass destruction.

effect from near-future AI may not be high. Thus, apriori, it makes sense to focus on threat models where we have higher confidence that the threat actor cannot do this today.

For pandemic pathogens, several sources suggest that moderately resourced groups and expert individuals plausibly already can. Thus, I focus on non-expert lone wolves, who, by contrast, are unlikely to be able to succeed at this today – but who might be 'uplifted' by future technology.

- Rose et al. (2024) [p45] provides the most explicit assessment. They assess 'novices' currently have a 'remote' (0-5%) chance of succeeding, 'highly skilled individuals' likely (55-75%), and 'somewhat capable groups' a 'realistic possibility' (40-50%).
- Montague (2023) [p10] notes that "very small organizations, disgruntled individuals, and lone wolf actors do not have the resources" whilst "ideologically motivated small organizations [...] remain the most relevant potential actors for biotechnological threats."
- National Academies (2018) [p39] notes that for re-creating known pathogen viruses, the requirements of an actor are of 'medium concern' that "would be achievable by an individual with relatively common cell culture and virus purification skills and access to basic laboratory equipment, making this scenario feasible with a relatively small organizational footprint."
- Gryphon Scientific (2016) [p241] notes that:
  - State actors "clearly often have the ability to acquire the equipment and expertise to use reverse genetics to create any strain of influenza or coronavirus described."
  - Expert individuals "with scientific training may have the ability to perform the manipulations necessary to obtain modified pathogens via simple methods."
  - Non-expert individuals mostly cannot do this today but "could leverage advancing technologies to gain a significant body of skills and knowledge" in the near future
- WHO Scientific Working Group (2015) [p8] noted it "would be possible to recreate variola virus, and that this could be done by a skilled laboratory technician or by undergraduate students working with viruses in a relatively simple laboratory" with a "sustained effort."
- Revill and Jefferson (2014) note, "Individuals involved are professionally trained scientists working on a specific enterprise. The extent to which 'amateurs' are, or

could be, involved is therefore questionable”. However, “advances in ICT have the potential to ‘unlock the master’s secrets’, or at least make these easier to discover.”



	Group Actors			Individual Actors	
	Highly capable groups	Moderately capable groups	Somewhat capable groups	Highly skilled individual	Novice
					relevant domain
Biological agents*					
Enhanced agents	Highly likely	Realistic possibility	Highly unlikely	Unlikely	Remote
Viruses	Almost certain	Likely	Realistic possibility	Likely	Remote
Bacteria	Almost certain	Likely	Realistic possibility	Realistic possibility	Remote
Toxins	Almost certain	Highly likely	Likely	Likely	Remote
Simple toxins (ricin)	Almost certain	Almost certain	Highly likely	Highly likely	Unlikely
Researchers	Thousands of researchers from academia and industry	Hundreds of researchers from academia and industry	Single digits to tens of researchers	Sole individual (potential for support via online forums)	Sole individual (potential for support via online forums)
Facilities	Highly sophisticated, purpose-built, state-of-the-art facilities *Realistic possibility to leverage academic laboratories and industry infrastructure*	Sophisticated, purpose-built, but not state-of-the-art facilities *Realistic possibility to leverage some academic laboratories and industry infrastructure*	Could have access to moderately sophisticated facilities via co-option or deceitful means	Slightly to moderately sophisticated facilities (via an at-home setup or access to university labs) *May have access to more sophisticated facilities through job*	Basic facilities (at-home setup)

*This section estimates the theoretical likelihood of success when a threat actor attempts to create and weaponise particular types of biological agents at baseline, i.e. without access to a foundation model. Specifically, this is the probability that a given threat actor group could achieve all the steps required to deploy a biological weapon of the agent type specified, within two years of active work, based on expert opinion. The following key provides corresponding probabilities for each term.[29]

Appendix 2 | Parameter Estimates

2.1 | Overview

The following appendices provide further details on each of the six parameters in the simple model. Whilst for each parameter the overall estimate is directly imputed via a subjective judgement, in several cases it is informed via additional calculations, a summary of which is given below.

Given many parameters have large uncertainty and long tails, I often assume lognormal distributions. For values bounded between 0–1 (e.g. % able to synthesize viruses), these are fitted onto a beta distribution – though this was not done in the survey since these can be less intuitive [see p. 99].

SIMPLE MODEL	ADDITIONAL ESTIMATION	DISTRIBUTION	METHOD
A: Number of individuals of this type	US People With Appropriate Experience Per Cohort	Lognormal	Data from survey results in NSF (2022) and NSF (2019)
	US Effective Number Of Such Cohorts Today	Lognormal	Subjective judgement based on NSF (2022), NSF (2019), and others
	US Fraction With Sufficient Disposable Wealth	Beta (Approx. LN)	Subjective judgement based on Fed (2022), IPUMS (2023), and others
	Generalising To The Rest Of The World	Lognormal	Subjective judgement based on World Bank (2023), CWUR (2021), and others.
	Overall Estimate	**Lognormal**	**Subjective judgement based on above**
L: % could synthesise virus in lab-setting	Perseverance Rate	Power Series with LN Uncertainty	Subjective judgement based on Williams et al. (2025), Gill et al. (2014), and others
	Laboratory Success Rate *conditional* on Perseverance	Beta (Approx. LN)	Subjective judgement based Williams et al. (2025), Rose et al. (2024), and others
	Overall Estimate	**Beta (Approx. LN)**	**Subjective judgement based on above**
O: % would still be caught	**Overall Estimate**	**Beta (Approx. LN)**	**Subjective judgement**

R: % would try to make a virus that year	Number Of People With Beliefs [Belief-Based Approach]	Lognormal	Subjective judgement based on START (2016), Westwood et al. (2022), and others
	Fraction Of People Who Take Action [Belief-Based Approach]	Beta (Approx. LN)	Subjective judgement based on ICCT (2023), ADL (2023), GTD (2024), and others
	Overall Estimate	**Beta (Approx. LN)**	**Subjective judgement based on above and others, including Williams et al. (2025)**
E: % attack "takes off"	**Overall Estimate**	**Beta (Approx. LN)**	**Subjective judgement based on Williams et al. (2025), Gryphon Scientific (2016), and others**
D: Potential deaths if a "take off" occurs	Average Epidemic Severity	Lognormal	Subjective judgement based on Marani et al. (2021), Glennester et al. (2023), and others
	Value Of Statistical Lives	n/a. Not in main model	Direct from FEMA (2022)
	Economic Output Loss	n/a. Not in main model	Direct from Glennester et al. (2023)
	Equity Adjustment	n/a. Not in main model	Sub. judgement based on Alon et al., (2022)
	Overall Estimate	**Lognormal**	**Subjective judgement based on above**

2.2 | Number Of Actors, A

Estimating Level Of Biological Experience [United States]

US People With Relevant Experience Per Cohort

Baseline Estimates [438K Non-Experts; 3K Experts]
I first describe how I arrived at my own estimate for the number of actors.

The United States has detailed information available on how many people complete university degrees. For relevant biologists, we can look at the NSF 'Survey of Earned Doctorates'. We want to only count people with PhDs that likely taught them highly relevant skills to help them build engineers. For example, many biology PhDs are informatics and don't involve 'wet lab' work. Classifying which PhDs to include requires some judgement and the survey's taxonomization was itself revised in 2021. But using data both before and after this change I get ~3K PhDs per year.[27]

Source	US Relevant Biology PhDs	US Total Biology PhDs	% Relevant
NSF (2022) [Table 3.1] "Detailed Field"	2.6K[28]	10.5K[29]	25%
NSF (2019) [Table 13] "Fine Field"	3.6K[30]	9.8K[4]	37%
Author Assumption	3K [2K-4K]	10K [9K-11K]	30%

For STEM Bachelors we can look at the NCES (2024) [Table 318.20], which make up 438K of the total 2.1M bachelor degrees total – i.e. 20% the total.[31] For non-STEM-Bachelors we can look at the OECD (2024), which notes there are ~210M people aged 15-64, and thus ~4M per cohort.

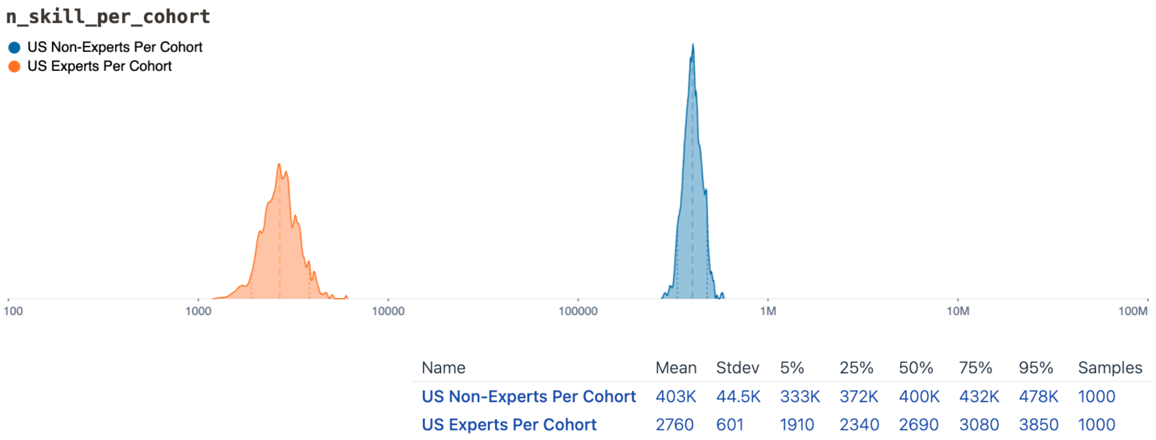
Adjusting For Non-US Students [~0.95X]
One concern raised is that the later 'Rest Of The World Multiplier' may double count people here who have a degree in the US but go on to live abroad. I think this is very likely a small effect. For non-experts, NCES (2024) [Table 318.45] finds that 8% of STEM bachelor students are non-residents. How many of these go on to leave the US is disputed – O'Brien (2024) claims it could be 93% across degrees, whilst Ruiz and Buidman (2018) appear to imply it could be higher. I note even if we take the higher number, the decrease is at most 7%

For experts, the NCES (2024) finds that 50% of STEM PhD students are international, which might suggest the haircut is indeed big. However, the NSF [i] breaks this down by field and finds it is a much lower 27% for biology[32]; and [ii] the proportion of people who stay 10 years later is consistently above 70% (see re-analysis by Corrigan et al, 2022). Thus, the decrease is at most 8% decrease.

Note also, these numbers don't consider the flipside of (e.g.) US citizens who study abroad and move back. Thus, for both groups I take a discount of 1%-10% (0.9X-0.99X) – which turns out to be trivial.

[27] It is worth noting that this latter figure is notably lower than Esvelt's (2022) estimate of >500 experts per year. He notes that "The U.S. grants 125 doctoral degrees in virology each year [and] at least four times as many individuals with degrees in related fields – such as my own PhD in biochemistry – possess similar skills". The data used is consistent. I interpret the difference is that Esvelt estimates a much stricter threshold of who he thinks already has all the necessary skills, whilst I consider a 'wider net' to include people who might be able to more easily acquire such skills with practice. I will account for this with a lower 'Laboratory Success Rate' later.
[28] I.e. those in "Biochemistry, biophysics, and molecular biology" and "Microbiology and immunology".
[29] I.e. "Biological and biomedical sciences" and "Bioengineering and biomedical engineering"
[30] I.e. "Plant pathology", "Bacteriology", "Biochemistry", "Biomedical sciences", "Cell, cellular biology, and histology", "Epidemiology", "Genetics, genomics, human and animal", "Immunology", "Microbiology", "Molecular biology", "Pathology, human and animal", "Structural biology", "Toxicology", and "Virology".
[31] Technically we want to subtract the number of relevant PhDs so as to not double count. But this effect is tiny.
[32] See NSF (2022) [Table 3.4] and NSF (2019) [Table 22]

Overall, this thus leaves me with the following probability distributions [note the log x-axis]:



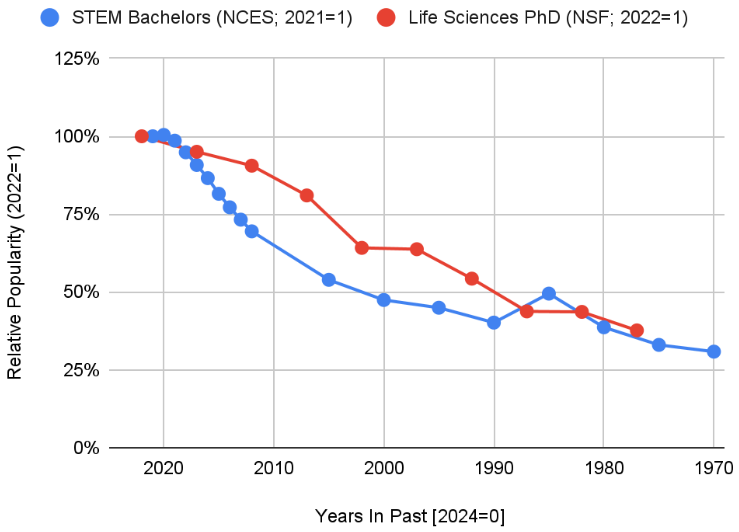
Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
US Non-Experts Per Cohort	403K	44.5K	333K	372K	400K	432K	478K	1000
US Experts Per Cohort	2760	601	1910	2340	2690	3080	3850	1000

Effective Number Of Above Cohorts Existing Today

Baseline Estimate [45X]
The above number includes the people in a 'cohort' who earned such an education in a recent year. If the working-age population is 20–64 that implies up to 45 such cohorts going back to 1980. Naively, we might then want to scale up the above by 45X. For 'non-STEM-Bachelors' we can just leave it at that.

Adjusting For STEM/Biology Popularity Over Time [~0.6X]
However, we might worry that the true number of STEM undergraduates and biology PhDs has changed over time. Usefully, the same sources as above do also report timelines – which I plot below.[33] We can see that in both cases the number of relevant individuals in the late 1970s compared to today seemed to only be 35-40% as much as it is today – corresponding to an approximate decrease of 2.2-2.5%pa for each year we go back. If we draw lines of best fit to find missing values, we get approximately 28 'effective' cohorts for Bachelors [0.6X] and 29 for PhDs [0.65X].

[33] See NCES (2024) [Table 322.10] and NSF (2024) [Table 1.6]



Adjusting For Relevance Of Skills [~0.55X for Experts]

Another concern is that older cohorts might lose their skills over time. This is less relevant for general STEM Bachelors but is especially relevant 'expert' PhDs where we care about specific skills. Regarding the latter we have seen both [i] changes in fundamental techniques like micro pipetting (Asimov, 2024), and [ii] synthetic biology as a whole has only emerged in the last 20 years (Meng & Ellis, 2020). On the other hand, we also don't want to discount this cohort group too much. People from older cohorts may in fact have more overall experience and not find it difficult to re-learn.

From speaking with experts, there was general uncertainty about how to operationalise this. To reflect this, I pick a wide credible interval between 0.5%pa and 5%pa for PhDs (and 0.1%-2% for Bachelors).[34] Doing so we see the following results:[35]

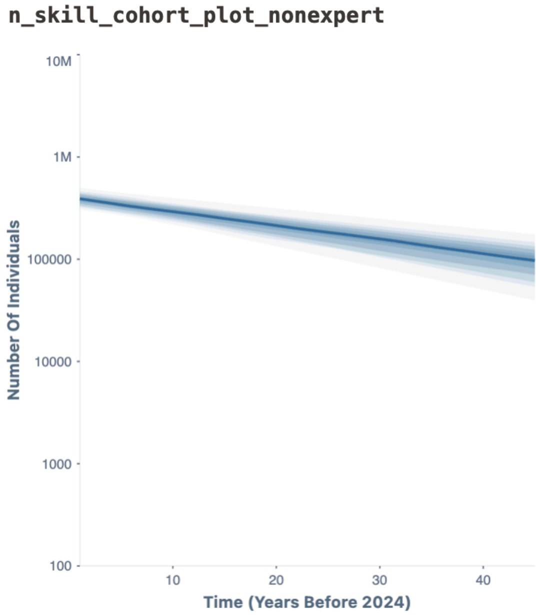
Group	No Decays	Just Time Decay	Just Skill Decay	Both Decays[36]
Wet lab biology PhD 'Effective Cohort'	45	29 [28-31]	22 [10-38]	**16 [8.4-24]**
STEM BSc. 'Effective Cohorts'	45	28 [26-31]	40 [31-45]	**25 [20-29]**
non-STEM 'Effective Cohorts'	45	n/a	n/a	**n/a**

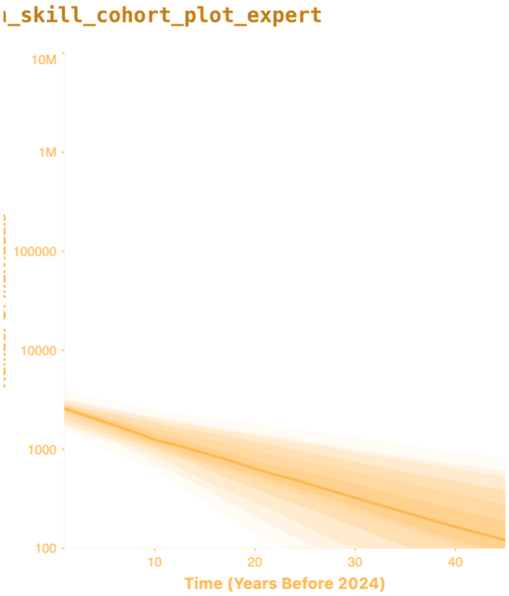
[34] Note that over time, this will naturally tend towards the lower end (Weitzman, 1998).

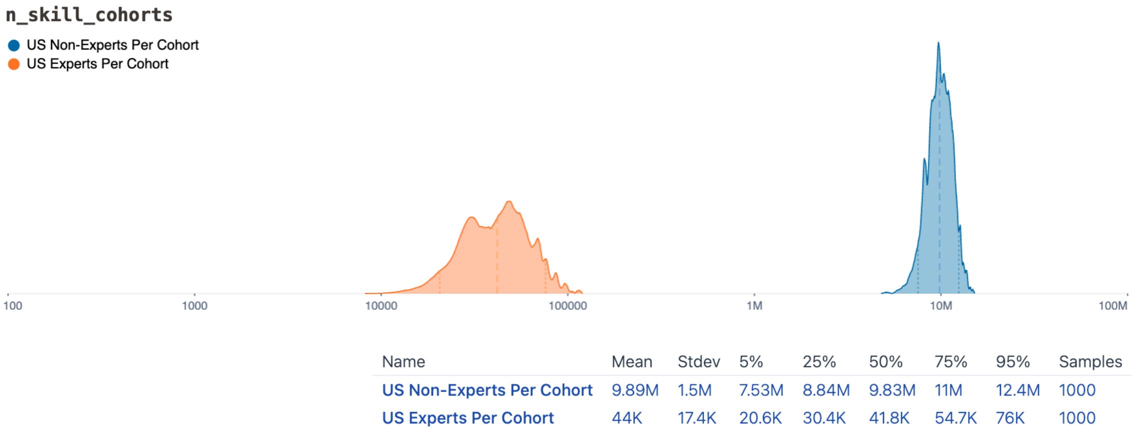
[35] This works out fairly similar to Esvelt (2022) who assumed 20 cohorts for his expert definition.

[36] Note that we want to explicitly include the interaction effect between the two factors.

Final Assumption
Combining all of the above factors we can show the expected number of people for each past cohort and thus the expected number of people overall. I find **~10M people in the US with the effective background of a '2024 STEM bachelor degree' and 44K with '2024 wet lab biology PhDs'.**


n_skill_cohort_plot_nonexpert


_skill_cohort_plot_expert


n_skill_cohorts

Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
US Non-Experts Per Cohort	9.89M	1.5M	7.53M	8.84M	9.83M	11M	12.4M	1000
US Experts Per Cohort	44K	17.4K	20.6K	30.4K	41.8K	54.7K	76K	1000

Estimating Level Of Financial Resources [United States]

Resources Needed To Set Up A Garage Lab

Even if somebody belongs to a relevant threat group, it is non-trivial to acquire the relevant scientific equipment and materials to synthesise a pathogen. Since a threat actor would have to use equipment over a prolonged period of time, they would most feasibly have to outright buy it.[37]

Experts seem to disagree on exactly how costly it is. DeFrancesco (2021) notes that on the one hand, there are people who claim they "sell a pretty complete molecular biology lab for $1,600, and we make a profit. Cost is not a limiting factor" and on the other hand that "horsepox virus 'only' cost $100,000 to produce, according to one estimate: that's still too much for a non-traditional lab."

From my own conversations with molecular biologists and searching for prices online, I think the disposable wealth needed is ~$30k, with a credible interval between $10K-$100K threshold. A lot of the uncertainty comes from how much can be found second-hand or leased.

Fraction With Sufficient Disposable Wealth

Baseline Estimates [0.6X]
Thus, I care that a fraction of our samples above plausibly have enough disposable wealth that they could afford the necessary equipment to synthesise a virus. I take two proxies: consider two proxies:

Firstly, I look at the distribution of US individuals' net worth, which is imperfect but can be adjusted accordingly.[38] Both the Survey of Consumer Finances (2022) [via DQDJ] and US Census (2023) provide estimates of this (the former usefully excludes primary homes that are hard to sell).

Secondly, I look at distribution of US individuals' income as per IPUMS (2023) [DQDJ] and extrapolate their disposable wealth. Per common financial advice, people should aim to have ~6 months of their salary stored in emergency savings and ~4 times their salary in retirement savings (Ally, 2024). I trust this methodology less, but it is a useful sanity check – especially at the 'ends'.[39]

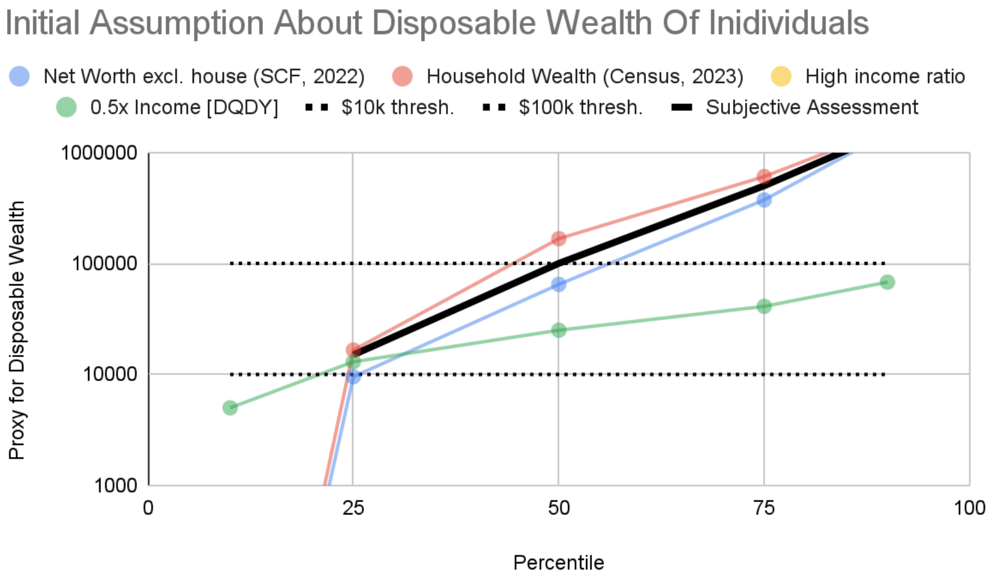
Plotting the reported percentiles, I get the following distributions. Thus, I subjectively assess that ~75% of individual actors exceed the lower $10k threshold and ~50% [10% – 75%] exceed

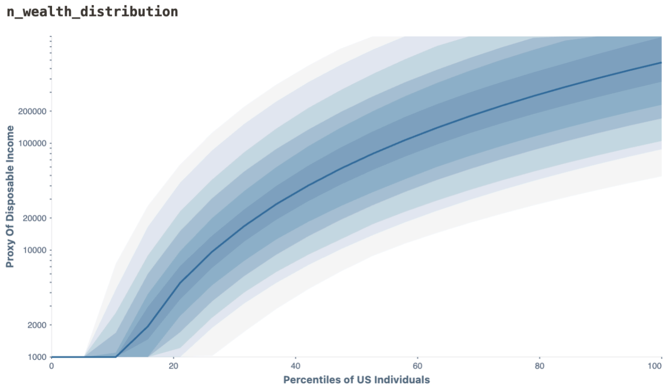
[37] Notably, 'experts' who have relevant biology PhDs may face somewhat lower costs if they already have access to laboratory facilities through their employment. However, following a discussion with some biosecurity experts, I concluded that this was not a big enough factor to be worth accounting for.

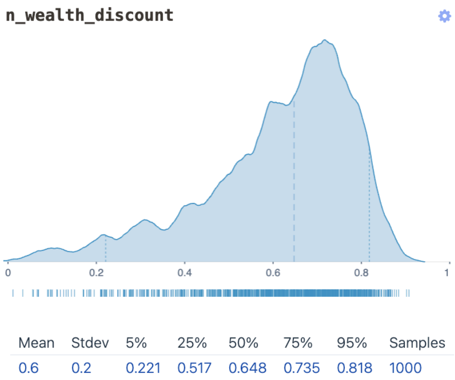
[38] On the one hand, this may be an overestimate. People may not be able to liquidate such assets, especially primary homes or if they are part of a large household. On the other hand, it may be an under-estimate as subtracting 'debt' does not seem to be a real blocker. People might also be able to get temporary finance by taking out loans or asking people in their network. I'd still guess that net worth is slightly too high a proxy for my use case.

[39] An issue with this proxy is I expect the wealth:income ratio to be increasing with income. I.e. People with higher incomes tend to over-proportionately accumulate wealth. This suggests the method might under-estimate the %-above $10k threshold and over-estimate the %-above $100k threshold.

the higher $100k threshold – with the truth somewhere between. I plot my own distribution to fit and get ~60%.







Mean	Stdev	5%	25%	50%	75%	95%	Samples
0.6	0.2	0.221	0.517	0.648	0.735	0.818	1000

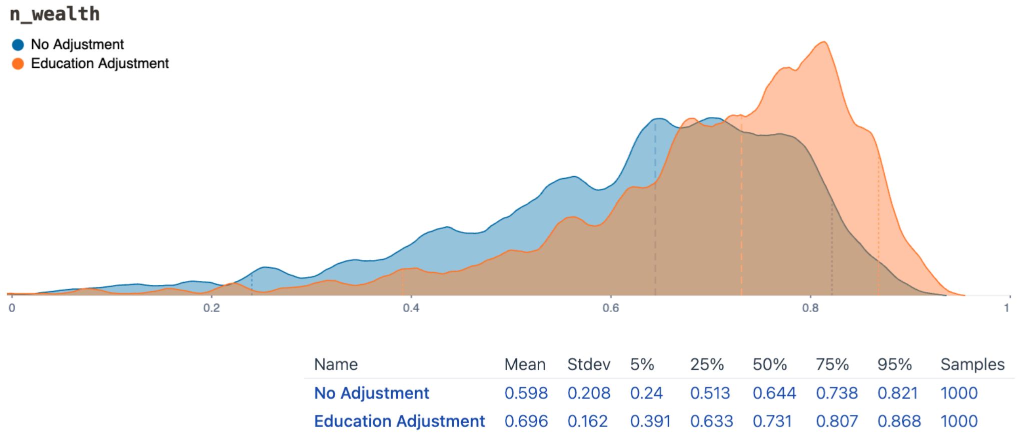
Adjusting Wealth For Potential Education Correlations [1.2X Standalone]

Importantly, we have assumed people have certain educational qualifications, which tend to imply more wealth. Thus, the population statistics above might under-estimate the true values. I could not find direct data on 'percentiles of wealth by education', but note the following adjustments:

- Emmons et al. (2019) [p9] find that the median net worth of four-year college graduate households is $300k compared to $100k of all families – so 3X higher. Unfortunately, they don't report how this shift might look like at other parts of the distribution. We might imagine the lower end increases more or less than that.

- Ruggles et al. (2023) [via DQDJ] do break down income percentiles by education level, and their medians seem to line up with our specific groups quite well (the US median is $50K; Bachelors earn $67K and PhDs $115K).[40] Doing so, we can see incomes increase 4.5X at the lower end and 2.2X at the higher end; for Bachelors 2X and 1.2X.

Overall, I'm still left pretty uncertain about the wealth increase to assume, so I take a wide range of 1-6X. To model this we can see what such a shift in the wealth distribution implies for how many people make it past the threshold rate. We can see it increases from 60% to 70% – i.e. a 1.2X increase.



Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
No Adjustment	0.598	0.208	0.24	0.513	0.644	0.738	0.821	1000
Education Adjustment	0.696	0.162	0.391	0.633	0.731	0.807	0.868	1000

Adjusting Wealth For Potential Age Correlations [0.75X Standalone]

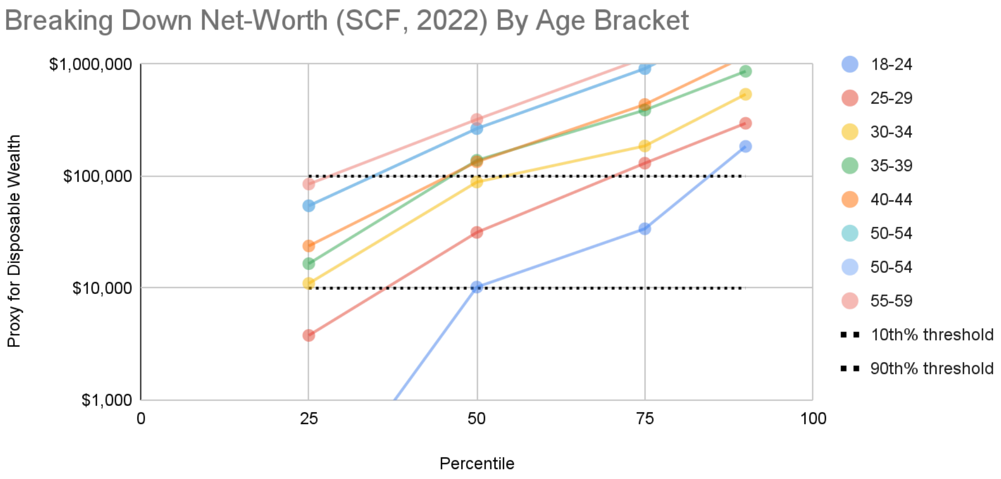
On the other hand, it is worth noting that we have already noted that the relevant STEM and PhD individuals tend to be younger – and younger people tend to have less money. It has likewise been suggested terrorists skew younger still (Russell & Miller, 1977). Thus, population statistics might overestimate the true number. Whilst I will later discount the intention, I want to account for potential interaction effects here.[41] We need to examine two factors: [i] how disposable wealth varies by age; and [ii] what what age-weighting to give to bioterrorists in our group

[40] We might think that STEM Bachelors are even higher than that (NCES, 2019). The NSF (2021) has the median STEM Bachelor of $64k and the NCSES (2020) [Table 50] the median biology PhD $110k/yr.

[41] The same sources also claim that terrorists skew more uneducated and lower income in general. I deal with this separately in I. I estimate the size of a socio-economic group first, then the likelihood of misuse – rather than the number of people committing misuse and then breaking this down by demographics.

The Survey of Consumer Finances (2022) usefully breaks down net-worth estimates by age bracket. (However note that this data, it no longer excludes primary equity, which makes it a slightly worse indicator and isn't directly comparable to the data above. Instead, I am mostly interested in seeing how the relative ratio changes as we change the age ratings.)

By doing so, we see from the graph below that there is a clear link between age and net worth, with 18-to-24-year-olds having an especially high fraction of people below even the $10k threshold. [Recall fn 6 for why net worth is likely still an imperfect proxy that likely lessens this].

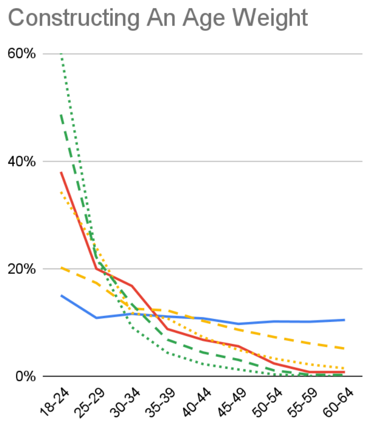


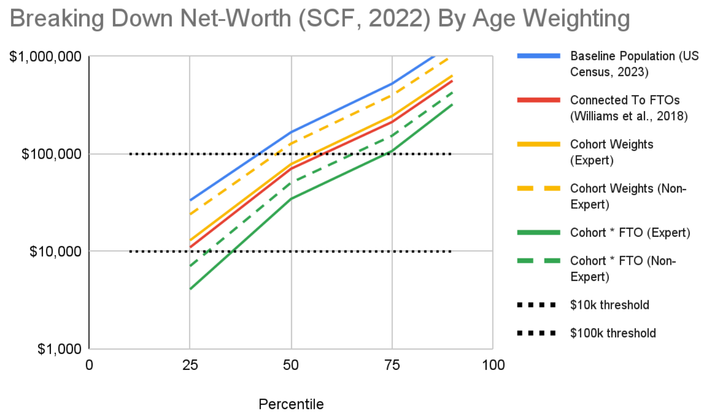
Thus, it matters what fraction of people we think are in the low age brackets versus others. We can then make the following adjustments:

- **'Baseline' Age:** We can begin by taking the US Census (2023), which gives us a baseline for age brackets if we assume that there is no skew at all
- **'Priced In' Age:** We can now overlay the age weights that we previously found in the section above when we discounted cohorts. This gives us a 0.4X-0.5X decrease in wealth for experts and 0.7X-0.75X for non-experts.
- **'Terrorist' Age:** We can now overlay the age weights Williams et al. (2018) find for 476 US citizens who joined Foreign Terrorist Organisations between 2001-2017.[42] If we fully take this haircut we get 0.1-0.25X for experts and 0.2X-0.3X for experts.

(The age weights are shown on the left and the resulting net worth distribution on the right)

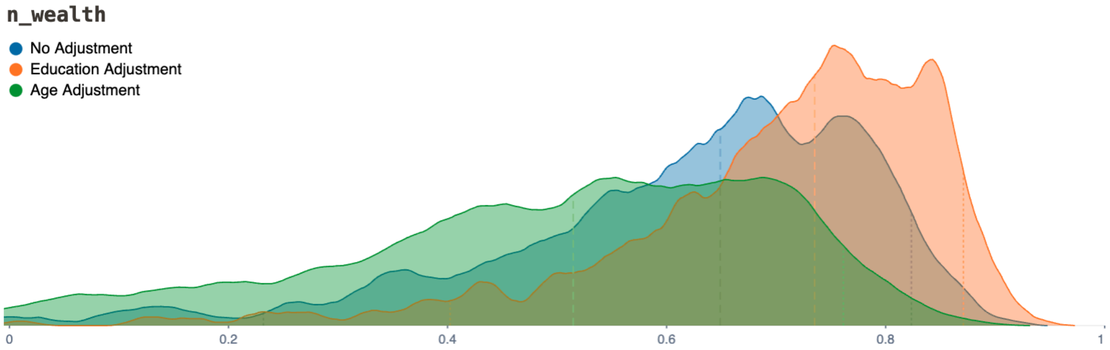
[42] I converted these from a PDF to CSV using Claude, so it may contain some irregularities.





With regard to the last factor, whilst I am sympathetic to an argument that terrorists generally skew young, I also do not want to over-update in the case of lone wolf bioterrorism. Bruce Ivins is likely the most infamous suspected US bioterrorist with the 2001 anthrax attacks, for which he was aged 55 (Wikipedia). Similarly, Ted Kaczynski seems like a prototypical lone wolf who is highly educated and engaged in a sophisticated project; and he was aged 35-55 for his attacks (Wikipedia).

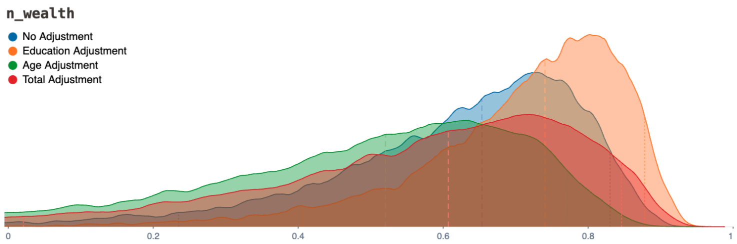
Thus, I only assign this later transformation half weight or so. Thus, I get a 0.25X-0.55X shift in the wealth distribution due to age. We can see that this by itself decreases the fraction of people who make it past the threshold from 60% to 45% – i.e. a 0.75X decrease.



Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
No Adjustment	0.602	0.206	0.232	0.525	0.649	0.742	0.824	1000
Education Adjustment	0.699	0.165	0.402	0.635	0.735	0.808	0.871	1000
Age Adjustment	0.447	0.298	-0.0947	0.343	0.515	0.649	0.761	1000

Final Assumption [0.85X joint]

For 'non-STEM-Bachelors' I choose the Age-Adjustment of 0.45X [0.01 – 0.77]. For both wet lab biology PhDs and STEM Bachelors. I take a mixture distribution with 2/3 'age' and 1/3 'education' that gives me 0.53X [0.02 – 0.85] .[43]

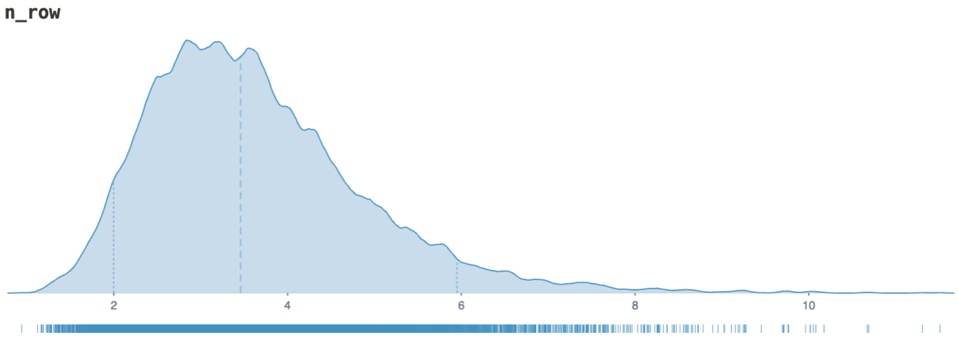


Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
No Adjustment	0.608	0.2	0.235	0.521	0.653	0.745	0.83	10000
Education Adjustment	0.704	0.161	0.406	0.637	0.74	0.812	0.878	10000
Age Adjustment	0.455	0.287	-0.0824	0.333	0.52	0.651	0.772	10000
Total Adjustment	0.542	0.28	0.0208	0.427	0.607	0.732	0.846	10000

Generalising To The Rest Of The World

Reference Classes [3.5X]

We can now 'multiply' this US-based number to get a global estimate. Naively, the US makes up 1/25th of the global population, i.e. a 25X multiplier (World Bank, 2022). However, the US share of people with relevant education and material resources is much higher – so the true multiplier is likely much lower. I consider various reference classes below and assume a 90%-CI of 2-6X multiplier. I also don't have too much evidence to think it greatly differs between experts and non-experts



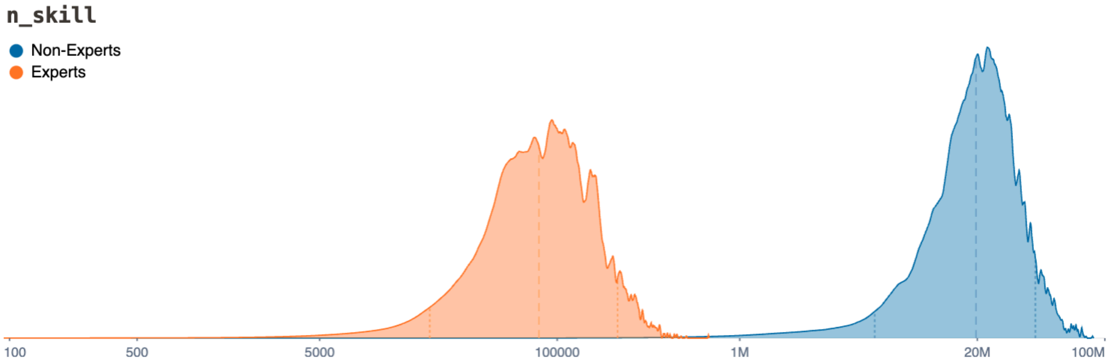
Mean	Stdev	5%	25%	50%	75%	95%	Samples
3.65	1.26	2	2.75	3.46	4.32	5.95	10000

[43] We could not easily find direct data that decomposes both. I choose these given that it seems likely that most of the education increase comes from older people, and we have fewer older people in our sample.

Global Multiplier Reference Classes	Multiplier
Wealth-Based Reference Class	~4X [3-6X]
When adjusting for purchasing power parity, the US has a GDP of $27.4T compared to a global GDP of $184.7T (World Bank, 2023). That gives it a 15% share and thus a 6.6X multiplier. I imagine this is an overestimate since the share of people with PhDs and [potentially] >$100k seems even higher than that.	<7X
The World Inequality Report (2022) reports that "North America & Oceania" contains ~25% of people who are in the global 1th-10th income percentiles ($37k–$124k) [Fig. 1.8] and likewise 15-30% for the global 1th-10th wealth percentiles ($126k-$807k). That gives a naive multiplier of 4x and 3-6x – and perhaps slightly higher when we exclude Canada, Australia, etc.	~4X
The US has a population of ~335M compared to all high-income countries [HICs] having ~1.2B (World Bank, 2022). If we assume HICs have a similar ratio, then that's a factor of 3.6X. On the one hand, we might think the US is still overproportionate. On the other hand, it also seems an under-estimate to exclude non-HICs like China and India.	>3X
STEM-Bachelors-Based Reference Classes	~3.5X [3-5X]
Oliss et al. (2023) report the STEM graduates for the top eleven countries. They find the US has 820k compared to 9M overall, i.e. 11X. This is an overestimate as it doesn't account for wealth. If we exclude China and India (~6M), we get 3.6X. I am no longer sure of the bias since it still includes non-HICs (e.g. Brazil) but excludes HICs like the UK.	~3-4X; <11X
CWUR (2021) reports that the United States contains 1/3 of the world's top 2,000 universities. Förster (2022) finds that 2,000 out of 10,000 listed universities are in the US. These suggest a factor of 5X. It does not adjust for cohort size, PhDs, or biology specifically, which I imagine would drive this down somewhat more.	~3X and <5X
PhD Science-Based Reference Classes	~3.5X [3-5X]
Esvelt (2022) claims that "The U.S. grants 125 doctoral degrees in virology each year, accounting for one-third of the total worldwide" – implying a 3X multiplier. This matches two other interviews I conducted with external biosecurity experts who have attempted to construct similar estimates for private reports.	~3X
The US spends 3.4% of GDP on science R&D compared to the world average of 1.9% (UNESCO, 2021; Table 9.5.1). If the US makes ~25% of world GDP that gives us a naive factor of 2X. This is plausibly an under-estimate as it does not account for purchasing power parity (i.e. the 1.9% of ROW can 'afford' more)	>2X
The US has 4,800 scientific researchers per M compared to the world average of 1,300 (UNESCO, 2021; Table 9.5.2). If the US is ~4% of the world population that gives us a multiplying factor of 6.8X. This is plausibly an overestimate as it does not take into account that amongst researchers, fewer people meet the $-threshold	<5.5X
In Williams et al. (2025) we asked highly credentialed forecasters and subject-matter experts to break down non-natural biological risk by region. It found that both groups estimated that ~20% of risk comes from the 'Regions of the Americas	~5X

[Result] Total Number Of Individuals In Threat Actor Class

Putting all of this together, we thus get the final distributions as follows: 21M [5.5M-42M] STEM Bachelors and 93K [20K-214K] wet lab biology PhDs:



Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
Non-Experts	21.1M	11.2M	5.48M	13.5M	19.7M	27.1M	41.6M	10000
Experts	93.4K	61.4K	20.1K	49.7K	79.5K	123K	214K	10000

[Checking] Comparison To Private Report

Redacted but may be available upon request

2.3 | Perseverance Level, P

Conceptual Framework

Several experts who I spoke to emphasised that 'perseverance' is an important dynamic to capture. I.e. that we shouldn't just think of people having a fixed probability of succeeding at a given step, but that people might try again-and-again. As time passes, they may become more skilled, but are also more likely to become discouraged, run out of money, be apprehended by authorities, etc. Thus, if (say) an AI can mean people can get things done in 3-6 months what might have otherwise taken 6-12 months, then that is an important acceleration.

To illustrate, consider the following stylistic example:

- Suppose there are 10 people who would be willing to put in 1-month of effort to build a bioweapon, but only 2 of those people who would be willing to put in 3-months of effort.
- Suppose with 1-month of effort you are able to only have one attempt, which gives you a 10% success rate. But with 3-months you are able to have three attempts – giving you 50% total.
- Thus, by default we might expect 1.8 successful attempts [=8*10%+2*50% = 1.8].
- Suppose now that AI means you can 'accelerate' to try three attempts in 1-month, so now everyone has a 50% success rate. I.e, we expect 5 successful attempts [10*50% = 5].

Importantly, we can see that the potential for AI to have a large counterfactual 'speed-up' is greater if:

i. The more willing people are to put in low versus high amounts of effort;

ii. The bigger the success rates are conditional on low versus high amounts of effort.

In the next section I will consider [ii.]. Here I want to get a better understanding of [i.] – How persistent should we expect threat actors to be? In particular there are two things to look out for here:

- If people have a very high drop-out rate, then maybe the number of people who are realistic threats is exceptionally low. AI would have to be incredibly powerful to change that.

- If people have very low drop-out rates, then maybe people are overdetermined to succeed. E.g. if everyone was a Ted Kaczynski operating over 15-years they just become experts.

Literature Review

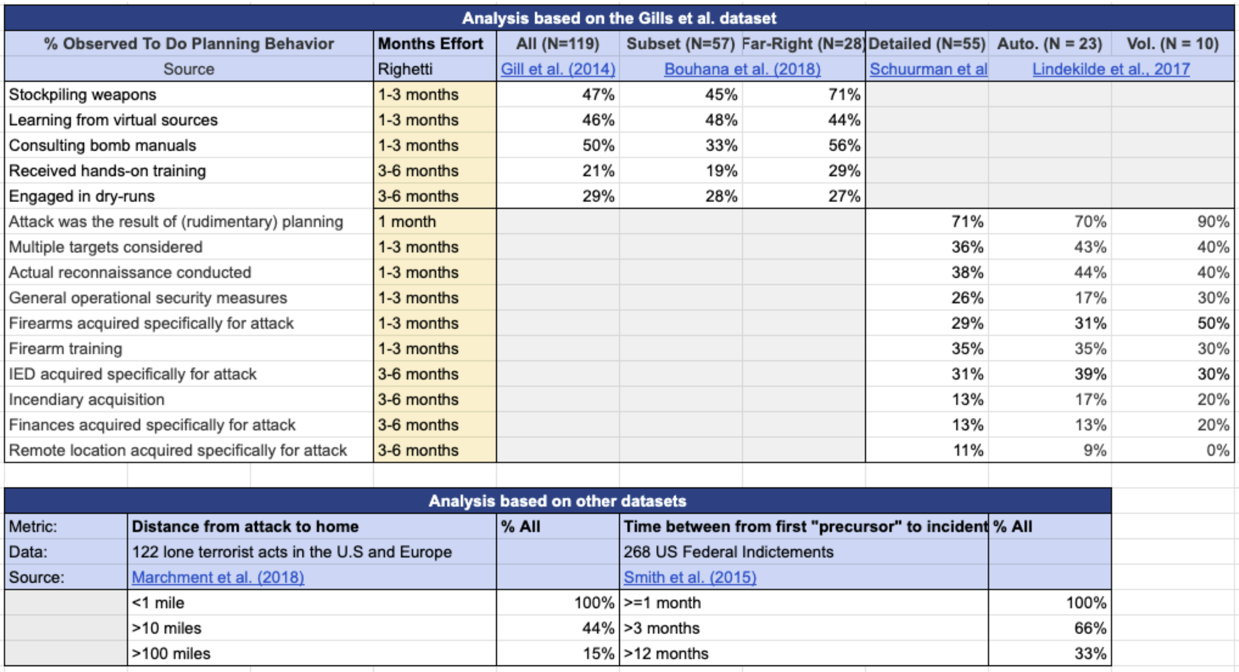
As noted, there is a literature that looks at what behaviour terrorists exhibit in planning their attacks – which we can use to infer how much effort they spent. I examined the key papers mentioned in Kenyon et al. (2023)'s literature review. Most analysis seems to draw on the same underlying dataset from Gill et al. (2014) – examining lone wolf attacks in North America and Europe between 1986–2015 – which was then given more detailed coding by Schuurman et al. (2018).

Whilst these insights are useful, I flag a few different issues that readers should keep in mind:

- **Limited datapoints.** Much of the literature itself highlights that the case studies they draw on are often too few to reach strongly generalizable insights, such as to do statistical tests. This appears an inherent feature of terroism being rare, let alone biological terrorism.

- **Databases plausibly self-select for higher effort threat actors.** E.g. threat actors who might spend 1-month planning but then "give up" may go undetected and thus not make it into such analyses. Thus, potential lone wolves might on the whole be 'lower effort' than suggested..

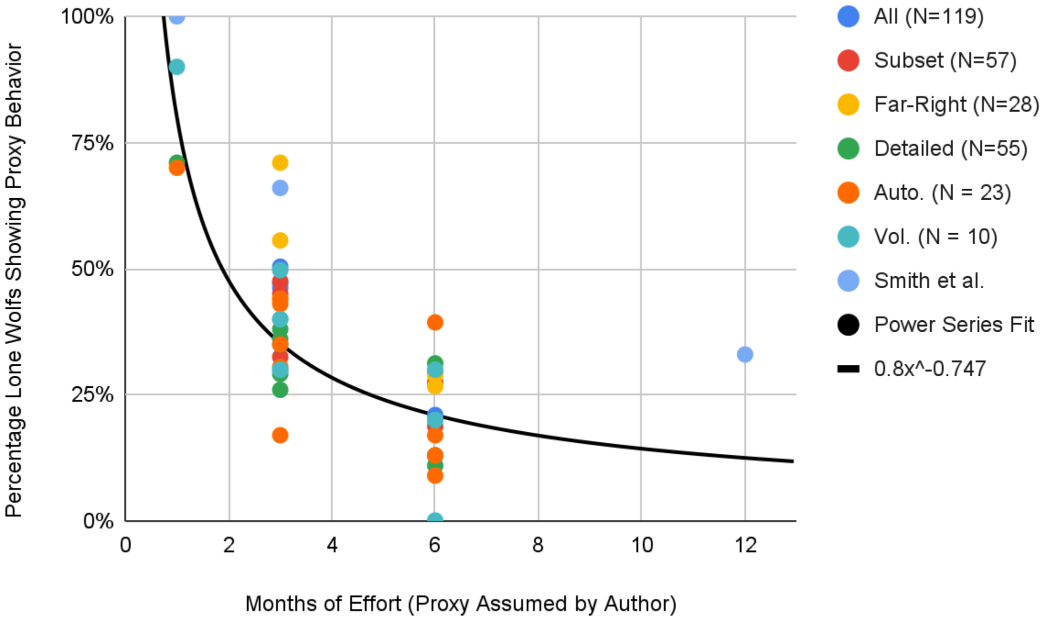
- **Actors motivated by biological weapons may be 'higher effort' to begin with.** E.g. threat actors pursuing a biological weapon may be more motivated than usual. This is perhaps an extension of studies finding 'autonomous' lone wolves plan attacks for longer and pursue more destructive weapons (Lindekilde et al., 2017).



Analysis based on the Gills et al. dataset							
% Observed To Do Planning Behavior	Months Effort	All (N=119)	Subset (N=57)	Far-Right (N=28)	Detailed (N=55)	Auto. (N = 23)	Vol. (N = 10)
Source	Righetti	Gill et al. (2014)	Bouhana et al. (2018)		Schuurman et al	Lindekilde et al., 2017	
Stockpiling weapons	1-3 months	47%	45%	71%			
Learning from virtual sources	1-3 months	46%	48%	44%			
Consulting bomb manuals	1-3 months	50%	33%	56%			
Received hands-on training	3-6 months	21%	19%	29%			
Engaged in dry-runs	3-6 months	29%	28%	27%			
Attack was the result of (rudimentary) planning	1 month				71%	70%	90%
Multiple targets considered	1-3 months				36%	43%	40%
Actual reconnaissance conducted	1-3 months				38%	44%	40%
General operational security measures	1-3 months				26%	17%	30%
Firearms acquired specifically for attack	1-3 months				29%	31%	50%
Firearm training	1-3 months				35%	35%	30%
IED acquired specifically for attack	3-6 months				31%	39%	30%
Incendiary acquisition	3-6 months				13%	17%	20%
Finances acquired specifically for attack	3-6 months				13%	13%	20%
Remote location acquired specifically for attack	3-6 months				11%	9%	0%

Analysis based on other datasets				
Metric:	Distance from attack to home	% All	Time between from first "precursor" to incident	% All
Data:	122 lone terrorist acts in the U.S and Europe		268 US Federal Indictements	
Source:	Marchment et al. (2018)		Smith et al. (2015)	
	<1 mile	100%	>=1 month	100%
	>10 miles	44%	>3 months	66%
	>100 miles	15%	>12 months	33%

I try to fit these behaviours onto how much effort they might indicate. This subjective, but in my mind does give some indication of a power series where each “step” in difficulty sees a 0.5X drop off: from ‘low’ ~80% [1 month?] → ‘low/medium’ ~40% [2 months?] → ‘medium’ ~20% [4 months?].



Reference Classes

> *I thank Rose Hadshar for most of the data and analysis here.*

Given some of the limitations in just relying on the sources in the above literature review, I also wanted to consider three other approaches:

- **Judgemental Forecasts:** Surveying a group of subject-matter experts and highly-credentialed people to directly answer this question. An issue is that these are very 'black box'
- **Terrorist-Based:** Going through case studies of how effortful previous terrorist attacks were, including bioterrorism specifically. An issue is it requires a lot of 'judgement'.
- **Other Contexts:** Looking at the 'survival functions' for tasks that are better studied even if very different. An issue is lack of external validity. However, they can help to sanity check.

These results are summarised in the table below, with further explanation being given in the table. Overall, we can see that the forecasts align almost exactly with my subjective judgments of the Gil et al. literature (which I did independently before knowing the results) – and towards the "quicker decay rate" end of the reference classes, but still within bounds.

Forecast	% Remaining In The Month...				Description
	3	6	12	24	
Judgemental Forecasts					
Subject-Matter Experts	50% (20-72%)	20% (8-46%)	8% (2-27%)	n/a	Williams et al. (2025) asked: "What proportion of lone wolves would invest at least X months on developing a biological weapon?" It is of course unclear whether such numbers turn out to be accurate, but it is one way to get a crowdsourced opinion.
Highly Credentialed Forecasters	50% (50-88%)	25% (15-63%)	8% (3-20%)	n/a	
Terrorist Based Indicators					

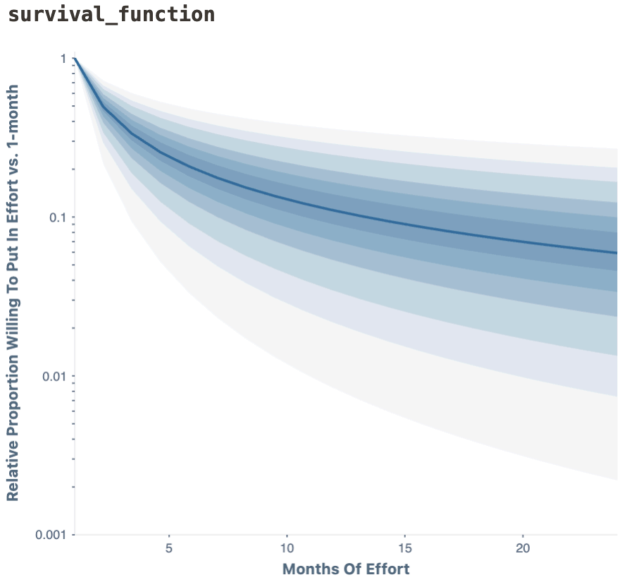
Bioterrorist Plots	>46%	>38%	>23%	>8%	Rose Hadshar went through 13 case studies of CBRN terrorist plots to assess post-WW2 to assess how much time these plots likely took [key sources were: Carrus (1998), Appendix A; Pilch and Zilinskas (2005), and all incidents in NASEM (2024)'s Appendix F]. Importantly many were 'stopped' because they got arrested, not because they were discouraged. So this is plausibly an under-estimate.
Lone Wolf Plots	>67%	>33%	>7%	>7%	Rose Hadashare went through 14 case studies of post-9/11 lone wolf plots, mostly with explosives [based on Hamm and Spaaj (2015)]. Again, many 'stops' were due to arrests.
Other Context Based Indicators					
Health Clinical Trials Drop-Outs	<71%	<44%			Landers and Landers (2009) look at how test-subjects drop out of a dieting study. Terrorists will have harder conditions, so I take this as an upper bound
New Year's Resolutions	<65%	<49%	<48%	n/a	Oscarsson et al. (2020) [Table 3] look at how many people are able to fulfil their new year's resolutions, which tend to have to do with dieting and health. This time they are self reports, so I expect actual perseverance to be lower. On the other hand, terrorists may be more motivated. Overall, I take this as an upper bound
Completing Massive Open Online Courses	64%	>9%	-	-	Jordan (2019) [Figure 5] looks at how many people complete courses depending on their length. On the one hand terrorists have a harder task; on the other I imagine they're more motivated. I take it as a lower bound

[Result] Rate Of Individuals With Meaningful Intent At Different Levels Of Effort

Overall, all of these estimates seem very well approximated by power-series [all have a R2>0.74 and 8/11 are >0.9]. Thus, I overall feel okay modelling this dynamic as a power-series with an exponent that we are uncertain over.

Looking at the different estimates it seems that the judgemental forecasts are actually fairly similar to the reference classes that I was able to find data on. Interestingly they imply an exponent pretty close to Zipf's law. Although I strongly caveat that we are fairly ignorant here, the uncertainty interval here seems appropriately wide.

As a best guess, I choose value "-1" with a 90% CI range "-1.55 to -0.5". In layperson's terms means that after 1-month of effort, for each doubling in time, 0.5X [0.35X–0.7X] people drop out. So if we begin with exactly 1-person at 1-month, for 2-months it's 0.5 [0.35 -- 0.7], for 4-months 0.25 [0.12 -- 0.5], for 8-months 0.125 [0.04-0.35] and so on.



2.4 | Laboratory Success, L

Defining A Difficulty Threshold For Synthesising A Dual-Use Virus

Level Of Virus Difficulty

For this variable, we want to know how likely a person is to synthesise a dangerous virus successfully. But to answer this question, we need to disentangle two components: [i] how 'difficult' is the dangerous virus in question, and [ii] how much might an individual be able to outsource steps to the synthetic biology ecosystem. I consider these in turn. I consider these in turn.

Importantly, molecular biology procedures differ across virus families (but not so much virus strains).[44] I.e., some viruses are easier to construct than others. Several factors affect this, but there is no easy rule to translate viruses into a single dimension of 'difficulty'.[45] Experts I spoke to generally agreed to a general ordinal ranking, something like "adenovirus" < "influenza" ~< "sars" < "smallpox". But there was great disagreement on exactly how pronounced these differences are.[46]

	Characteristics relevant to how hard it is to assemble a virus					Expert difficulty
Virus	Material	Strand	Sense	Genome	Segment	
Adeno	DNA	Double	n/a	36kb	1	Low
Influenza	RNA	Single	Neg.	13kb	8	Low-Med.
Sars	RNA	Single	Positive	30kb	1	Med.
Variola	DNA	Double	n/a	190kb	1	High

[44] I.e. the difference in difficulty between Influenza A and Influenza B might be small; but the difference between either an a coronavirus strain large

[45] Factors include:

- Viruses with larger genomes tend to be harder to work with. Intuitively, the larger the genome, the more fragments need to be stitched, creating a higher chance of failure. A genome with >30k base pairs is often cited as a difficult threshold (NASEM, 2018, [p40]).
- Viruses that consist of RNA tend to be harder to work with than those that consist of DNA. Intuitively, RNA is typically 'single-stranded' whilst DNA is 'double-stranded', making the former a less stable genetic material and thus more fragile to do experiments with.
- Viruses that consist of negative-sense RNA tend to be harder to work with than those that consist of positive-sense RNA. Intuitively, negative-sense RNA is 'backward' from the way enzymes need to read it, so they need to be flipped first, which requires more steps.
- Viruses that need to be inserted into mammalian host cells tend to be harder to work with. Intuitively, mammalian host cells are much more difficult to keep healthy and alive than other hosts – so the likelihood of them being a good environment for the virus to form in is lower.

[46] For example, one person said "I would expect adopting a protocol for Rabies, COVID, flu, etc. to be ten times more difficult than adenovirus" whilst another said, "Making a defective adenovirus is a close proxy to flu for all that counts. It seems possible that specific flu experts are overvaluing their specific skills."

It thus actually matters where the 'scary line' is. We don't want to set this too low. To take a stylised example, it could be that a non-expert has a (say) 50% chance of synthesising an adenovirus. But these are also mostly harmless (CDC) and anything pandemic scary is much harder. So their 'true' success rate is lower. Yet we also don't want to set it too high. A nonexpert might have a 0.05% chance of synthesising smallpox – which was only done more recently and taken. Yet, whilst smallpox is scary (Johns Hopkins, 2001; NASEM, 2024), there are also many other threats that come well before – such as 1918 influenza. So their 'true' success rate is much higher.

In theory, we might want to estimate the 'laboratory success rate' [L] separately for each virus and then weigh it by how pandemic-credible each virus family is [E] and thus, in turn, how likely an average threat actor is to pursue it. See the stylised table below. However, this is too elaborate for this report, and the E exercise is especially information hazardous.

Stylistic Example	Weight	B	E	B*E
Influenza	70%	5%	5%	0.25%
Smallpox	25%	0.05%	50%	0.025%
etc.				

Instead, I will take a single threshold to estimate the overall success rate, which is "synthesising an influenza strain from DNA fragments with deliberate mutation". I justify this decision accordingly:

- 1918 Influenza (and to a lesser degree H5N1) is already known to be potentially dangerous. Bad actors will gravitate towards the 'easiest credible seeming route', so these carry weight;
- Some experts I spoke to mentioned other viruses they were worried about, but – with the exception of smallpox – thought that flu was a decent proxy in the right order of magnitude.

However, others may disagree and I qualitatively increase my uncertainty. Note that this uncertainty suggests that doing so might be harder than the threshold my question is posing.

Level Of Outside Help

Biosecurity experts have warned that threat actors could perhaps outsource some of the hard steps of building a virus beyond ordering DNA fragments. For example, with influenza, there are third-party providers that 'pre-assembled' influenza plasmids, which is easier than stitching together DNA fragments yourself (Soice et al., 2023). More generally the WHO (2021) describes how cloud laboratories and contract research organisations can be contracted to "de-skill the research needed by reducing some of the knowledge requirements for conducting sophisticated research protocols."

I spoke to people who agreed that a threat actor able to use such services would need somewhat less biological skill and thus have a noticeably higher laboratory success rate. Such outsourcing pathways are concerning and have thus become an increased focus of regulation (U.S. NSTC, 2024). However, there are three important nuances that prevent this from lowering the skill threshold too much:

Firstly, such third-party vendors check orders and would know if they are being asked to build something dangerous. Such organisations do have 'know your customer' policies in place, and whilst these aren't standardised, a lot appear non-trivial – involving requesting written proposals and phone calls. Not having an official academic or business procurement contract can be very problematic. A threat actor risks getting caught [see O].

Secondly, as a result, a threat actor would have to obfuscate their orders to make it look more benign – which requires biological skill to 'undo'. For example, they might request two assembled genomes that are close to something dangerous but 'slightly wrong' in their own way – which they would have to fix themselves. This can still be easier than making a genome from fragments – but note that it would still require some biological skill to then cut and ligate these two genomes together (Soice et al., 2023). Designing such an 'obfuscated' order also requires its own kind of skill. Outsourcing lowers the skill barrier too much.

Lastly, (on a more minor note) there are limited pathogens for which such outsourcing pathways would work in the first place. For example, virologists regularly order pre-assembled plasmids for influenza, so there is an existing ecosystem in place that a threat actor might try to 'piggyback' on. But the same does not exist for other viruses – such as smallpox, which is both [i] made up of one very large plasmid that would be much harder for anyone to build; and [ii] much strictly regulated.

Overall, I thus think that "synthesising an influenza strain from DNA fragments with deliberate mutation" is probably the right threshold. Again, others may disagree and I do qualitatively increase my uncertainty. Note that this uncertainty suggests that doing so might be easier than the threshold my question is posing – and thus also somewhat cancels with the section above.

Description Of The Difficulty In Biological Steps

So, say a threat actor has a lab with the relevant materials and identifies the correct sequence. What concrete problems will they face? As I understand it, biological challenges can be categorised as follows. The hardest appears to be 'troubleshooting' and 'tacit knowledge'.

Identifying The Correct Sequence To Follow: In order to assemble a specific pathogen you need to know the specific DNA it consists of. Increasingly, such genome sequences are now publicly accessible that most experts I spoke to do not consider this a bottleneck anymore. However, I would still note two points of nuance that point at some imperfections:

- NASEM (2018) notes that when scientists create novel RNA viruses, these mutate as they replicate. Thus, scientists end up with a 'mixture' of slightly different copies. It can thus be unclear when the scientists deposit the sequence of a single member of that mixture into a [public] database "it is possible that the starting sequence may not generate a "wild type," fully virulent population after booting".
- Rasmussen (2022) notes that "most "complete" viral genome sequences actually have pretty poor coverage at the ends and in highly structured regions."

This is a particular problem because the feedback loops here are very slow. Somebody trying to create a pathogen won't be able to tell if the sequence they chose actually results in a virulent population or if they have to restart until the very end.

Identifying The Correct Protocol To Follow: Synthesising a virus involves many steps of specific instructions, including what biological materials to use, what temperature to set them at, and how long reactions take. Several of these protocols can be found online, in books, or ready-to-use kits (e.g. the publication of the methodology of synthesising horsepox (Noyce et al., 2018) and SARS-CoV-2 (Xie et al., 2021)) both created controversy for just how detailed they were (see Koblentz, 2017, and Pannu et al., 2021, respectively).

Sometimes, specific knowledge is kept purposefully 'concealed' (Collins, 2010). In the case of H5N1 gain-of-function experiments, the US National Science Advisory Board for Biosecurity recommended "the basic result be communicated without methods or details, [so] that the benefits to society are maximized and the risks minimized" (Berns et al, 2012). Historically, there was deliberate withholding of information about creating anthrax weapons even amongst different Soviet laboratories – likely due to competition between them (Vogel, 2006).

Other times, information that is 'logistically demanding'. I.e. it can be shared but might not be because describing procedures in detail is a lot of work. Scientists may choose not to spell everything out, especially since their intended audience is already familiar with some basic concepts and they want their article to remain readable. A protocol might say "Prepare the plasmid transfection cocktail in 50 µl of OptiMEM media" but it might be unclear to a non-expert what exactly "prepare" means here.

Notably, when I spoke with molecular biologists, they often had much more detailed 'private' lab-notes describing protocols that were shared within their research groups than what I was able to find available in publications. Still, on the whole, experts I spoke to did not think this barrier was insurmountable for pathogens of concern, but they did note it might trip some non-experts up. In particular, the increasing existence of online videos has done a lot to lower this barrier (e.g. JOVE).[47]

Protocol 'Troubleshooting' To Apply To Own Setting: A more serious issue is that even if somebody begins with a corresponding protocol, these still often require modifications and troubleshooting to make them work for the specific lab environment. As DeBenedicts (2023) notes: "Often if you're trying a new protocol in biology you may need to do it a few times to 'get it working.' It's sort of like cooking: you probably aren't going to make perfect meringues the first time because everything about your kitchen – the humidity, the dimensions, and power of your oven, the exact timing of how long you whipped the egg whites – is a little bit different than the person who wrote the recipe."

Part of why this appears so problematic for individual threat actors is not just that it suggests an individual might have to give it several attempts – but also that overcoming this requires some underlying understanding of the biological processes, not just following instructions. Somebody would have to [i] notice that something in their process is going wrong (else they risk wasting too much time and materials proceeding), [ii] correctly diagnose what caused this (or at least have a reasonable guess), and [iii] know what to change about their approach (Ouagrham-Gormley, 2014).

Somatic skills to execute the protocol: Even if somebody has a correct and reasonably detailed protocol, some experts still argue this might be insufficient due to the importance of 'tacit knowledge' (Revill & Jefferson, 2013) – skills that are acquired through experience and cannot just be passed down in writing. As MacKenzie and Spinardi (1995) note: "Most of us, for example, know perfectly well how to ride a bicycle yet would find it impossible to put into words how we do so."

In the context of molecular biology we might consider the following examples:

[47] As Revill and Jefferson (2014) explain, this should also not be overstated: "Because video protocols only reveal as much as the editors deem necessary, it is possible that some information is lost in the editing process, not least because completely following every single step in the process would make a rather boring video." The authors also note that such videos don't exist for biological weapons. This is true for things like anthrax. However, reverse genetic protocols for virus families do exist, and these can be repurposed if a pandemic credible agent is found.

- An intuitive sense for measurements, such as to “crush the cells with just the right amount of pressure” (Ouagrham-Gormley, 2014) or using a Dounce homogenizer machine (Vogel, 2013)
- Motor skills to ‘pipet’ which requires using the correct angle, immersion depth, constant rhythm, and more (Mettler-Toledo, 2013)
- General ‘good laboratory practice’, such as sterile technique or upkeep of instruments and materials – which can require a ton of small behaviour (Cell Signalling, 2022)

Historically, Danzig et al. (2012) proposed that such lack of tacit knowledge may have been what caused Aum Shinrikyo [a Japanese bioterror group in the 1990s] to have have failed at successfully following a protocol for plasmid insertion to turn anthrax vaccine strain pathogenic. The report suggests that it took a skilled researcher six months of practice in a leading laboratory to be able to successfully learn how to perform this procedure. Ouagrham-Gormley, 2014 similarly describes the failure of many bioweapon programmes in the 1980s and 1990s to tacit knowledge.

However, I also want to be somewhat cautious about thinking this barrier is impossible to overcome. In particular, synthetic biology has just advanced a lot in ways that ‘erodes’ the tacit knowledge needed to complete procedures (Revill & Jefferson, 2014). For example, E-gel devices now let people purify DNA in “as little as 10 minutes” when this used to be much more elaborate (Thermo Fisher, 2018; Addgene, 2019). As some experts note, new technology can require its own new skills (Ouagrham-Gormley, 2014), but on the whole I think all the advances we have seen in synthetic biology (Meng & Ellisa, 2020) should make us review the extent of tacit knowledge barriers.

Expert Discussion

As Revill and Jefferson (2014) note “the ease through which individuals can synthesise life remains contested”. Here, I briefly collect some quotes regarding what different experts think about the difficulty for non-experts and experts to complete reverse genetic protocols.

On, the one hand, several experts have raised concerns:

- The WHO (2015) convened a scientific working group, which concluded that “it would be possible to recreate variola virus, and that this could be done by a skilled laboratory technician or by undergraduate students working with viruses in a relatively simple laboratory”. The hardest part would be assembling DNA fragments. See diagram below.
- NASEM (2018) put the concern for re-creating known pathogenic viruses at ‘medium’:

  - "The production of most DNA viruses would be achievable by an individual with relatively common cell culture and virus purification skills and access to basic laboratory equipment" and " the level of skill and amount of resources required to produce an RNA virus is not much higher"
  - "The J. Craig Venter Institute (JCVI) was able to develop a viable seed stock within just 3 days of learning the sequence of a new strain of influenza A virus (a negative-strand virus). Although JCVI has extensive resources and expertise that would not be available to every actor, the demonstration nonetheless underscores current capabilities regarding booting both DNA and RNA viruses."
  - But "depending on the resources and expertise available to the actor, there may be difficulties in building and testing a fully virulent RNA virus."
- Esvelt (2022) noted in his congressional testimony that "A large number of scientists, engineers, and lab technicians have the skills required to obtain many types of infectious viruses from publicly available genome sequences". He suggested maybe >1-in-10 would receive training mammalian cell culture might succeed and >1-in-20 life science PhDs would.

On the other, several experts have countered this conclusion:

- Rasmussen (2022) replied to the Esvelt claims by noting "there are 1000s of virologists but far fewer with these skills. We aren't cooking up novel viruses all the time for several reasons [...] reverse genetic systems are *very* technically challenging. I spent half my PhD trying to get an infectious clone of rhinovirus!"
- Carter et al. (2023) noted that
  - Assembly: "Experts familiar with assembly of viral genomes argued that an individual or group with basic molecular biology skills (including bacterial and yeast culture) could likely assemble a viral genome that was 10,000–12,000 base pairs. Assembly of larger viral genomes (up to 30,000 base pairs) requires additional know-how, including virus-specific expertise and troubleshooting capabilities, and is thus more likely to be a group effort." [Influenza is 13,000 base pairs]
  - Booting: "A few experts expressed concern that the ability to boot up viruses might someday become broadly accessible or even available as a kit. Still, most experts argued that in most cases, generating an infectious agent from viral genomes would continue to be challenging and would require virus-specific expertise and training".

- In DeFrancesco (2021) a synthetic virology expert noted on the horsepox experiment that "[saying] you don't need exceptional biochemical knowledge or skills [...] would then mean that [...] being one of the leading orthopox researchers in the world [like David Evans] isn't exceptional". David Evans himself later noted "the skill set needed to do this work requires advanced scientific training, insider knowledge, and infrastructure" (Noyce & Evans, 2018) – although it is worth noting horsepox is a more complex pathogen.

- An interviewee I spoke to with a background in synthetic biology noted that the knowledge is "taught in undergraduate degrees but to execute the experiments you'd need a level of independence found at the post doc level." A critical part of the experiment is high variance – whereby some PhD students get it on first try or post-doc does the one crucial step for them. When asked to give a range they said most people might need 9-12 months to learn the skills.

Recreating known pathogenic viruses				
	Usability of the technology	Usability as a weapon	Requirements of actors	Potential for mitigation
Level of concern for re-creating known pathogenic viruses	High	Medium-high	Medium	Medium-low
Making existing viruses more dangerous				
	Usability of the technology	Usability as a weapon	Requirements of actors	Potential for mitigation
Level of concern for making existing viruses more dangerous	Medium-low	Medium-high	Medium	Medium

Figure A.2.3 | Summary Of Concern For Re-creating And Making More Dangerous Viruses. Table is recreated from NASEM (2018). Note that for Potential for Mitigation, "the concern level is higher for viruses that spread rapidly and efficiently." I.e., pandemics.

Estimating The Difficulty

Human Error Reference Classes

To help sanity-check these numbers we can also try to look at some examples outside of reverse genetics, which help us to calibrate these numbers. In particular, the Human Error Assessment & Reduction Technique is an established methodology for "evaluating the probability of a human error occurring throughout the completion of a specific task".

I apply the 'generic task' estimates suggested by EPD and also spend some time looking for potential biology-specific and non-biology analogies to help sanity check these answers. Details on each of these can be found in the longer table below.

Reference Classes	Base Rate
Success Rates	
Generic: HEART suggests that the error rate as follows: [E] Routine, highly practiced, rapid task involving relatively low skill: 2% (0.7%–4.5%) and C: Complex task requiring high level of comprehension and skill: 16% (12%-28%)	16% error rate per step for experts. If 14 step protocol get 9% success
Looking through a reverse genetics protocol, I note that there are 14 complex 'steps' – but that breaking these down further I get approx. more 'routine' 70 individual actions, which I round up to 100. I consider if molecular biology could be described as either.	2% error rate per action for experts. If 100 get 13%
Biology: Moni et al. (2007) find that 77-90% of first year science undergraduates could operate a micropipette sufficiently on their first attempt when taught by graduate students. On the one hand, people in real life would have more than one attempt. On the other hand, this is just one of the basic skills in molecular biology. But we can construct a lower-bound from this:	<10% error rate per action for non experts. If 100 actions in protocol get >0.003% success
iGEM is an annual synthetic biology competition for high-schoolers and students. Around half of participants win gold medals that are not limited in number. If we assume that a reverse genetics protocol is harder than these projects *and* iGEM selects for people disproportionately good at molecular biology, this gives an upper-bound	<50%(??) protocol success for non experts over 6-12 months?
I could not find good data to assess differences in 'quality' amongst PhDs. Potential approaches include 50-85% of STEM PhD candidates succeed (Glorieux et al., 2023; Duke, 2016); 20% of research project applicants receive NIH grants (NIH, 2024); "anecdotal evidence suggests" 30% of postdocs receive tenure (Bonnetta, 2008)	>20%(??) protocol success for experts
Non-Biology: Fabri and Zayas-Castro (2008) looked at 9,830 medical surgeries and found that "The overall complication rate was 3.4%. Overall, 78.3% of the complications were reported to be related to a medical error." This gives an error rate of 2.7% per surgery. Presumably a molecular biology protocol is harder and has less visceral 'stakes'.	>3% error rate per step for experts. If 14 step protocol get <65% success
Martin et al. (2012) look at how "Individuals without previous experience in ultrasound" are able to complete examinations using virtual guidance as their only training tool. With a small sample (2* n=10 and 2* n=9) they find a failure rate of 6-20%. Presumably a molecular biology protocol is a harder procedure to complete than this, so we can use this as a lower-ish bound for non-experts.	>11% error rate per step for non experts. If 14 step protocol get <20% success

Error Rates Differences Between Experts And Non-Experts	
Generic: HEART suggests the ‘maximum increase in unreliability’ as follows: 3X if operator inexperienced (e.g. a newly qualified tradesman, but not an ‘expert’) 2X if a mismatch between the educational achievement […] and the requirement 1.6X if a need for absolute judgements which are beyond the capabilities or experience Some or all of these might apply. Molecular biology might be more or less good.	Non-experts have ~4X [1.6–9.6X] higher error rate per action than experts
Biology: Kim et al. (2024) conduct the first ‘human reliability in the life sciences laboratory’. They find people with little laboratory experience make pipetting errors 1-in-4k times, whilst those with sig. do 1-in-8k times. Pipetting is one of the more routine activities where experts have less of an edge (see also Lippi et al., 2016). I.e. I’d expect a larger ‘true’ error rate and expertise gap for molecular biology as a whole.	Non-experts have a >2X higher error rate per action than experts
Non-Biology: Jarvis and Harris (2008) break down the accident rate of UK glider pilots with different levels of experience across six flight phases. Overall, they pilots with under 10 hours of experience have an accidence 1-in-6.3k times; whilst those with over 10 hours do 1-in-29.2k times – i.e. 4.6X lower. A 10-hour threshold seems too low a dividing line, but I am actually not confident which way this cuts.	Non-expert have 4.6X higher error rates per action than experts

DIFFERENT POTENTIAL REFERENCE CLASSES FOR “ERROR RATE”					
Source	Error	For Each X		Success	
Expert (i.e., wet lab biology PhD)					
HEART suggests routine tasks have an error rate of 2% (0.7%–4.5%)	2%	100	Actions	~	13%
HEART suggests complex tasks have an error rate of 16% (12%–28%)	16%	14	Sections	>	9%
Fabri and Zayas-Castro (2008) suggest 2.7% of medical surgeries have accidents	3%	14	Sections	<	65%
Kim et al. (2024) suggest pipetting errors happen 1-in-8k times	0.01%	1,000	Sub-Actions	<	88%
Jarvis and Harris (2008) suggest gliders have accidents 1-in-30k flights	0.00%	1,000	Sub-Actions	<	97%
Non-Expert (i.e., STEM Bachelor)					
HEART meta-analysis suggests non-experts are 4X worse at routine tasks	8%	100	Actions	~	0.02%
HEART meta-analysis suggests non-experts are 4X worse at complex tasks	64%	14	Sections	>	0.00%

Martin et al. (2012) suggest novices fail ultrasound exam 6–20% of the time	11%	14	Sections	<	20%
Moni et al. (2007) find 10–23% of BScs fail at micro-pipetting the first time	0.03%	1,000	Sub-Actions	<	78%
Jarvis and Harris (2008) suggest novice gliders have accidents 1-in-6k flights	0.02%	1,000	Sub-Actions	<	85%

2.5 | Operational Success, O

Description Of The Difficulty At Operational Steps

From speaking to experts, I have the impression there are three main operational challenges:

- **Obtaining synthetic DNA for a specifically dangerous virus**, when this is currently “produced by centralized providers that screen their customers” (Carter et al., 2023);
- **Obtaining general equipment to construct a ‘garage’ laboratory** that can then build a virus from this DNA – or covertly using an existing shared facility (DeFrancesco, 2021);
- **Learning additional skills may require searching the web or contacting professionals**, who could alert authorities as part of their ‘culture of responsibility’ (PHE, 2015; NSABB, 2011).

I now want to consider how likely both experts and non-experts are at each of these challenges, which in turn contribute to an overall success rate. As discussed in the main text, I have already conditioned on laboratory success. I want to avoid a multiple-stage-fallacy when choosing a final number.

Case Study: Synthetic DNA Material

DNA material can be ordered directly via gene synthesis companies or indirectly via non-traditional providers like Contract Research Organisations (U.S. NSTC, 2024). Doing so is plausibly the most difficult operational step, given it requires going through centralised DNA providers, many of whom screen orders or have know-your-customer policies – unlike other biological equipment (U.S. HHS, 2023).[48] Still, such safeguards seem at best mixed. Whilst there is limited public analysis, I note the following:

Experts suggest that currently a non-expert obtaining dangerous DNA is at most 'difficult'. The Defense Science Board (2009) said "The single overarching finding of this investigation is that a determined adversary cannot be prevented from obtaining very dangerous biological materials intended for nefarious purposes" and referenced synthesis as one of the easier routes. Nicholas Evans stated, "I doubt the actual physical materials would be hard to get". David Evans noted, "Academic institutions and commercial industries would have few difficulties, private individuals much more so" but still did not rule this out more strongly (DeFrancesco, 2021).

This is driven by the fact that screening orders for synthetic materials are voluntary. There is no evidence of laws requiring laboratories to follow those guidelines in any country (Piper, 2020). The International Gene Synthesis Consortium has 34 members that together screen orders, and reportedly make up 80% of global commercial capacity (IGSC, 2017) – although that figure is "little more than an educated guess" (Schulson, 2023). It would not be hard for a threat actor to find a provider that does not screen. Moreover, Kane and Parker (2024) "observed significant heterogeneity in security practice throughout the field" – including whether suspicious orders just get rejected or reported to law enforcement.

It is unclear if recent policy changes will do much to strengthen screening. The U.S. National Science and Technology Council (2024) announced its intention to strengthen synthesis screening of both direct and indirect providers. However, its enforcement relies on requiring research projects that receive federal research funding to buy DNA from compliant companies.[49] Federal funding only makes up ~61% of biological and biomedical science (NCSES, 2021; Table 12). Whilst this would likely have some effect, there is potentially enough non-federal demand for at least some non-compliant companies to continue to operate –

[48] Unlike general equipment, DNA material can be more easily identified as to whether an order has dual-use concerns. For example, an order for a tissue culture hood could be used for any kind of experiment, the vast majority of which would be non-dangerous, so screening each order might be too costly. By contrast, an order for DNA similar to 1918 influenza clearly has more narrow uses-casts, so there is more justification to check.
Secondly, DNA synthesis is a more centralised industry making enforcing screening easier (at least for now). This feature may change in the future as "new benchtop DNA synthesis devices will enable users to obtain synthetic DNA more rapidly by synthesising it in their own laboratories" – but this still appears 5-10 years away. Carter et al. (2023) describes that currently benchtop synthesis devices can reliably print DNA up to 200 bases in length, which is too short to credibly be used for most viruses. For reference, influenza has a genome of ~13k and variola 190k. However, the report notes that it is "very likely" that newer devices will be able to produce 5,000–7,000 base pairs in length within the next 2–5 years – and 10,000 base pairs over 5-10 years. This would certainly be worrying and we should revisit this bottleneck before then.
[49] This is akin to how the 'Common Rule' regulates human subject research via federal funding (HHS, 2023)

which threat actors could continue to use.[50] It is also unclear if these measures will stay under a new administration (NextGov, 2024) – or have any knock-on effect elsewhere in the world.

It is unclear how effective strengthened screening would be. There is also the question of how much risk might continue with compliant providers due to 'false negatives' – whereby threat actors manipulate their orders so they incorrectly get screened as safe. Kane and Parker (2024) report many providers do red-teaming to test the effectiveness of their screening measures and if there are loopholes. Public information is limited, in large part due to wanting to avoid discussion that would help threat actors circumvent measures (Beal et al., 2023). However, I am able to note the following:

- Edison et al. (2024) note dangerous sequences can be 'camouflaged'. Doing so, they found that 36 out of 38 synthesis companies shipped orders for 500 base pair fragments of the 1918 influenza. This included "12 out of 13 IGSC companies" despite screening. The researchers claim they could reconstruct the 1918 virus with the fragments of these multiple vendors.

- IGSC (2024) criticised the above study, stating that many of its companies flagged the order and only shipped it because it was ordered by SecureBio, a trusted organisation. Thus, a non-affiliated threat actor would not succeed. However, the article admits there is "no screening solution" for a threat actor ordering many small pieces from multiple providers.

- Beal et al. (2023) note BLAST is the most popular screening method but can incorrectly categorise things due to taxonomic errors or database ambiguities. They find BLAST has a false positive rate of 5%–20% when common genetic tools are 'extended'; other methods 0.5%. They note "a similar dynamic applies to the more dangerous issue of false negatives".

- Wheeler et al. (2024) compare how four screening tools categorise 200–10,000 base fragments of 3 select agents. They found the proportion of 'optional flags' or 'undetermined' was 0% for Orbivirus, 44% for Francisella tularensis, and 83% for Coccidioide. Such cases would need several hours of manual investigation (though they aren't pandemic pathogens).

[50] The overall effect here is highly non-linear. Even if federal funding only makes up 61% of total R&D, plausibly a much higher share of actors receive at least *some* kind of federal funding.It currently appears unclear how the EO deals with this: I.e. if a company participated in just one NIH grant, does it have to comply? Even if the EO ends up applying to all partial cases, Gryphon Scientific seems to suggest ~5% of laboratories are "invisible" to federal oversight (Greene et al., 2022). This may be enough to keep a few non-compliant DNA providers in business. Thus, stylistically, instead of choosing between ~10 non-compliant providers, threat actors might now choose between ~2. But that doesn't reduce risk much, as they still have a viable option.

Other Challenges

Compared to the above, obtaining other equipment appears much less difficult. Indeed, micro-pipetters can be ordered on Amazon, tissue culture hoods on Alibaba, and there are also large second-hand markets for such equipment.

The increasing existence of biotech start-ups and community laboratories also shows the increasing proliferation of getting this. A lot of US biosecurity oversight and enforcement only applies to people who receive federal funding, a lot is out of regular purview. Gryphon estimated that >6% of US labs today are "invisible" in this way. One extreme example of this is Reedley Biolab, which operated illegally until it was shut down in 2023 and had worked with HIV, COVID and other agents it received from a US supplier [page 50]. Allegedly, they also worked with Ebola (a Category 1 Select Agent), although unclear how it got this.

The two main reasons for scepticism I found are:

- Maintaining sterile working conditions is notably harder to do in 'garage' settings than a professional laboratory. Aum Shinrikyo created 'fermentors' to produce C. botulinum for a terrorist attack but failed, plausibly to not being able to maintain sterile conditions (Danzig et al., 2012). It is unclear how difficult an obstacle this creates for reverse genetics.

- Sometimes there is regulation that requires equipment, like tissue culture hoods, to be certified upon installation. A supplier might insist on this as part of shipping their order, drawing attention. However, from speaking to experts, this rarely seemed to be an issue.

2.6 | Radicalization Rate, R

Defining A 'Meaningful' Level Of Intent To Cause Harm

Quantifying these above arguments is hard. In order to move towards a numerical estimate I first need to define what I mean when trying to measure intent. There is an important difference between people who report on a survey they think a pandemic would be good, who might actually press a button to make this happen, and would of their own volition take steps in the real world to build a pathogen.

Since reverse genetics protocol takes at least a month to run, I will consider meaningful intent as 'at least 1-month of serious sustained effort'. The actual time needed is likely even longer than that, so later I will consider how this number might decay for 3-months, 6-months etc.

Estimating The Intention For >1-Month Of Effort

Incidence-Based Reference Classes

First, I try to bound the intent rate by looking at related forms of violence:

Incidence-Based Reference Classes	#/People/Yr
CRBN-Based Indicators	
CBRN Database (2024) notes 48 instances of people pursuing agents 1993–2024, of which 13 are also not labelled ‘toxins’. When requesting additional data, the education level is unknown for ~half, but for those we do know at least half appear to have at least an undergraduate education. If there are ~21M non-experts [see N] throughout this 30-year period, we get a rate of 1-in-100M [= (13*0.5)/(21M*30yr)]. On the one hand, we might think that the subset of people who would consider deadlier pandemic agents if they could is much smaller. On the other hand, there are likely many more people who had the *intention* to commit a biological attack but not picked up in this database (only 3 events count as “proto-plots”) which seems too low). I think the latter consideration mostly offsets this, and perhaps slightly dominates.	~1-in-100M
I thank Rose Hadsahr for constructing a database of organisations that pursued CBRN agents. She examines the case studies and makes a qualitative assessment on whether they would have pursued pandemic agents if synthetic biology had made it feasible and their education levels. She concludes that she expects ~2.5 cases between 1970-2024 (1 strong case; 6 maybes, of which 2/3 had a STEM Bachelor).[51] I.e. 1-in-450M. Compared to the method above, this method is much more plausibly an overestimate – since it attempts to account pandemic intention but does not capture ‘proto-plots’ that do not function in the database or otherwise. Thus, I take it as an upper bound.	More common than 1-in-500M
In Williams et al. (2025) we asked 46 subject-matter experts and 22 superforecasters to estimate “What proportion of each of the following groups will, in 2026, spend at least 1-month trying to deliberately create an epidemic bioweapon, i.e. with a meaningful intention to kill more than 10,000 people via a human transmissible pathogen”. This multiplication is hacky and I don’t fully trust the results. Some implied numbers appear absurd (e.g. two people say 1-in-1000). So I take this more as a maximal range.	Experts: 1 in ~100K [30K to 30M] Non-Experts: 1 in 10M [100K to 100M]
Violence-Based Indicators	
Wikipedia (2024) notes that there have been 32 plots to assassinate a US president post-WWII. If we assume a ~175M US adult population [growing from 100M to 250M] (Census, 2021) that gives us 1-in-400M. There might also be additional plots that involved 1-months of effort but that are not recorded here. I imagine that bioterrorism is still rarer but assassinating an incumbent president is a very specific motif,	~1-in-400M

[51] The ‘strong interest’ was Aum Shinrikyo. The ‘maybes’: Unnamed Christian Millenarian groups; Islamic State; World Islamic Front for Fighting Jews and Christians; Aryan Republican Army; RISE; and al-Qaeda

so I'm actually not sure it is obviously >10X rarer. Thus, on the whole I take this number as reasonably close to a best guess.	
One crude heuristic is to look at analogous examples of people committing crimes with some kind of omnicidal or pandemic intent, and reasoning that some fraction would plausibly pursue epidemic terrorism if it were feasible. Torres (2018) provides 15 case studies. On the one hand, I find this list plausibly misses several cases; on the other hand I actually only take about 3 of the case studies as close proxies.[52] Overall, I think it seems reasonable to assume that across a generation (20yrs), 0.5 – 3 people with STEM bachelor degrees would pursue epidemic terrorism. I.e. a rate of 1-in-140M to 1-in-800M [=0.5/(21M*20yr)]	More common than 1-in-800M
Wikipedia notes six murder-suicide attempts by commercial flight pilots between 1970 and 2024, each of which causes numerous casualties – plus a further suspected cases (see also Kenedi et al., 2016 for more details). CAE (2020) estimates that there are ~333k commercial airline pilots. This gives a base rate of 1-in-3M per pilot year. However, I imagine that epidemic terrorism is much rarer than pilot-suicide. I also note that such pilot-suicides presumably take much less than 1-month of effort to do. Thus, I take this as an indicative lower bound – with epidemic terrorism being >10X rarer.	Less common than 1-in-3M
Duwe (2020) noted that 1976–2018 the US had 845 mass shootings (of which 158 were public). Assuming there are ~250M adults in the US, this translates to 1-in-12M (and 1-in-66M) I am less sure how to adjust this for STEM Bachelors, if this is equivalent to 1-month of effort, but on the whole these don't seem like critical considerations. I do imagine that epidemic terrorism is a rare incentive than mass shooting. Duwe et al. (2023) also notes the US rate is 4-6X higher than the rest of the world. Thus, I take this as an indicative lower bound – with epidemic terrorism being >10X rarer.	Less common than 1-in-12M/66M
Aamodt (2016) looks at the average number of serial killers committing 2 separate events of murder. They found a peak in the 1980s of 768 and a recent decline in the early 2010s to 117. Assuming there are ~250M adults (Census, 2021) that translates to rates of 1-in-3M and 1-in-20M respectively.. On the one hand, I am somewhat unsure about the STEM and 1-month adjustments here. These may move the rate down, but I don't think critically (the average IQ was 94.5, slightly lower than the US average of 98). On the other hand, the data only seems to include those caught and identified – which may move the rate up.	Less common than 1-in-3M/20M

[52] 1978: Ted Kaczynski, a professor, conducted a series of mail bombings to bring down industrial society.
1995: Aum's Shynrikio' Seiichi Endo tried to produce botulinum and anthrax to trigger an apocalyptic event.
1999: Eric Harris built several explosives and did a school shooting, writing about his hate for humans.
I'd add 2001: Bruce Ivins, a microbiologist, suspected to have committed the 2001 Anthrax Attacks.

Belief-Based Reference Classes

> *I thank Damon Binder for much of the framework and data; and Rose Hadshar for some of the data*

A more micro approach to the one above is to estimate [i] how many people plausibly endorse extremist ideologies that would support cause somebody to pursue epidemic terrorism, and then [ii] estimate how many people are likely to take action based on how many people 'follow through' for other kinds of extreme beliefs.

It should be noted that I think that this approach –even if done well– only covers a fraction of the 'true' risk. It excludes accidents that might occur without malevolent intent, or people who act non-ideologically.[53] I think that all else equal the numbers from such an approach are likely under rather than overestimating the true intention rate. Still, I find this a useful sanity check to do.

Number Of People With Beliefs, Whose Extremist Version Could Justify Epidemic Terrorism

Torres (2018) notes the following groups as especially relevant:

- **Apocalyptic:** A subset of these Apocalyptic Groups might actively try to "destroy the world to save it" – such as Aum Shinrikyo or Heavens Gate
  - "I want the present world, which is so full of pain, to be extinguished" [p10] and that they "would be the sole survivors" [p8]. See also Lifton (2000)
- **Misanthropic:** A subset of these might hate humanity or more systematically believe destroying life would lower suffering in the world
  - "The human animal is the only evil animal in the animal kingdom. We destroy everything. . . . I email the president weekly and beg him to push the button" [p38]
- **Radical Environmentalism:** A subset of these might support the idea of industrial collapse or lowering global population levels – such as Ted Kaczynski
  - "We would also welcome [...] any new anti-Human viruses" [p21] "Any means to decreasing human population would be welcomed [...] even [...] disease. " [p22]

[53] Many threat actors might not actually have well thought out ideological beliefs for why they are doing this. I.e. Many lone wolves might want to construct an epidemic and then rationalise an ideology backwards from that. For example, think of Thomas Matthew Crooks' attempted assasination of Donald Trump without a clear political motivation, but instead "somebody intent on perpetrating mass violence, and they happened to pick a political rally" (New York Times, 2024).

- **Other:** There are also other motivations or beliefs systems people might have. For example, Bruce Ivins is suspected to have committed the 2001 Anthrax attacks to get political sympathy for continuing his anthrax vaccine research programme.

Quantifying these group sizes into numbers is hard – both because they are niche so there is limited reporting *and* because any data likely contains more 'unserious' responses or 'trolling' behaviour. One highly imperfect approach is to compare the relative size of relevant online communities to more mainstream ones for which we have data. This approach is limited because the numbers might be non-constant, i.e. there is a higher proportion of 'troll' deep ecologists than 'troll' environmentalists.

Nonetheless, I find this exercise somewhat useful to do. I conclude ~450K [100K-1M] people in the US would potentially endorse beliefs that might justify epidemics – or 0.2% of the adult population [0.05%-0.5%]. Note, this does not yet discount what fraction endorse a violent version of these beliefs, or would take any violent action to act on these beliefs. This number should be thought of as, who on a poll might say they support a pandemic punishing humans.

Variable	Mainstream Communities		Groups With Potential Epidemic Beliefs		
Group	**Environment**	**Far Right**	**Misanthropy**	**Deep Ecology**	**Apocalyptic**
Reddit Size	~1.6M [r]	0.6M-1.1M [rw]	90K-200K [r;r]	5K-25K [r;r]	<3K-16K [r]
Relative Score:	100%	70%	5%-12%	0.3%-1.5%	<0.1%-1%
Internet Search	10	25	24	1	1/715
Relative Score:	100%	250%	240%	0.02% – 1%	0.01%
US Size	~20-40M each [Pew and Gallup]		1M-4M	4K-800K	<2K-40K

Fraction Of People Who Hold Extremist Version *And* Take Actions On Such Beliefs

Now I do want to discount what fraction of believers people are likely to take radical action. To do so, I first look at three reference groups that most likely *don't* want to cause an epidemic, but for which there is more ample data: Islamic Jihadism, Far-Right Extremism, and Eco Civil Disobedience. Looking at the below table, I feel a rate of "1-in 6K-600K per person per year rate" seems like the roughly correct range for how many people act violently on their beliefs.

Variable	US Jihadi Terrorism	US Far Right Murder	UK Eco Civil Disobedience
# Of People In Country	210,000,000	210,000,000	45,500,000
	OECD (2024) est. of the US population aged 15-64 years	OECD (2024) est. of the US population aged 15-64 years	ONS (2024) est. of the UK population aged 15-64 years
% In Demographic Group	1.1%	10%	20%
	Pew (2018) est. Muslims make up 1.1% of the US population	Pew (2021) est. of most right-wing typology	YouGov (2024a) 18% UK say eco is #1 issue. Also Gallup (2021)
% Supporting Extremist View	8%	2%	5%
	Both START (2016) and Pew (2011) ask if suicide bombing in defence of Islam is justified	Westwood et al. (2022) est. for political murder as justified [adjusting for disengaged]	YouGov (2024b) est. 5% of US support defacing property and riot as okay for their #1 issue
# Extremists	184,800	451,500	455,000
# Extremist Crimes Per Year	29	20	549
	Williams et al. (2018) finds at least 471 US people joined Jihadi FTOs 2001-20217	CSIS (2021) finds ~550 US far-right terrorism attacks and plots between 1994-2021	London Assembly (2022) reported 4,481 charges from XR 2019-2022 (LDN only
1-in-X per person per yr	**6,278**	**22,165**	**828**
# Terrorist Killings Per Year	1.6	0.7	0.1
	ICCT (2023) finds 24 Jihadi terrorist attacks in the US 2004-2019 (not incl. plots)	ADL (2023) found 47 extremist killings 1971-2022, of which ~3/4 far-right (not incl. plots)	GTD (2024) finds 1 eco-terrorist in UK 2009-2019 (not incl. unknowns, plots etc.)
1-in-X per person per yr	**115,500**	**653,234**	**4,550,000**

Putting This Together

Having established these reference classes, we can then construct this simple back of the envelope calculation. See the table below for the assumptions:

Variable	US Epidemic Terrorism	
	Low-End Assumptions	High-End Assumptions
# Of People In Country	210,000,000	
	OECD (2024) est. of the US population aged 15-64 years	
% In Demographic	0.05%	0.5%
% Supporting Extremist View	2%	8%
Additional Multiplier	1.0x	4.0x
	I think the extremism rate for extreme beliefs is presumably higher	
1-in-X per person per yr	600,000	6,000
% Attempts That Are Pandemic	10%	50%
	I think this is highly uncertain so I take a wide credible interval	

In turn, this gives us the following results below. Intuitively, out of the 210M US adult population we expect to see ~0.1 people want to cause an epidemic in a given year [0.006 – 1.6]. Ergo, we need 10 years to expect to see one attempt. **Or, in other words, 2.1B people per year.**

In reality, we think the actual population that could conceivably pull this off is almost 10X smaller – only 20M non-expert STEM Bachelors. So in fact we expect ~0.01 threat actors to want to cause an epidemic in a given year; so we expect to need 100 years to reach this threshold.

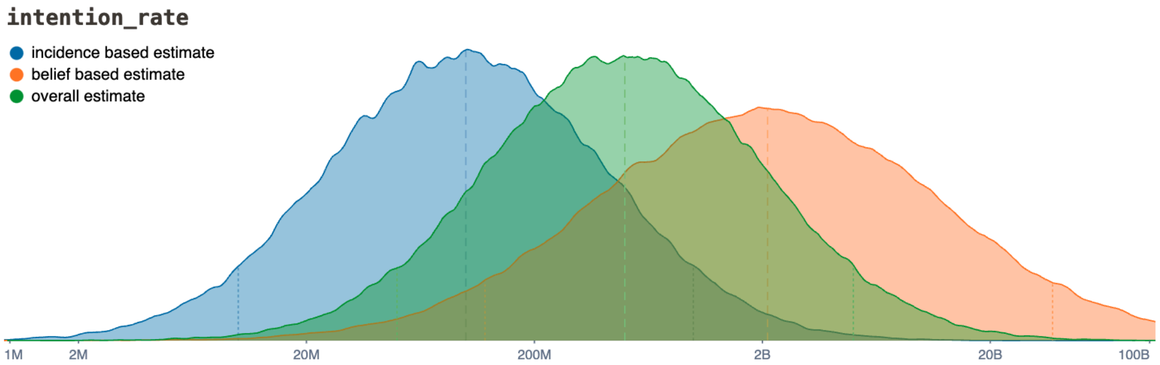
Variable	Belief-Based US Epidemic Terrorism			
	Mean	5th Percentile	**Median**	95th Percentile
# People Per Year	210M			
# Believers Per Year	430K	110K	330K	1.1M
# Extremists Per Year	41K	5900	27K	120K
# Bioterrorism Attempts Per Year	0.46	0.0056	0.1	1.7
E(Years For 1 Attempt By A US Adult)	47	0.58	10	180
E(People For 1 Attempt By A US Adult)	9.8B	120M	2.1B	38B
# Threat Actors	20M			
E(Years For 1 Attempt By A Threat Actor)	490	6.1	110	1900

[Result] Average Rate Individuals Have Meaningful Intent

Overall we have two methods. The incidence_based rate –which likely somewhat overestimates the risk– and the belief_based rate – which likely somewhat under-estimates it. Thus, to arrive at an overall best guess, I take an average of these [green], as shown in the graph below.

Overall, I conclude that for a given person there is a **1-in-1B [50M – 5B]** chance that they would seriously pursue epidemic terrorism in a given year with at least a month of effort. Note that if we assume up to 20M non-experts who could be lonewolves, that gives us one attempt per 25 years.

Clearly a huge amount of uncertainty persists, with the 5th and 95th percentiles differing by three orders of magnitude [100X]. I think this does just reflect our general ignorance on this topic. I don't expect future work to massively cut this down, although I am working with FRI to collect more superforecaster responses. I think an upshot is that it does make sense.



Name	Mean	Stdev	5%	25%	50%	75%	95%	Samples
incidence based estimate	263M	642M	10M	38.7M	100M	257M	996M	100000
belief based estimate	9.79B	43.8B	121M	648M	2.11B	6.87B	37.5B	100000
overall estimate	1.33B	3.18B	50M	194M	498M	1.28B	5.01B	100000

2.7 | Epidemic "Take Off", E

In order to engineer a pathogen you need to know the specific DNA sequence it consists of. Increasingly, such genome sequences are publicly available online (e.g. GenBank). There may be some issues with such data's accuracy,[54] but most experts I spoke to do not consider this a bottleneck.

Instead, the main 'bottleneck' is whether one of these known genome sequences actually has pandemic potential – meaning that it could spread within a human population (i.e. R_0>1). This assumption is non-obvious, as most viruses do not fit this criterion.

- We can perhaps draw some inference from the literature around lab-originated outbreaks, although it is important to note that much of this remains disputed:
- Both 1977 pandemic flu is believed to have not been a natural occurrence, but either a failed vaccine trial or a lab accident (Rozo & Gronvall, 2015). Similarly UK's 2007 FMD outbreak in cattle (Enserink, 2007). We might see these as "proofs of concept"

[54] NASEM (2018) [p41] notes when a scientist creates novel RNA viruses, it continues to mutate and replicate. Thus, scientists end up with a 'mixture' of slightly different copies. When the scientists enter the sequence of a single member of that mixture into a [public] database "it is possible that the starting sequence may not generate a "wild type," fully virulent population after booting". Rasmussen (2022) notes that "most "complete" viral genome sequences actually have pretty poor coverage at the ends and in highly structured regions." Somebody trying to create a pathogen won't be able to tell if they are making a virulent population until the very end.

2.8 | Potential Deaths, D

For people interested in converting "expected deaths" into a "willingness to pay to prevent" number, I suggest using the following calculation:

Variable	Description	Value
Expected fatalities	Epidemics follow a Pareto distribution, declining more rapidly after ~1M deaths (Marani et al., 2021). I take a range mean between Ebola and ones slightly smaller than COVID.	~2.5M [0.1M-10M]
Valuing fatalities	This is a normative assumption. I follow US Government guidance of $5M-$9M and adjust from 2020 to 2025 dollars. Other values appear defensible.	$8.75M per death
Economic Damages	Pandemics also create economic damage. Reviewing Glennester et al. (2023) it appears that an average pandemic death corresponds to a further ~$0.57M in damages	+$0.57M per death
Equity Adjustment	I exclude this for my result but note that other normative approaches may want to further adjust if they believe damages accrue regressively in ways that skew econ statistics.	n/a (1X)
Total Damages	*[Calculation: Fatalities * VSL + Economic * Equity]*	$23T WTP

Value Of Statistical Lives [$3M per death]

If we want to be able to combine these fatalities into a single number with the economic damages below, it is typical to convert these fatalities into willingness-to-pay-to-avoid. In cost-benefit analysis, this is typically done using a so-called "value of statistical life" [VSL] (Colmer, 2020; Bressler, 2022).

Choosing a VSL requires several normative assumptions, none of which are obvious, such as whether to value all lives equally or adjusting by income. For simplicity, I settle on valuing deaths at ~$8.75M in line with the US Government, although lower values appear defensible for a global context. If there are 2.5M expected deaths, that gives us $9T in willingness-to-pay-to-avoid. (Given such assumptions are non-obvious, I will also continue to report the fatalities separately in the bottom-line results.)

Method/Source	Description	Result
US Government	OMB noted "most federal agencies are using VSLs between $5 million and $9 million and that values outside of this range would be difficult to justify" (FEMA, 2022). But note that rich countries may be willing and able to "pay more" than the globe.	~$7M (2020), converting to $8.75M with inflation
Banzhaf (2021)	Does a meta-analysis and finds "baseline model yields a central VSL of $7.0m, with a 90% confidence interval of $2.4m to $11.2m". But again may skew high-income	~$7M
Favaloro & Berger, (2020) [Preferred]	Values all statistical life *years* at $100K based on both OECD and LMIC data. So if each death is associated with ~30 years of life lost,[55] that gives us $3M [$100K*30]	$3M
Bressler (2021)	Headline estimate values all statistical life *years* at 4X avg. global consumption. So if each death is associated with ~30 years of life lost, that gives us $1.5M [4*12K*30]	$1.5M
Sweis (2022) [used by Glennester, 2023]	Scales the value of life by income, which deals with the USG critique above. However, it's not clear if international actors should thus value rich lives more (Sunstein, 2004).	$1.3M

[55] Glennester et al. note the historical average is 29.5 YLL, with a range of 15-50.

Economic-output loss [$0.6M of additional economic damages per death]

Average Pandemic Severity [Additional $0.6M per death – or 1.3% GWP per attack]

In addition to fatalities, pandemics can also create harm by damaging the economy – such as hurting tourism, or, in extreme cases, requiring lockdowns. Although the literature here is less robust, there are still several estimates to draw on looking at how GDP deviates from trends in years where pandemics occur. I note the following sources that seemed the most useful:

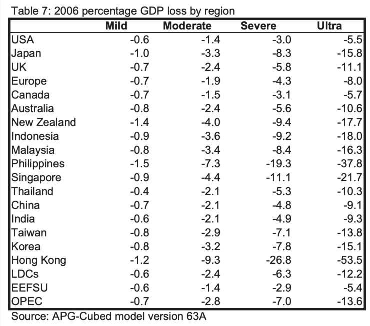
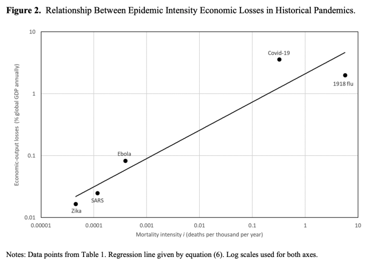
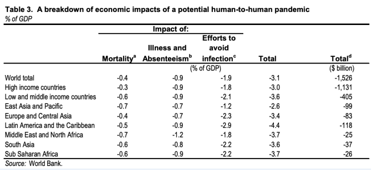
M&S (2006) [used in NASEM, 2016]	Glennester et al. (2023)	World Bank [used in Fan et al. (2016)
Estimated that an ex post of a "mild" event [1.4M deaths] costs 0.8% of GDP in the first year, "moderate" [14M deaths] 2.5%, and "ultra" [142M deaths] 10.5%.	Find a log-log relationship between mortality intensity and economic intensity. Suggests that if the average is ~1M-2M deaths that's ~1% of GDP in the first year	Estimates a total impact of 3% of world GDP, of which 2%-point is due to "efforts to avoid infection". It assumes a pandemic similar to 1918, i.e. 1% of the world dies.
		
~2.5% GWP per year for ~2yrs due to a 14M deaths event	~1.5% of GWP loss in avg. year due to a 1-3yrs for ~3M death event	~3% GWP per event due to a ~80M death event

Table 7: 2006 percentage GDP loss by region

	Mild	Moderate	Severe	Ultra
USA	-0.6	-1.4	-3.0	-5.5
Japan	-1.0	-3.3	-8.3	-15.8
UK	-0.7	-2.4	-5.8	-11.1
Europe	-0.7	-1.9	-4.3	-8.0
Canada	-0.7	-1.5	-3.1	-5.7
Australia	-0.8	-2.4	-5.6	-10.6
New Zealand	-1.4	-4.0	-9.4	-17.7
Indonesia	-0.9	-3.6	-9.2	-18.0
Malaysia	-0.8	-3.4	-8.4	-16.3
Philippines	-1.5	-7.3	-19.3	-37.8
Singapore	-0.9	-4.4	-11.1	-21.7
Thailand	-0.4	-2.1	-5.3	-10.3
China	-0.7	-2.1	-4.8	-9.1
India	-0.6	-2.1	-4.9	-9.3
Taiwan	-0.8	-2.9	-7.1	-13.8
Korea	-0.8	-3.2	-7.8	-15.1
Hong Kong	-1.2	-9.3	-26.8	-53.5
LDCs	-0.6	-2.4	-6.3	-12.2
EEFSU	-0.6	-1.4	-2.9	-5.4
OPEC	-0.7	-2.8	-7.0	-13.6

Source: APG-Cubed model version 63A

Figure 2. Relationship Between Epidemic Intensity Economic Losses in Historical Pandemics.

10
1
0.1
0.01
Economic-output losses (% global GDP annually)
Covid-19
1918 flu
Ebola
SARS
Zika
0.00001
0.0001
0.001
0.01
0.1
1
10
Mortality intensity *i* (deaths per thousand per year)

Notes: Data points from Table 1. Regression line given by equation (6). Log scales used for both axes.

Table 3. A breakdown of economic impacts of a potential human-to-human pandemic
% of GDP

	Impact of:				
			Efforts to		
		Illness and	avoid		
	Mortality[a]	Absenteeism[b]	infection[c]	Total	Total[d]
			(% of GDP)		($ billion)
World total	-0.4	-0.9	-1.9	-3.1	-1,526
High income countries	-0.3	-0.9	-1.8	-3.0	-1,131
Low and middle income countries	-0.6	-0.9	-2.1	-3.6	-405
East Asia and Pacific	-0.7	-0.7	-1.2	-2.6	-99
Europe and Central Asia	-0.4	-0.7	-2.3	-3.4	-83
Latin America and the Caribbean	-0.5	-0.9	-2.9	-4.4	-118
Middle East and North Africa	-0.7	-1.2	-1.8	-3.7	-25
South Asia	-0.6	-0.8	-2.2	-3.6	-37
Sub Saharan Africa	-0.6	-0.9	-2.2	-3.7	-26

Source: World Bank.

Importantly, we need to scale the size of economic damages relative to the size of the mortality in question. Per Glennester et al. (2023) there appears good reason to think that there is a log-log relationship here. We can use its equation to calculate a "death per year" and then scale this up to our average sized pandemic of 2.5M deaths. This is done in the table below.

Converting "per death" economic harm in Glennester et al. (2023)

Variable	Unit	Source
Annual % GDP lost per year per "i" unit of mortality	2.10	Equation 6
Year duration of a given pandemic	2	Assumption
Dollar value per % GDP lost	$1T	World Bank
Deaths per "i" unit of mortality	8.14M	World Bank
Additional dollars lost per pandemic death	**$0.57M**	Calculation
Applied to scenario:		
Expected deaths for an 'average' pandemic	2.50M	Previous
Expected total dollars lost	$1.43T	Calculation
Expected % GDP lost	**1.3%**	Calculation

Equity Adjustment [n/a]

Whilst many estimates just cite the "raw" dollar damages, I note that this is not necessarily the same as "willingness to pay". For example, $100 of damage to a New York banker is likely much less "bad" than $100 of damage to a rural Kenyan farmer. Economic damages against poorer communities can get undervalued using such "raw" dollar damages. This is an issue since pandemics often hit the economies' of lower income countries by even more.

In cost-benefit analysis, we can address this issue using an "income weighting" (Prest et al., 2024). I exclude this from my analysis (especially since I chose a higher United States based VSL) but note that others may want to – especially if using a global VSL like $3M in via Favaloro and Berger, (2020).

How bad is it if we assume pandemics are proportional, i.e. everyone lost 2% of their income?

I adapt the approach in in making the following assumptions:

- Let's assume that "utility" is approximately log-income. I.e. a person earning $100K losing $2K is roughly as bad as someone earning $1K losing $20.

- Let's assume the normalization point in which we measure "income-weighted" dollars is someone earning ~$14K which is the same as global GDP per capita [IMF] or salary [WIP]
- A decision-maker valuing all 10B people losing 3% of their income is worth $3.5T [8.5B people * $14K set-point * 3% income loss]. This is the same as raw dollars.[56] But…

How bad is it if we note pandemics are regressive, i.e. the average person actually loses more?

We can look at how much we might be under-estimating economic damages due to inequity. I aim to do a quick sanity-check here rather than a big deep-dive. The more "intense" economic damages are in lower-income countries, the higher the adjustment. I try to illustrate this dynamic in the table below.

Background Statistics [Source: World Bank, 2023]				Examples Of Damage Severity	
Country Group	**GDP/capita**	**% GDP**	**% Pop**	**"Proportional"**	**"Regressive"**
Low	$920	0.6%	9.0%	2%	10%
Lower-Middle	$2,196	7.0%	43.1%	2%	10%
Upper-Middle	$11,754	28.0%	32.0%	2%	0%
High	$54,498	64.4%	15.9%	2%	0%
[1] If you just add "raw" dollars (GDP-weighted average)				2.0%	0.8%
[2] If treat all % losses equally (population-weighted average)				2.0%	5.2%
Equity adjustment: [2]/[1]				1.0x	6.8x

To see how much this dynamic might matter, we now need to look at how "regressive" pandemic economic damages might be. I couldn't easily find data on global households, but there are two rough proxies we can use: [i] how much pandemics affected GDP and/or unemployment across countries; and [ii] how pandemics affected mortality, using this as a proxy. Both of these methods are imperfect:

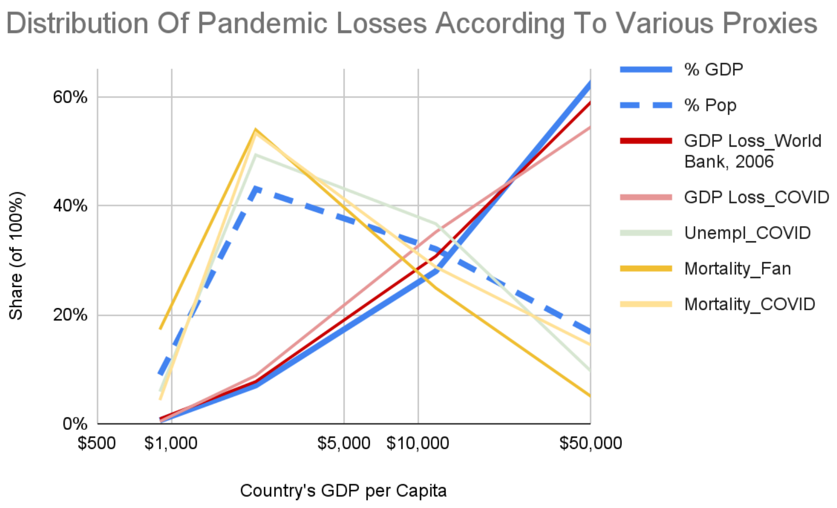
[56] That's not a coincide, as we roughly just did "loss" = "loss / global $" * "global people" * "global $ / people". The small increase is because the global population is set to increase.

Proxy Method	Issues With Proxy
Global Household Data	Couldn't easily find this data
Country-Level GDP Loss and/or Unemployment Loss	**Under-estimate**: The country-level GDP statistics might leave out inequality within countries, which increases the skew by more (on the other hand… the country-level GDP loss might be heavily affected by large firm losses that don't "trickle down" to ordinary people). Unemployment seems a decent proxy.
Country-Level Mortality	**Under-estimate**: Richer countries are "older" and so might have a higher mortality:economic ratio than poor countries (on the other hand… richer countries might have more economic value that could be destroyed though lockdowns etc.)

For all methods I try to find an academic study and look at COVID-19. The results are as follows. We can see that all sources suggest pandemic losses are at least slightly regressive – i.e. rich countries have a lower "severity" than poor countries. But it only causes a meaningful adjustment in two cases.

Further work could help narrow this down, but for now I take a crude 1.25X [1.1X – 1.5X]. If we previously valued economic damages at $3.5T, this now suggests we should raise it to ~$4.5T.

Variable	GDP Loss		Unemploy.	Mortality Loss [SMU]	
Country Group	World Bank, 2006	COVID [ILO]	COVID [ILO]	Fan et al. (2016)	COVID [WHO]
Low	-2.80%	-3.6%	-3.1%	4.95	0.91
Lower-Middle	-2.10%	-6.70%	-5.40%	3.22	2.34
Upper-Middle	-2.10%	-6.70%	-5.40%	2.00	1.70
High	-1.80%	-4.60%	-2.40%	0.63	1.62
GDP-weight	-1.9%	-5.3%	-3.5%	1.22	1.69
Pop.-weight	-2.1%	-6.1%	-4.7%	2.57	1.89
Equity Adj.	1.1x	1.1x	1.4x	2.1x	1.1x


Distribution Of Pandemic Losses According To Various Proxies
Share (of 100%)
60%
40%
20%
0%
$500
$1,000
$5,000
$10,000
$50,000
Country's GDP per Capita
% GDP
% Pop
GDP Loss_World Bank, 2006
GDP Loss_COVID
Unempl_COVID
Mortality_Fan
Mortality_COVID

Appendix 3 | Scenario Estimates

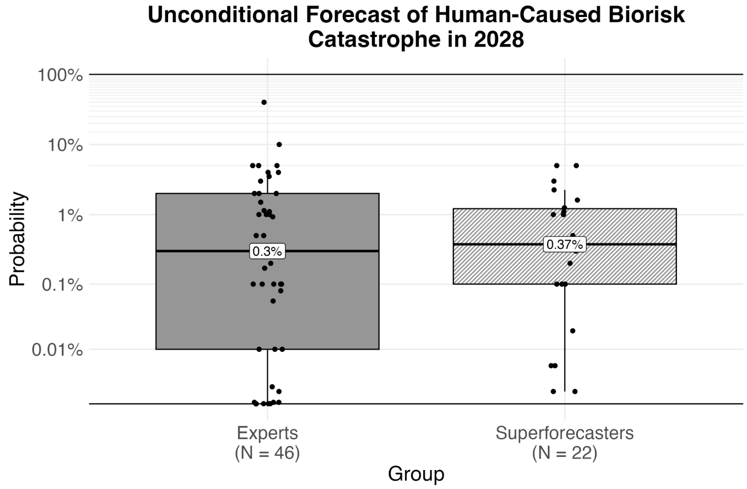
3.1 | Laboratory study

The evaluation in question is more specifically defined as follows:

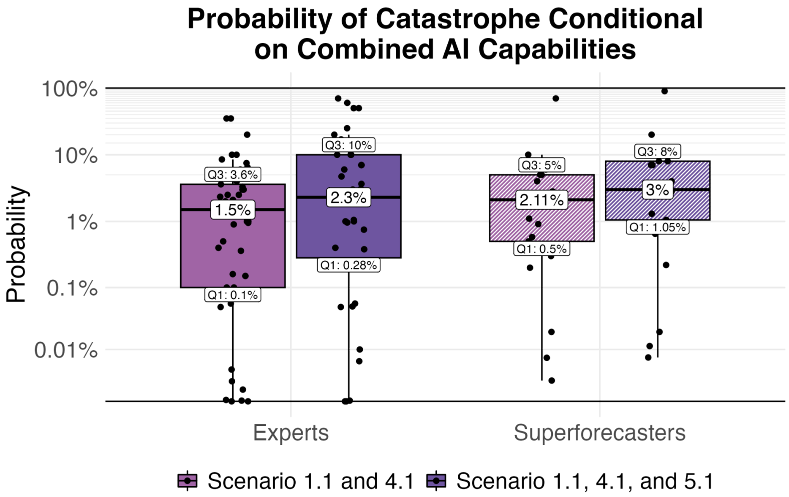
- A randomised controlled trial is conducted with 100 participants, who hold undergraduate degrees in STEM. The study is conducted in the first quarter of 2026.
- The control group has access only to AI systems roughly in line with the best-performing models of 2023, plus unrestricted access to Internet content apart from frontier AI (search results, videos, protocols, etc.). They will not be provided access to any experts. They may reach out to experts independently, but are not allowed to say that they are attempting to synthesise the virus for a study.
- The treatment group has unrestricted access to the best-performing multimodal AI systems available at that time, including versions of the models that respond directly to risky questions without refusals (like the research-only version of GPT-4 that was used in OpenAI, 2024), as well as unrestricted access to Internet content. The models can also access the Internet and external tools. The treatment group will not be provided access to any human experts; as with the control group they may reach out to them independently but are not allowed to say that they are attempting to synthesise the virus for a study.
- The model being evaluated demonstrates several advantages over 2023 models but does not show extreme changes in AI self-improvement capabilities, e.g. it cannot replace a top ML engineer.
- Both groups aim to, starting from scratch with basic wet lab access (e.g. with access to the equipment listed by University Lab Partners, but needing to purchase their own consumables, including any synthetic DNA, and non-commonly shared equipment), obtain live, viable samples of some pathogen similar in complexity to replicating influenza, likely through reverse genetics.
- They are given three months of calendar time and $100,000 to spend. Both groups are monetarily incentivised to achieve success.
- Please assume that the study is conducted to a high standard, and that another group is able to replicate the findings.

3.2 | Forecaster Results

To sanity check these results I compared them to the survey respondents in Williams et al. (2025). We can see that my estimate of ~0.15% is close to their median results of ~0.3% (which I was not aware of). I speculate that part of why my number is lower can be explained because I only considered biological misuse, whilst there's also epidemics caused by lab accidents.



Similarly for Scenario A, I compared my own estimates with the structured survey in Williams et al. (2025). What I believe to be the closest comparable point is shown below. I can see that participants' total risk estimates are 0.3% → ~2%, which is largely similar to those of my model predictions. Scenario B did not have an equivalent question.



Appendix 4 | Forecaster Survey

4.1 | Instructions Given To Reviewers[57]

Whilst the report has endeavored to synthesise many differing opinions, arriving at any bottom-line estimate still requires significant judgment. Therefore, as an additional check, we are assembling a panel of expert reviewers to engage with this report and produce their own forecasts using the ALORED model.

We ask reviewers to complete their initial responses by March 21. We will then update the report to incorporate reviewers' feedback and present an overview of everyone's results. Participants will then have a chance to update their estimates, which will constitute the main results of the final publication.

The following gives you an overview of the rest of the document. **You are not yet asked to complete any of the exercises.** Section 1 begins on page 5:

- After you finish reading Section 1, you will be asked to complete **Exercise 1**, which means initial "flash" estimates of how many people in theory could engineer a virus and the likelihood of such an attack. This **survey should take ~10 minutes** to complete.
- As you make your way though Section 2 you will be asked to complete **Exercise 2**. At the end of each sub-section laying out the evidence for a variable, please come up with your own 'baseline' estimate (i.e. absent AI progress). This **should take 2-4 hours [i.e. main part]**.
- After having finished reading Section 3, you will be asked to complete **Exercise 3**, which asks you to similarly adjust your previous answers to reflect on the new "AI scenarios" – and to then come up with a new overall forecast. This **should take you ~30 minutes**.

[57] These instructions differed from those given to the double-blind reviewer that reviewed a near-final version of the report.

When answering, please keep in mind the following:

Caveats For Reviewers Around Filling In The Survey
There is limited evidence; this report asks for your best guess under uncertainty. We fully expect ourselves and others to change forecasts as more evidence emerges over time. We are asking people to make their best guess given what we know now (and will caveat any such answers appropriately). For example, there is disagreement about just how difficult it is to avoid being caught by law enforcement [O]. There is ongoing red-teaming work to see how effective DNA synthesis screening is (Esvelt, 2024; IGSC, 2024), and we can revisit estimates.
When estimating variables, please consider the "multiple-stage fallacy." One concern about the estimation method is that its multi-premise structure biases towards lower numbers. Participants may fail to adequately condition all of the previous premises to be true and properly account for their correlations, or they might hesitate to assign suitably extreme probabilities to individual premises. Please try to account for this in your estimates.[58] For example, suppose you assume a given STEM Bachelor only has a 1% chance of "succeeding at wet lab biology" [L]. The 1% that do succeed might be very skilled and determined. So when you go on to estimate "not being caught by law enforcement" [O], you will likely want to use a higher value having *conditioned* on [L].
When constructing confidence intervals, please keep in mind the "tails." Given this uncertainty, it is useful to estimate not just a point estimate but a confidence interval.[59] The survey asks you to specify your 5th and 95th percentile outcomes. The more uncertain we are, the larger the confidence intervals tend to be. By default, the model shows lognormal distributions, though you may in fact refer to any.[60]

Note that we need to be careful in choosing our tails:

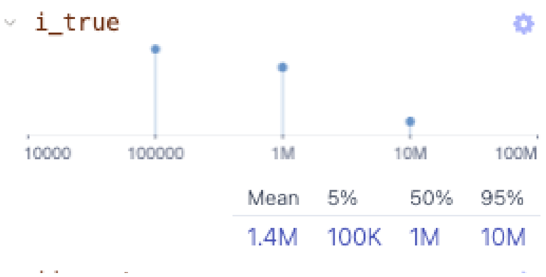
- It is plausible that most of the expected damage comes from the "tails" – and it's important to pay attention to those. I.e., the 90th percentile might only have a 10% chance of occurring but drive >10% of damages [1]
- The survey lets you use any distribution. Log-normal distributions are easy to specify and decent models for skewed data (Briggs et al., 2006), so these are the default. But it is important that we don't make the tail "too small" [2] or "too big" [3].
- To help, we also let you "trim" the tails (i.e. set max/min values) [4]. It can be useful to sanity check what they say about the "worst-case scenario" [5]

[58] This is why I ask reviewers to estimate inputs and outputs separately, allowing me to add "model uncertainty".

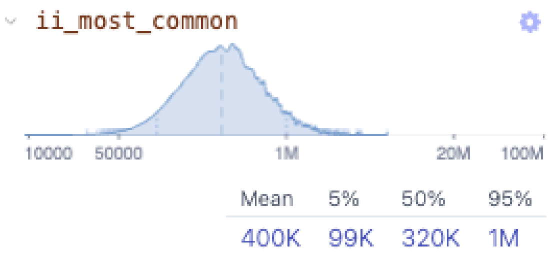
[59] We have seen that accounting for uncertainty has been important in other fields. For example, climate change (Weitzman, 2011), financial risk (Taleb, 2007), and the Fermi paradox (Sandberg et al., 2018).

[60] Note, I was later corrected that a more accurate term would be to refer to these ranges as credible intervals as opposed to confidence intervals, given the model is taking a Bayesian approach. For details see Statsig (2024).

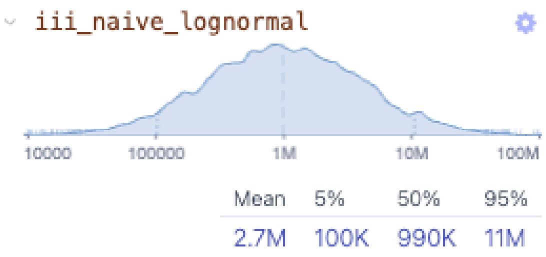
Approach	Description
[1] "True" Distribution	Stylistically, suppose the "true" distribution is a 50% chance it kills 10K people, 40% it kills 100K people, and 10% it kills 10M people. This gives **1.4M expected deaths** (2/3rds coming from the "tail")
[2] Too small Log-normal	If we only consider the most likely outcomes and, say, pick a log-normal distribution between 100K [10th percentile] and 1M [90th percentile], we miss the entire tail. Expected damages are **0.4M [3X too small].**
[3] Too big Log-normal	If given a log-normal distribution of 100K–10M, we assign *too* much weight to the tail (e.g. a 1% chance of 100M deaths, which matters a lot). Expected damages are **2.7M [2X too big].**
[4] Trimmed Log-normal	But if we "trim" this log-normal distribution to make it clear damages are log-normal between 100K–10M but *no bigger* than 10M, then we get **1.7M. [basically replicating the "True" Dist. above]**
[5] Worst Case Check	A good sanity check would be to explicitly consider the worst case we are worried about: a ~10% chance of killing ~10M. That is easier to think about and close to the truth: **1M [pretty close!]**



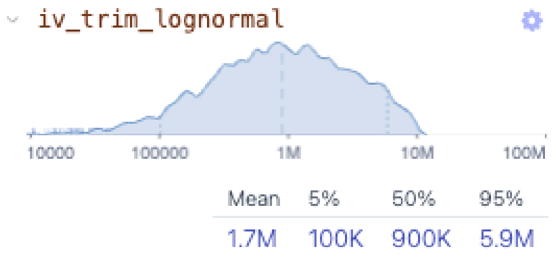
Mean	5%	50%	95%
1.4M	100K	1M	10M



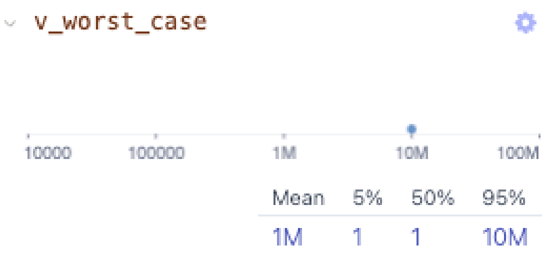
Mean	5%	50%	95%
400K	99K	320K	1M



Mean	5%	50%	95%
2.7M	100K	990K	11M



Mean	5%	50%	95%
1.7M	100K	900K	5.9M



Mean	5%	50%	95%
1M	1	1	10M

By default the model will display lognormal distributions, although you may in fact refer to any distribution you like. We will later also fit ‘beta distributions’ onto parameters that are probabilities – the issue is that these are computationally more expensive to include in the survey.

None

```
1. 1. Prob. a swine-flu-size pandemic in a year [100K death] est. 1 in 20 (5%)

2. Prob. a COVID-size pandemic in a year [30M death]: est. 1 in 100 (1%)

3. Prob. a random global person is a millionaire: est. 1 in 140 (0.7%)

4. Prob. a random US person is a Harvard student: est. 1 in 15,000 (0.007%)

5. Prob. a random US person is a neurosurgeon: est. 1 in 60,000 (0.002%)

6. Prob. a random US person is an elected Fed. politician: est. 1 in 600,000

7. Prob. of being struck by lightning in a year: est. 1 in 1,000,000

8. Prob. of an 2km asteroid hitting earth in a year: est. 1 in 1,000,000
```

4.2 | Overview of Qualitative Responses

Selected Qualitative Responses
Baseline: Number of Actors
Arguments For Lower Estimate: *“Biosecurity and law enforcement processes are evolving and nobody knows exactly how likely a big company (e.g. Twist) is to flag a pathogen sequence”* *“The person who handles the science may be terrible at operational security issues. You can't assume that the technical person will automatically overcome other issues.”* *“Non-STEM Bachelors [number too high. Est for number of years should be lower across all categories”*
Arguments For Higher Estimate: *“It makes no sense to exclude cohorts of people aged over 65. Outside the US, the proportion of graduates in STEM fields is significatively higher than 20-25%.”* *“Is it necessary for them to have financial resources themselves - couldn't they take out a loan?”*
Other: ***“***My estimates are highly dependent on assumptions about capability and cost of supplies/reagents/technology.”* *“I don't know how to "condition" the operational success on the laboratory success -- they are too intertwined.”*

Baseline: Likelihood Of Attack

Arguments For Lower Estimate:

Radicalization**:** *“Assumed those with technical degrees are less likely to radicalize (probably more financially successful and also less prone to religious extremism)” [3 others also made a point on this line – 2 people made the opposite case]*

“Extremely unlikely that a non-STEM person would attempt this in any meaningful capacity that we could detect or would need to worry about.”

“Most [historical cases of bioterrorism] were not trying to kill 10,000 people.”

“Covid experience will push people away from the use of bioweapons”

Epidemic Takeoff: *“Just because you synthesise it does not mean it becomes a weapon that can be intentionally used […] State-level BW programs went through arduous processes of stabilising pathogens”*

“You could have had someone engineer Ebola and release it, and whether it would actually cause greater than 10,000 deaths would depend on these factors, not whether someone had the technical skill to make [it]”

Arguments For Higher Estimate:

Radicalization**:** *“General personality types associated with high-level scientists that might make them more susceptible to radicalization than the average person.”*

“Radicalization may be higher at this moment than in the past, due to defunding of science in the US”

Epidemic Take Off**:** *“Different skill set, probably, to operationalise the pathogen distribution. That probably makes it more difficult for a lone wolf. Consider accidental release records here - most often spread is quite limited.”*

I don’t agree with the notion that non-STEM Bachelors have an intention rate that is 30 times lower than that of PhDs.”

Baseline: Ex Ante Damages

Arguments For Lower Estimate:

“Again, host-environment interaction with the pathogen (which also is connected to socio-cultural-economic factors).”

“I think it's fair to apply some downfiltering on the STEM bachelor and other background categories to account for their relative ability to purposely engineer higher virulence in some way”

Arguments For Higher Estimate:

*“A biologist is more likely to pick and properly culture (so that it retains is pathogenicity) a more pathogenic virus than others” I'd expect mol bio PhDs to be able to cause 2-3 OOMs more harm than STEM Bachelors” [*4]*

“Was COVID-19 really a once-in-a-thousand-years event? Historical precedents suggest it wasn't.”

“Potential for AI or other tech to assist in creating genome blueprints for novel pathogens that are both highly transmissible and highly lethal.”

Scenario A

Arguments For Raising By More:

“I believe AI might have an even stronger catalytic effect in the biothreat space than others do”

“The likelihood that someone attempts this seems like it would go up across the board, but the biggest jumps might be in the STEM bachelor group”

Arguments For Raising By Less:

“The cost/access-to-resources factor is still huge and I think puts some ceiling on the upper limit of increase. If the cost of DNA fragments, reagents, lab equipment, etc were also to fall, and ease of access to them to rise, that would also increase risk substantially”

“I've slightly upped the probability of operational success, though I remain unconvinced about how useful it is as a parameter to estimate.” [4 others raised this point]

“I considered raising the Radicalization numbers but ultimately did not. The range contains a great amount of uncertainty and raising those numbers here strikes me as too speculative.” [2 others raised]

“I don't see an AI solving the technical problems for a lone wolf actor”

Scenario B

Arguments For Raising By More:

I doubled the mean for D by raising the 5% values since the death toll from Covid-19 was higher than average for modern pandemics.

If that virus has, say, a 20% chance of becoming a new global pandemic, then E increases dramatically.

Arguments For Raising By Less:

I didn't meaningfully distinguish between a lab coach and a virology coach, because the lab coach sounded like, in the scenario where it was providing uplift, already able to do this.

The process seems like it would compound any errors in estimation, and I lose track of what I think as I work through

Host-pathogen-environment interaction and socio-cultural-economic factors are not factored in here.

I want to be careful not to "double-dip" since the outcome is multiplicative.

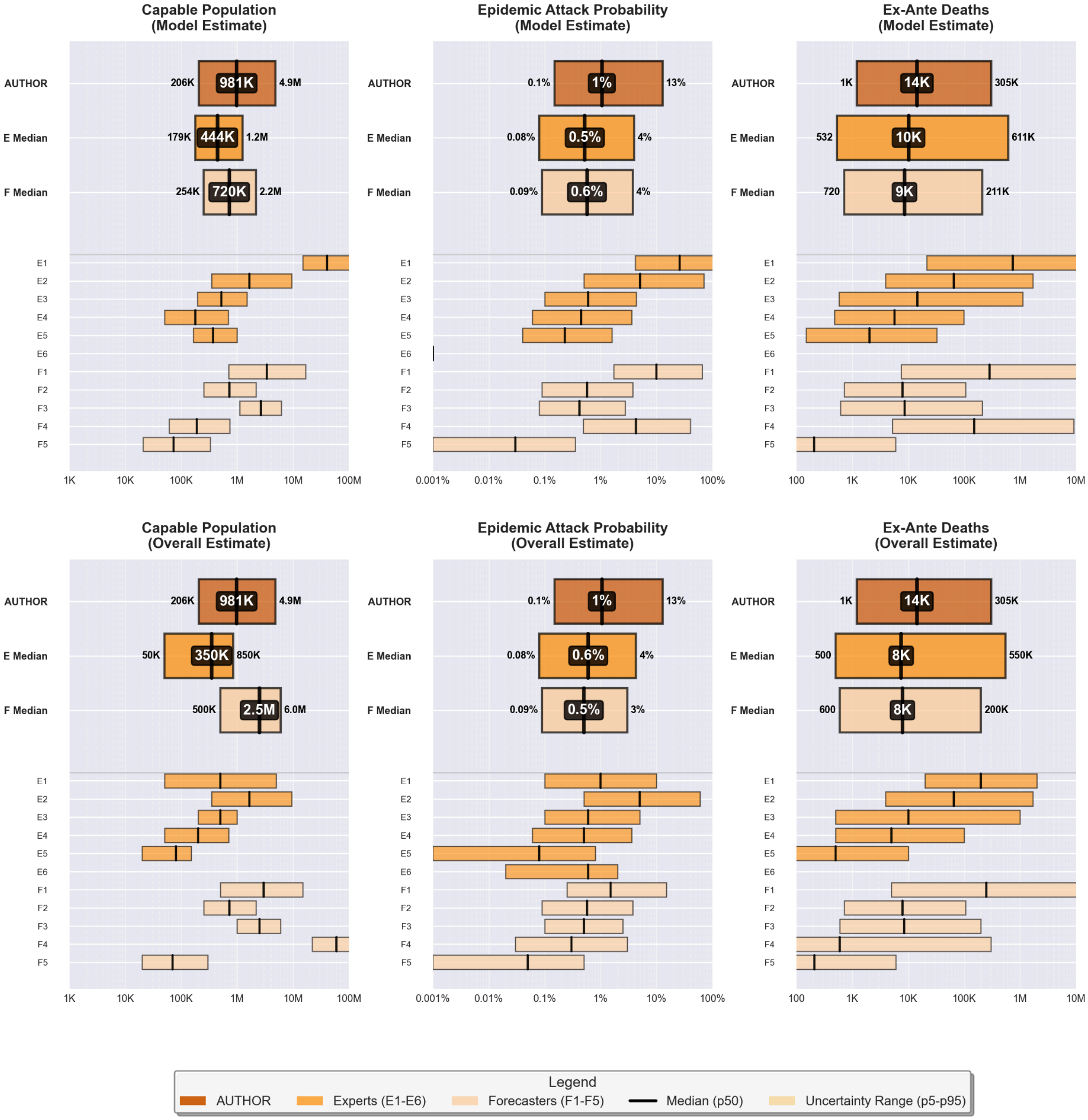
My forecasts here don't change because, to my mind, smallpox already meets this bar.

if the sequence is known from the start, the path to a vaccine could be even faster than before

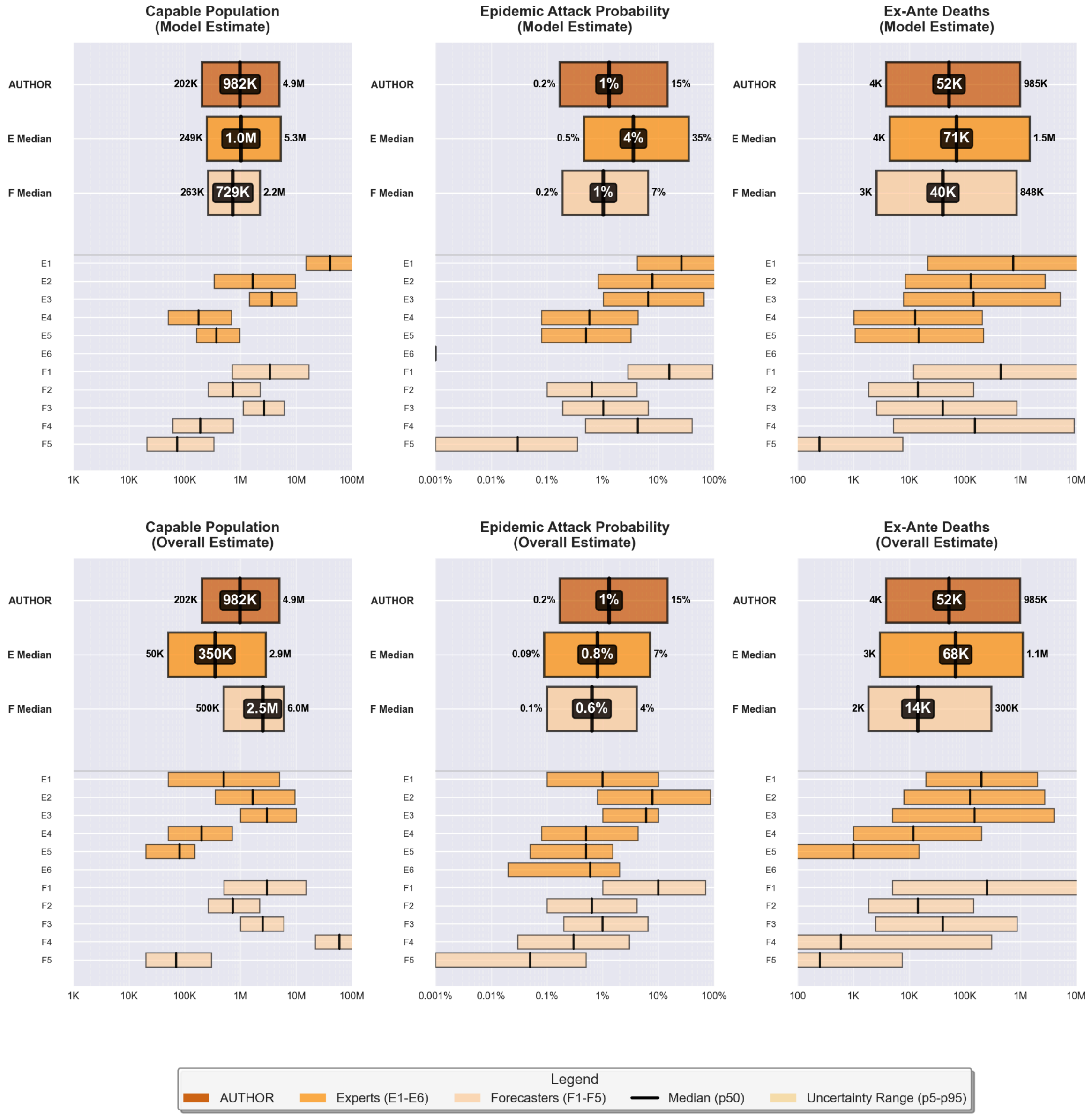
If something similar to Covid-19 emerges in 2028 (less than 10 years from the preceding pandemic), one should assume that the response will be similar in proportion. How could a new similar virus cause 2.5 times the deaths caused by Covid-19

4.3 | Additional Results

Scenario A

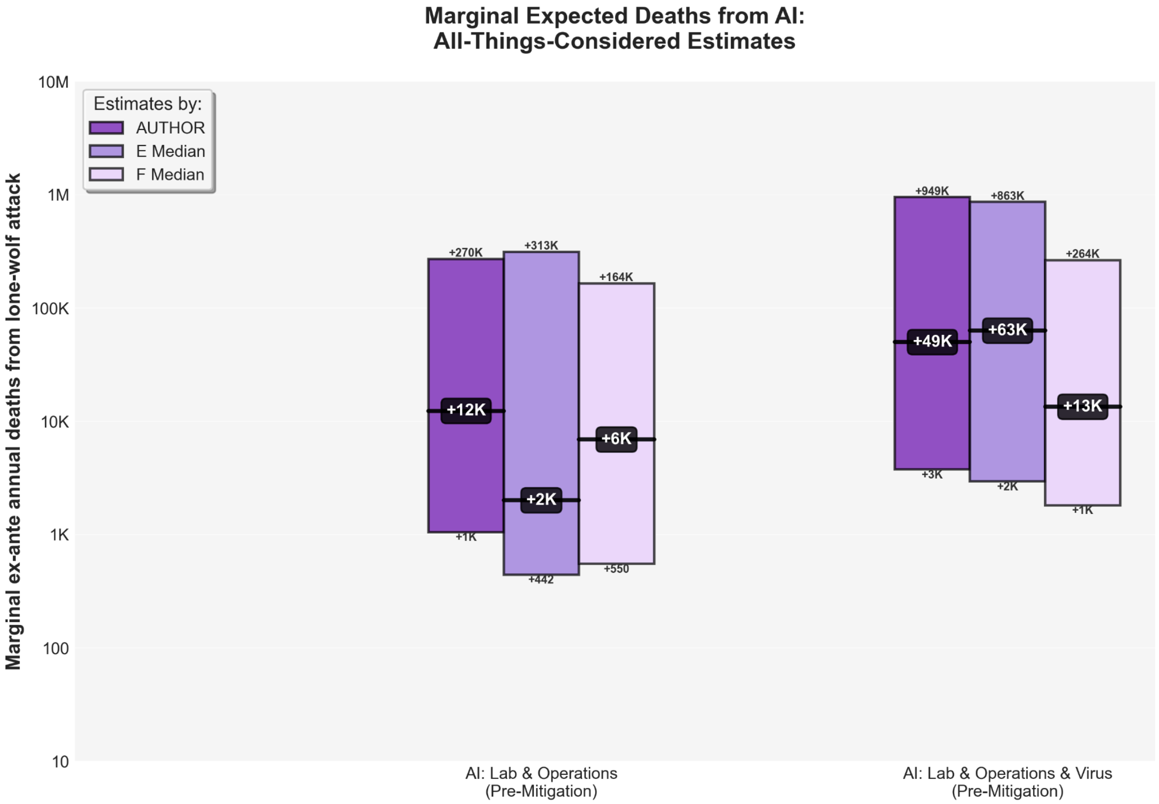


Scenario B


Capable Population (Model Estimate)
Epidemic Attack Probability (Model Estimate)
Ex-Ante Deaths (Model Estimate)
AUTHOR
E Median
F Median
202K
982K
4.9M
249K
1.0M
5.3M
263K
729K
2.2M
0.2%
1%
15%
0.5%
4%
35%
0.2%
1%
7%
4K
52K
985K
4K
71K
1.5M
3K
40K
848K
E1
E2
E3
E4
E5
E6
F1
F2
F3
F4
F5
1K
10K
100K
1M
10M
100M
0.001%
0.01%
0.1%
1%
10%
100%
100
1K
10K
100K
1M
10M
Capable Population (Overall Estimate)
Epidemic Attack Probability (Overall Estimate)
Ex-Ante Deaths (Overall Estimate)
202K
982K
4.9M
50K
350K
2.9M
500K
2.5M
6.0M
0.2%
1%
15%
0.09%
0.8%
7%
0.1%
0.6%
4%
4K
52K
985K
3K
68K
1.1M
2K
14K
300K
Legend
AUTHOR
Experts (E1-E6)
Forecasters (F1-F5)
Median (p50)
Uncertainty Range (p5-p95)

Implied Marginal Risk

We can similarly present these results by subtracting each scenario from the baseline:



References

Ackerman, Gary, and Lauren E. Pinson. “An Army of One: Assessing CBRN Pursuit and Use by Lone Wolves and Autonomous Cells.” *Terrorism and Political Violence* 26, no. 1 (2014). https://doi.org/10.1080/09546553.2014.849945.

Adamala, Katarzyna P., Deepa Agashe, Yasmine Belkaid, Daniela Matias de C. Bittencourt, Yizhi Cai, Matthew W. Chang, Irene A. Chen, et al. “Confronting risks of mirror life.” *Science* 386, no. 6728 (2024): 1351–1353. https://www.science.org/doi/10.1126/science.ads9158.

Adalja, A. A., Watson, M., & Inglesby, T. V. (2018). *Characteristics of Pandemic Pathogens*. Johns Hopkins Center for Health Security. https://centerforhealthsecurity.org/sites/default/files/2022-12/180510-pandemic-pathogens-report.pdf

Amodt, M. G. (2016). *Serial killer statistics*. Radford University/Florida Gulf Coast University. http://maamodt.asp.radford.edu/serial%20killer%20information%20center/serial%20killer%20statistics.pdf

Amodei, Dario. “Machines of Loving Grace.” 2024. https://www.darioamodei.com/essay/machines-of-loving-grace.

Anthropic. *Claude* 3.7 *Sonnet System Card*. 2025. https://www.anthropic.com/claude-3-7-sonnet-system-card.

Apostolakis, George E. “How useful is quantitative risk assessment?” *Risk Analysis* 24, no. 3 (2004): 515–520. https://doi.org/10.1111/j.0272-4332.2004.00455.x.

Asimov. “Making the Micropipette.” Metacelsus on *Asimov Press* (Blog), 2024. https://www.asimov.press/p/making-the-micropipette.

Aven, Terje., Ortwin Renn, and Eugene A. Rosa. “On the ontological status of the concept of risk.” *Safety Science* 47, *no.* 6 (2011): 1074-1079. https://doi.org/10.1016/j.ssci.2011.04.015.

Bakker, Edwin. “Forecasting Terrorism: The Need for a More Systematic Approach.” *Journal of Strategic Security* 5, no. 4 (2012): 69–84. http://www.jstor.org/stable/26463974.

Baum, Seth D., Roman I. de Vere-Mould, and Austin M. Barrett. “A Model for The Probability of Nuclear War.” GCRI Working Paper 042. Global Catastrophic Risk Institute, 2018. https://gcrinstitute.org/papers/042_nuclear-probability.pdf.

Baxter, Lori, Jame Booker, and James Scouras. “Constructing an Elicitation on the Risks of Weapons of Mass Destruction: Lessons from Analyzing the Lugar Survey.” Technical Report. Johns Hopkins University Applied Physics Laboratory, May 2024. https://www.jhuapl.edu/sites/default/files/2024-05/LugarSurveyLiteratureAnalysis.pdf.

Ouagrham-Gormley, Sonia Ben. *Barriers to Bioweapons: The Challenges of Expertise and Organization for Weapons Development*. Cornell University Press, 2014. http://www.jstor.org/stable/10.7591/j.ctt1287dk2.

Boiko, D. A., R. MacKnight, B. G. Kline, and G. Gomes. “Autonomous chemical research with large language models.” *Nature* 624, no. 7992 (2023): 570–578. https://doi.org/10.1038/s41586-023-06792-0.

Broad, William J., and David Johnston. “Anthrax sent through mail gained potency by the letter.” *The New York Times*, May 7, 2002. https://www.nytimes.com/2002/05/07/us/anthrax-sent-through-mail-gained-potency-by-the-letter.html.

Brown, Gerald, and Louis Anthony Cox, Jr. “How Probabilistic Risk Assessment Can Mislead Terrorism Risk Analysis.” *Risk Analysis* 31, no. 2 (2010): 196–204. https://doi.org/10.1111/j.1539-6924.2010.01492.x.

Bunn, Matthew. “A Mathematical Model of the Risk of Nuclear Terrorism.” *The Annals of the American Academy of Political and Social Science* 607, no. 1 (2006): 103–120. https://matthewbunn.scholars.harvard.edu/sites/g/files/omnuum3466/files/matthew_bunn/files/bunn_a_mathematical_model_of_the_risk_of_nuclear_terrorism.pdf.

Carus, Seth W.. *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents Since* 1900. Center for Counterproliferation Research, National Defense University, February 2001. https://wmdcenter.ndu.edu/Portals/97/Documents/Publications/Articles/Bioterrorism-and-Biocrimes.pdf

Carus, W. Seth. A *Short History of Biological Warfare: From Pre-History to the 21st Century*. CSWMD Occasional Paper, no. 12. Washington, D.C.: National Defense University Press, August 2017. https://ndupress.ndu.edu/Portals/68/Documents/occasional/cswmd/CSWMD_OccasionalPaper-12.pdf.

Casper, Stephen and Casper, Stephen and O'Brien, Kyle and O'Brien, Kyle and Longpre, Shayne and Seger, Elizabeth and Klyman, Kevin and Klyman, Kevin and Bommasani, Rishi and Nrusimha, Aniruddha and Nrusimha, Aniruddha and Shumailov, Ilia and Mindermann, Sören and Mindermann, Sören and Basart, Steven and Basart, Steven and Rudzicz, Frank and Rudzicz, Frank and Rudzicz, Frank and Pelrine, Kellin and Ghosh, Avijit and Ghosh, Avijit and Strait, Andrew and Strait, Andrew and Kirk, Robert and Kirk, Robert and Hendrycks, Dan and Hendrycks, Dan and Henderson, Peter and Kolter, J. Zico and Irving, Geoffrey and Irving, Geoffrey and Gal, Yarin and Gal, Yarin and Bengio, Yoshua and Hadfield-Menell, Dylan and Hadfield-Menell, Dylan. *Open Technical Problems in Open-Weight AI Model Risk Management* SSRN, October 2025. https://ssrn.com/abstract=5705186

Christopher, G. W., T. J. Cieslak, J. A. Pavlin, and E. M. Eitzen. “Biological Warfare: A Historical Perspective.” *JAMA* 278, no. 5 (1997): 412–417. https://jamanetwork.com/journals/jama/fullarticle/417896.

Collins, Harry. *Tacit and Explicit Knowledge*. University of Chicago Press, 2010. https://www.degruyterbrill.com/document/doi/10.7208/9780226113821/html.

Corrigan, Jack R., Joshua Dunham, and Remco Zwetsloot. *The Long-Term Stay Rates of International STEM PhD Graduates*. Center for Security and Emerging Technology, 2022. https://cset.georgetown.edu/publication/the-long-term-stay-rates-of-international-stem-phd-graduates/.

Danzig, Richard, Marc Sageman, Terrance Leighton, Lloyd Hough, Hidemi Yuki, Rui Kotani, and Zachary Hosford. *Aum Shinrikyo: Insights into How Terrorists Develop Biological and Chemical Weapons*. Center for a New American Security, 2011. https://www.cnas.org/publications/reports/aum-shinrikyo-insights-into-how-terrorists-develop-biological-and-chemical-weapons.

DeBenedictis, E. A. “Language is not enough.” Substack, 2023. https://erikaaldendeb.substack.com/p/language-is-not-enough.

Defense Science Board. *Department of Defense Biological Safety and Security Program*. May 2009. https://aspr.hhs.gov/S3/Documents/ADA499977.pdf.

DeFrancesco, Laura. “Synthetic virology: the experts speak.” *Nature Biotechnology* 39, no. 10 (2021): 1185-1193. https://www.nature.com/articles/s41587-021-01078-0.

Department for Environment, Food, & Rural Affairs (DEFRA). “Foot and mouth disease 2007: a review and lessons learned.” UK Government. March 11, 2018. https://www.gov.uk/government/publications/foot-and-mouth-disease-2007-a-review-and-lessons-learned.

Dev, Sunishchal, Charles Teague, Kyle Brady, et al. *Toward Comprehensive Benchmarking of the Biological Knowledge of Frontier Large Language Models*. RAND Corporation, 2025. https://www.rand.org/pubs/working_papers/WRA3797-1.html.

Duke University. “All Departments: PhD Completion Rates Statistics.” 2016. https://gradschool.duke.edu/about/statistics/all-departments-phd-completion-rates/.

Duwe, Grant. “Patterns and Prevalence of Lethal Mass Violence.” *Criminology & Public Policy* 19, no. 1 (2020): 17–35. https://doi.org/10.1111/1745-9133.12478.

Duwe, Grant, Nathan E. Sanders, Michael Rocque, and James Alan Fox. “Estimating the Global Prevalence of Mass Public Shootings.” *International Journal of Offender Therapy and Comparative Criminology*, 67, no. 16 (2023): 1642-1658. https://doi.org/10.1177/0306624X221139070.

Ellis, Clare, Raffaello Pantucci, Benoit Gomis, Simon Palombi, and Melanie Smith. “Analysing the Processes of Lone-Actor Terrorism: Research Findings.” *Perspectives on Terrorism* 10, no. 2 (2016): 33–41. https://pt.icct.nl/article/analysing-processes-lone-actor-terrorism-research-findings

Emmons, William R., Ana H. Kent, and Lowell R. Ricketts. “Is College Still Worth It? The New Calculus of Falling Returns.” *Federal Reserve Bank of St. Louis Review* 101, no. 4 (2019): 297–329. https://doi.org/10.20955/r.101.297-329.

Esvelt, Kevin. “How a deliberate pandemic could crush societies—and what to do about it.” *Bulletin of the Atomic Scientists*, November 15, 2022. https://thebulletin.org/2022/11/how-a-deliberate-pandemic-could-crush-societies-and-what-to-do-about-it/.

Esvelt, Kevin. “Credible pandemic virus identification will trigger the immediate proliferation of agents as lethal as nuclear devices.” *Testimony before the Senate Committee on Homeland Security and Governmental Affairs*, 2022. https://www.hsgac.senate.gov/wp-content/uploads/imo/media/doc/Esvelt%20Testimony.pdf.

Esvelt, Kevin. “It shouldn’t be this easy to buy the synthetic DNA fragments to recreate the deadly 1918 flu virus.” STAT, 2024. https://www.statnews.com/2024/05/08/shouldnt-be-easy-buy-synthetic-dna-fragments-recreate-deadly-1918-flu-virus/.

Ezell, Barry Charles, Steven P. Bennett, Detlof von Winterfeldt, John Sokolowski, and Andrew J. Collins. “Probabilistic Risk Analysis and Terrorism Risk.” *Risk Analysis: An International Journal* 30, no. 4 (2010): 575–589. https://doi.org/10.1111/j.1539-6924.2010.01401.x.

Fabri, Peter J., and José L. Zayas-Castro. “Human error, not communication and systems, underlies surgical complications.” *Surgery* 144, no. 4 (2008): 557–565. https://doi.org/10.1016/j.surg.2008.06.011.

Fan, Victoria Y., Dean T. Jamison, and Lawrence H. Summers. “The Loss from Pandemic Influenza Risk.” In *Disease Control Priorities: Improving Health and Reducing Poverty*, 3rd ed., edited by Dean T. Jamison, Helena Gelband, Susan Horton, et al. The International Bank for Reconstruction and Development / The World Bank, 2017. https://www.ncbi.nlm.nih.gov/books/NBK525291/.

Fischhoff, B., de Bruin, W.B., Güvenç, Ü., Caruso, D., & Brilliant, L. “Analyzing disaster risks and plans: An avian flu example.” *Journal of Risk Uncertainty* 33, 131–149 (2006). https://doi.org/10.1007/s11166-006-0175-8

Flade, Florian. “The June 2018 Cologne Ricin Plot: A New Threshold in Jihadi Bio Terror.” CTC Sentinel 11, no. 7 (August 2018). https://ctc.westpoint.edu/june-2018-cologne-ricin-plot-new-threshold-jihadi-bio-terror/.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. *IPUMS USA: Version* 13.0 [dataset]. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D010.V13.0.

Förster, Klaus.. *Universities Worldwide*. 2022. https://univ.cc/.

Fouchier, Ron A. M. “Studies on Influenza Virus Transmission between Ferrets: The Public Health Risks Revisited.” *mBio* 6, no. 1 (2015): e02560–14. https://doi.org/10.1128/mBio.02560-14.

Frontier Model Forum (FMF). *Issue Brief: Preliminary Taxonomy of AI Bio-Safety Evaluations*. 2024. https://www.frontiermodelforum.org/updates/issue-brief-preliminary-taxonomy-of-ai-bio-safety-evaluations/

Frontier Model Forum. *Latest from the FMF: Grant-Making to Address AI-Bio Risk Challenges*. 2025. https://www.frontiermodelforum.org/updates/latest-from-the-fmf-grant-making-to-address-ai-bio-risk-challenges/

FutureHouse. “LAB-Bench: Measuring Capabilities of Language Models for Biology Research.” *arXiv*. Submitted July 14, 2024. https://doi.org/10.48550/arXiv.2407.10362

Gill, Paul, John Horgan, and Paige Deckert. “Bombing Alone: Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorists.” *Journal of Forensic Sciences* 59, no. 2 (2014): 425–435. https://doi.org/10.1111/1556-4029.12312.

Glennerster, Rachel, Christopher M. Snyder, and Brandon Joel Tan. “Calculating the Costs and Benefits of Advance Preparations for Future Pandemics.” NBER Working Paper No. 30565. National Bureau of Economic Research, October 2022. https://doi.org/10.3386/w30565.

Götting, Jasper, Pedro Medeiros, Jon G. Sanders, et al. “Virology Capabilities Test (VCT): A Multimodal Virology Q&A Benchmark.” *arXiv*. Submitted April 21, 2025. https://doi.org/10.48550/arXiv.2504.16137.

Good Judgment. “Superforecasting the Origins of the COVID-19 Pandemic.” *Good Judgment Substack*, March 11, 2024. https://goodjudgment.substack.com/p/superforecasting-the-origins-of-the.

Goodwin, Paul, and George Wright. *Decision Analysis for Management Judgment*. Wiley, 2009. https://www.wiley.com/en-us/Decision+Analysis+for+Management+Judgment%2C+5th+Edition-p-9781118740736.

Gopal, Anjali, Nathan Helm-Burger, Lennart Justen, et al. “Will releasing the weights of future large language models grant widespread access to pandemic agents?” *arXiv*. Submitted October 25, 2023. https://doi.org/10.48550/arXiv.2310.18233.

Google DeepMind. *Gemini 2.5 Pro Model Card*. Last updated June 17, 2025. https://modelcards.withgoogle.com/assets/documents/gemini-2.5-pro.pdf.

Grant, Matthew, and Mark G. Stewart. “A systems model for probabilistic risk assessment of improvised explosive device attack.” *International Journal of Intelligent Defence Support Systems* 5, no. 1 (2012): 75–93. https://doi.org/10.1504/IJIDSS.2012.053664.

Greene, Dan, Jassi Pannu, and Allison Berke. "The Danger of 'Invisible' Biolabs Across the U.S." *Time*, August 31, 2023. https://time.com/6309643/invisible-biolabs/.

Gryphon Scientific. *Risk and Benefit Analysis (RBA) of Gain of Function Research*. 2015. https://osp.od.nih.gov/wp-content/uploads/Risk_and_Benefit_Analysis_of_Gain-of-Function_Research.pdf

Gryphon Scientific. *Risk and Benefit Analysis (RBA) of Gain of Function Research*. 2016. https://gryphonsci.wpengine.com/wp-content/uploads/2018/12/Risk-and-Benefit-Analysis-of-Gain-of-Function-Research-Final-Report-1.pdf

Hamm, Mark S., and Ramón Spaaij. *Lone Wolf Terrorism in America: Using Knowledge of Radicalization Pathways to Forge Prevention Strategies*. Final Report, NCJ 248691. Washington, D.C.: National Institute of Justice, January 2015. https://www.ojp.gov/pdffiles1/nij/grants/248691.pdf.

Herre, Bastian, and Fiona Spooner. "Homicide data: how sources differ and when to use which one." Our World in Data, 2023. https://ourworldindata.org/homicide-data-how-sources-differ-and-when-to-use-which-one.

Hummel, Stephen. "The Islamic State and WMD: Assessing the Future Threat." *CTC Sentinel* 9, no. 1 (2016): 1–9. https://ctc.westpoint.edu/the-islamic-state-and-wmd-assessing-the-future-threat/.

International AI Safety Report 2025. London: Department for Science, Innovation and Technology (DSIT), 2025. https://www.gov.uk/government/publications/international-ai-safety-report-2025.

International Gene Synthesis Consortium (IGSC). "Why a misleading 'red-team' study of the gene synthesis industry wrongly casts doubt on industry safety." *The Bulletin of the Atomic Scientists*. June 3, 2024. https://thebulletin.org/2024/06/why-a-misleading-red-team-study-of-the-gene-synthesis-industry-wrongly-casts-doubt-on-industry-safety/.

Inglesby, Thomas V., and David A. Relman. "How likely is it that biological agents will be used deliberately to cause widespread harm?" *EMBO Reports* 17, no. 2 (2016): 127–130. https://doi.org/10.15252/embr.201541674.

JASON. *Rare events*. The MITRE Corporation, 2009. https://fas.org/irp/agency/dod/jason/rare.pdf.

Jefferson, Catherine, Filippa Lentzos, and Claire Marris. "Synthetic Biology and Biosecurity: Challenging the 'Myths'." *Frontiers in Public Health* 2 (2014): 1–15. https://doi.org/10.3389/fpubh.2014.00115.

Johari, Md Radzi. "Anthrax - Biological Threat in the 21st Century." *The Malaysian Journal of Medical Sciences* 9, no. 1 (2002): 1-2. https://pmc.ncbi.nlm.nih.gov/articles/PMC3436101/.

Joiner, Emily. "Rethinking the value of a statistical life." *Common Resources* (Blog), September 21, 2023. https://www.rff.org/publications/explainers/rethinking-the-value-of-a-statistical-life/.

Justen, Lennart. "LLMs Outperform Experts on Challenging Biology Benchmarks." *arXiv*. Submitted May 9, 2025. https://doi.org/10.48550/arXiv.2505.06108.

Kane, Arianne, and Michael T. Parker. "Screening State of Play: The Biosecurity Practices of Synthetic DNA Providers." *Applied Biosafety* 29, no. 2 (2024): 85–95. https://doi.org/10.1089/apb.2023.0027.

Kapoor, Sayash, Rishi Bommasani, Kevin Klyman, et al. "On the Societal Impact of Open Foundation Models." Stanford University Center for Research on Foundation Models, 2024. https://crfm.stanford.edu/open-fms/.

Kenyon, Jonathan, Christopher Baker-Beall, and Jens Binder. "Lone-Actor Terrorism – A Systematic Literature Review." Studies in Conflict & Terrorism 46, no. 10 (2023): 2038–2065. https://doi.org/10.1080/1057610X.2021.1892635.

Klotz, Lynn C. "Comments on Fouchier's Calculation of Risk and Elapsed Time for Escape of a Laboratory-Acquired Infection from His Laboratory." *mBio* 6, no. 2 (2015): e00268–15. https://doi.org/10.1128/mBio.00268-15.

Koblentz, Gregory D. “Predicting Peril or the Peril of Prediction? Assessing the Risk of CBRN Terrorism.” *Terrorism and Political Violence* 23, no. 4 (2011): 501–520. https://doi.org/10.1080/09546553.2011.575487.

Koblentz, Gregory D. “The De Novo Synthesis of Horsepox Virus: Implications for Biosecurity and Recommendations for Preventing the Reemergence of Smallpox.” *Health Secur.* 15, no. 6 (2017): 620-628. https://doi.org/10.1089/hs.2017.0061.

Koblentz, Gregory D., and Stevie Kiesel. “The COVID-19 Pandemic: Catalyst or Complication for Bioterrorism?” *Studies in Conflict and Terrorism* 47, no. 2 (2021): 154-180. https://doi.org/10.1080/1057610X.2021.1944023.

Koessler, Leonie, Jonas Schuett, and Marcus Anderljung. “Risk thresholds for frontier AI.” . *arXiv*, June 20, 2024. https://arxiv.org/abs/2406.14713.

Lancet Infectious Diseases. (2024). “What is the pandemic potential of avian influenza A(H5N1)?” *The Lancet Infectious Diseases* 24, no. 5 (2024): 437. https://doi.org/10.1016/S1473-3099(24)00238-X

Ledford, Heidi. “Garage biotech: Life hackers.” *Nature* 467, no. 7316 (2010): 650–652. https://doi.org/10.1038/467650a.

Lentzos, Filippa, Jez Littlewood, Hailey Wingo, and Alberto Muti. “Apathy and hyperbole cloud the real risks of AI bioweapons.” *The Bulletin of the Atomic Scientists*, September 12, 2024. https://thebulletin.org/2024/09/apathy-and-hyperbole-cloud-the-real-risks-of-ai-bioweapons/.

Lewis, Gregory, Paul Millett, Anders Sandberg, Andrew Snyder-Beattie, and Gigi Gronvall. “Information Hazards in Biotechnology.” *Risk Analysis* 39, no. 5 (2019): 975–981. https://doi.org/10.1111/risa.13235.

Lipsitch, Marc. “Why Do Exceptionally Dangerous Gain-of-Function Experiments in Influenza?” In *Influenza Virus: Methods and Protocols*, edited by Yohei Yamauchi, 589–608. New York, NY: Humana Press, 2018. https://doi.org/10.1007/978-1-4939-8678-1_29.

Lipsitch, Marc, and Thomas V. Inglesby. “Moratorium on Research Intended To Create Novel Potential Pandemic Pathogens.” *mBio* 5, no. 6 (2014). https://doi.org/10.1128/mBio.02366-14.

Lugar, Richard G. *The Lugar Survey on Proliferation Threats and Responses*. Report. US Senate, June 2005. https://irp.fas.org/threat/lugar_survey.pdf.

Marani, Marco, Gabriel G. Katul, William K. Pan, and Anthony J. Parolari. “Intensity and frequency of extreme novel epidemics.” *Proceedings of the National Academy of Sciences* 118, no. 35 (2021): e2105482118. https://doi.org/10.1073/pnas.2105482118.

Martínez-Sobrido, Luis, and Adolfo García-Sastre. “Generation of recombinant influenza virus from plasmid DNA.” *Journal of Visualized Experiments*, 42 (2010): 2057. https://doi.org/10.3791/2057.

Mashaw, Jerry L., and David L. Harfst. *The Struggle for Auto Safety*. Harvard University Press, 1990.

Meselson, M., J. Guillemin, M. Hugh-Jones, A. Langmuir, I. Popova, A. Shelokov, and O. Yampolskaya. “The Sverdlovsk anthrax outbreak of 1979.” *Science* 266, no. 5188 (1994): 1202–1208. https://www.science.org/doi/10.1126/science.7973702.

Metaculus. “Pandemic series: a significant bioterror attack by 2020.” Forecast opened in June 2016; accessed October 18, 2025. https://www.metaculus.com/questions/254/pandemic-series-a-significant-bioterror-attack-by-2020/.

Mettler-Toledo. “Pipetting Tips and Tricks.” n.d. https://www.mt.com/us/en/home/campaigns/product-organizations/pipe/Rainin_Tips_Tricks.html?elq_mid=4950&elq_cid=22297913.

Mollick, Ethan. “The Present Future: AI's Impact Long Before Superintelligence.” *One Useful Thing*, November 4, 2024. https://www.oneusefulthing.org/p/the-present-future-ais-impact-long.

Montague, Michael. “Towards a Grand Unified Threat Model of Biotechnology.” *PhilSci Archive*, 2023. https://philsci-archive.pitt.edu/22539/.

Mouton, Christopher A., Caleb Lucas, and Ella Guest, The Operational Risks of AI in Large-Scale Biological Attacks: A Red-Team Approach. Santa Monica, CA: RAND Corporation, 2023. https://www.rand.org/pubs/research_reports/RRA2977-1.html.

Mouton, Christopher A., Caleb Lucas, and Ella Guest, The Operational Risks of AI in Large-Scale Biological Attacks: Results of a Red-Team Study. Santa Monica, CA: RAND Corporation, 2024. https://www.rand.org/pubs/research_reports/RRA2977-2.html.

Mueller, John, *Atomic Obsession: Nuclear Alarmism from Hiroshima to Al-Qaeda* (New York, NY, 2009) Oxford Academic: 2023, https://doi.org/10.1093/oso/9780195381368.001.0001.

Murray, Malcolm, Henry Papadatos, Otter Quarks, Pierre-François Gimenez, and Simeon Campos. “Mapping AI Benchmark Data to Quantitative Risk Estimates Through Expert Elicitation.” *arXiv*. Submitted March 6, 2025. https://doi.org/10.48550/arXiv.2503.04299.

Narayanan, Arvind, and Sayash Kapoor. “AI existential risk probabilities: A skeptical view.” *AI As Normal Technology* (blog). https://www.normaltech.ai/p/ai-existential-risk-probabilities.

National Academies of Sciences, Engineering, and Medicine (NASEM). *Department of Homeland Security Bioterrorism Risk Assessment: A Call for Change*. The National Academies Press, 2008. https://www.nap.edu/catalog/12206/department-of-homeland-security-bioterrorism-risk-assessment-a-call-for-change

NASEM. 2008 Amendments to the National Academies' Guidelines for Human Embryonic Stem Cell Research. The National Academies Press, 2008. https://nap.nationalacademies.org/catalog/12260/2008-amendments-to-the-national-academies-guidelines-for-human-embryonic-stem-cell-research.

NASEM. *The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises*. The National Academies Press, 2016. https://www.ncbi.nlm.nih.gov/books/NBK368393/.

NASEM. *Biodefense in the Age of Synthetic Biology*. The National Academies Press, 2018. https://doi.org/10.17226/24890.

NASEM. *Risk Analysis Methods for Nuclear War and Nuclear Terrorism*. The National Academies Press, 2023. https://nap.nationalacademies.org/catalog/26609/risk-analysis-methods-for-nuclear-war-and-nuclear-terrorism.

NASEM. *Future State of Smallpox Medical Countermeasures*. The National Academies Press, 2024. https://nap.nationalacademies.org/catalog/27652/future-state-of-smallpox-medical-countermeasures.

NASEM. *Chemical Terrorism: Assessment of U.S. Strategies in the Era of Great Power Competition*. The National Academies Press, 2024. https://doi.org/10.17226/27159.

National Commission on Terrorist Attacks upon the United States. *The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks upon the United States*. Washington, D.C.: Government Printing Office, 2004. https://www.govinfo.gov/app/details/GPO-911REPORT.

NCES. *Digest of education statistics*. 2024. https://nces.ed.gov/programs/digest/.

Neumann, Gabriele. “Influenza Reverse Genetics-Historical Perspective.” *Cold Spring Harbor Perspectives in Medicine* 11, no. 4 (2021):a038547. https://doi.org/10.1101/cshperspect.a038547.

Neumann, Gabriele, and Yoshihiro Kawaoka. "Which Virus Will Cause the Next Pandemic?" *Viruses* 15, no. 1 (2023): 199. https://doi.org/10.3390/v15010199.

NIST. *Managing Misuse Risk for Dual-Use Foundation Models*. NIST AI 800-1 ipd2 (Second Public Draft). National Institute of Standards and Technology, U.S. AI Safety Institute, January 2025. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.800-1.ipd2.pdf.

NIST. "Updated Guidelines for Managing Misuse Risk for Dual-Use Foundation Models." National Institute of Standards and Technology, U.S. AI Safety Institute, January 15, 2025. https://www.nist.gov/news-events/news/2025/01/updated-guidelines-managing-misuse-risk-dual-use-foundation-models.

Noyce, Ryan, Seth Lederman, and David H. Evans. "Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments." *PLoS One* 13 no. 1 (2018): e0188453. https://doi.org/10.1371/journal.pone.0188453.

OpenAI. *O1 System Card*. 2024. https://assets.ctfassets.net/kftzwdyauwt9/67qJD51Aur3eIc96iOfeOP/71551c3d223cd97e591aa89567306912/o1_system_card.pdf.

OpenAI. *Building an early warning system for LLM-aided biological threat creation*. 2024. https://openai.com/research/building-an-early-warning-system-for-llm-aided-biological-threat-creation

OpenAI. *Preparing for future AI capabilities in biology*. 2025. https://openai.com/index/preparing-for-future-ai-capabilities-in-biology/.

OpenAI. *Deep Research System Card*. 2025. https://openai.com/index/deep-research-system-card/.

OpenAI. *Preparedness Framework*, *Version 2*. 2025. https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbddebcd/preparedness-framework-v2.pdf.

OpenAI. *Accelerating life sciences research*. 2025. https://openai.com/index/accelerating-life-sciences-research-with-retro-biosciences/.

Oscarsson, Martin, Per Carlbring, Gerhard Andersson, and Alexander Rozental. "A large-scale experiment on New Year's resolutions: Approach-oriented goals are more successful than avoidance-oriented goals." PLoS One 15, no. 12 (2020): e0234097. https://doi.org/10.1371/journal.pone.0234097.

Pannu, Jaspreet. "Protocols and risks: when less is more." *Nature Protocols* 17, no. 1 (2021): 1. https://doi.org/10.1038/s41596-021-00655-6.

Pannu, Jaspreet, Sarah L. Gebauer, Henry Alexander Bradley, et al. *Defining Hazardous Capabilities of Biological AI Models: Expert Convening to Inform Future Risk Assessment*. RAND Corporation, 2025. https://www.rand.org/pubs/conf_proceedings/CFA3649-1.html.

Parachini, John V., and Rohan Kumar Gunaratna. *Implications of the Pandemic for Terrorist Interest in Biological Weapons: Islamic State and al-Qaeda Pandemic Case Studies*. RAND Corporation, 2022. https://www.rand.org/pubs/research_reports/RRA612-1.html.

Paris AI Action Summit. *Program*. 2025. https://aiconference.ip-paris.fr/program/#:~:text=Panel%202%3A%20Setting%20thresholds%20in%20practice.

Paskov, Patricia, Michael J. Byun, Kevin Wei, and Toby Webster, *Preliminary suggestions for rigorous GPAI model evaluations*. RAND Corporation, 2025. https://www.rand.org/pubs/perspectives/PEA3971-1.html.

Peppin, Aidan, Anka Reuel, Stephen Casper, et al. "The Reality of AI and Biorisk." *arXiv*. Submitted December 2, 2024. https://doi.org/10.48550/arXiv.2412.01946.

Piper, Kelsey. "Can we stop the next pandemic by seeking out deadly viruses in the wild?." *Vox*, May 7, 2022. https://www.vox.com/future-perfect/2022/5/7/22973296/virus-hunting-discovery-deep-vzn-global-virome-project/.

PREDICT Consortium. (n.d.). https://p2.predict.global/

Revill, James, and Catherine Jefferson. "Tacit knowledge and the biological weapons regime." *Science and Public Policy* 41, no. 5 (2014): 597–610. https://doi.org/10.1093/scipol/sct090.

Rose, S., Moulange, R., Smith, J., & Nelson, C. (2024). *The near-term impact of AI on biological misuse. The Centre for Long-Term Resilience.* https://www.longtermresilience.org/reports/the-near-term-impact-of-ai-on-biological-misuse/

Roth, John, Douglas Greenburg, and Serena Wille. *Monograph on Terrorist Financing*. National Commission on Terrorist Attacks Upon the United States (9/11 Commission), 2004. https://govinfo.library.unt.edu/911/staff_statements/911_TerrFin_Monograph.pdf.

Rozo, Michelle, and Gigi Kwik Gronvall. "The Reemergent 1977 H1N1 Strain and the Gain-of-Function Debate." *mBio* 6, no. 4 (2015). doi:10.1128/mBio.01013-15. https://pmc.ncbi.nlm.nih.gov/articles/PMC4542197/.

Russell, Charles A., and Bowman H. Miller. "Profile of a Terrorist." Military Review 57, no. 8 (August 1977): 21–34. https://www.ojp.gov/ncjrs/virtual-library/abstracts/profile-terrorist.

Salama, Sammy, and Lydia Hansell. "Does Intent Equal Capability? Al-Qaeda and Weapons of Mass Destruction." *The Nonproliferation Review* 12, no. 3 (2005): 615–53. https://doi.org/10.1080/10736700600601236.

Sandberg, Anders, and Cassidy Nelson. "Who Should We Fear More: Biohackers, Disgruntled Postdocs, or Bad Governments? A Simple Risk Chain Model of Biorisk." *Health Security* 18, no. 3 (2020): 155–163. https://doi.org/10.1089/hs.2019.0115.

Sandberg, Anders, Eric Drexler, and Toby Ord. "Dissolving the Fermi Paradox." *arXiv*, June 8, 2018. arXiv:1806.02404. https://arxiv.org/abs/1806.02404.

Sandbrink, Jonas B. "Artificial intelligence and biological misuse: Differentiating risks of language models and biological design tools." *arXiv*, June 24, 2023. arXiv:2306.13952. https://arxiv.org/abs/2306.13952.

Schuurman, Bart, Edwin Bakker, Paul Gill, and Noémie Bouhana. "Lone Actor Terrorist Attack Planning and Preparation: A Data-Driven Analysis." *Journal of Forensic Sciences* 63, no. 4 (2018): 1191–1200. https://doi.org/10.1111/1556-4029.13676.

Seger, Elizabeth, Noemi Dreksler, Richard Moulange, et al.. "Open-Sourcing Highly Capable Foundation Models: An Evaluation of Risks, Benefits, and Alternative Methods for Pursuing Open-Source Objectives." Research Paper. *Centre for the Governance of* AI, September 29, 2023. https://www.governance.ai/research-paper/open-sourcing-highly-capable-foundation-models.

Select Committee on the Strategic Competition Between the United States and the Chinese Communist Party. *Investigation into the Reedley Biolab: Report*. November 15, 2023. https://selectcommitteeontheccp.house.gov/sites/evo-subsites/selectcommitteeontheccp.house.gov/files/evo-media-document/scc-reedley-report-11.15.pdf.

Simons, Erica. *Faith, fanaticism, and fear: Aum Shinrikyo, the birth and death of a terrorist cult*. U.S. Department of Justice, Office of Justice Programs, 2006.

https://www.ojp.gov/ncjrs/virtual-library/abstracts/faith-fanaticism-and-fear-aum-shinrikyo-birth-and-death-terrorist.

Simon, J. (2013). *What makes lone-wolf terrorists so dangerous?* https://newsroom.ucla.edu/stories/what-makes-lone-wolfe-terrorists-245316

Smith, Brent L., Paxton Roberts; Jeff Gruenewald; Brent R. Klein. *Patterns of Lone Actor Terrorism in the United States: Research Brief*, NCJ 304688. National Institute of Justice, 2015. https://www.ojp.gov/ncjrs/virtual-library/abstracts/patterns-lone-actor-terrorism-united-states-research-brief.

Soice, Emily H., Rafael Rocha, Kimberlee Cordova, Michael Specter, and Kevin M. Esvelt. "Can large language models democratize access to dual-use biotechnology?" *arXiv*, June 6, 2023. https://arxiv.org/abs/2306.03809.

Sperandei, Sandro, Marcelo C. Vieira, and Arianne C. Reis. "Adherence to physical activity in an unsupervised setting: Explanatory variables for high attrition rates among fitness center members." *Journal of Science and Medicine in Sport* 19, no. 11 (2016): 916–920. https://doi.org/10.1016/j.jsams.2015.12.522.

STATSIG, "Credible intervals vs confidence intervals: A Bayesian twist". https://www.statsig.com/perspectives/credible-vs-confidence-intervals

Stewart, Mark G., & John Mueller. "Risk and economic assessment of U.S. aviation security for passenger-borne bomb attacks." *Journal of Transportation Security* (2018): 11: 117-136. https://doi.org/10.1007/s12198-018-0196-y.

Tetlock, Philip E. *Expert political judgment: How good is it? How can we know?* Princeton University Press, 2025. https://psycnet.apa.org/record/2005-09833-000.

Tetlock, Philip E., Barbara A. Mellers, and J. Peter Scoblic. "Bringing probability judgments into policy debates via forecasting tournaments." *Science* 355, no. 6324 (2017): 481–483. https://doi.org/10.1126/science.aal3147.

The White House. "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI." September 12, 2023, https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

Tin, Derrick, Pardis Sabeti, and Gregory R. Ciottone. "Bioterrorism: An analysis of biological agents used in terrorist events." *The American Journal of Emergency Medicine* 54 (2022): 117–121. https://www.sciencedirect.com/science/article/pii/S0735675722000602

Torres, Phil. "Who would destroy the world? Omnicidal agents and related phenomena." *Aggression and Violent Behavior* 39 (2018): 129-138. https://doi.org/10.1016/j.avb.2018.02.002.

Thadani, Nicole N., Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J. Rollins, Daniel Ritter, Chris Sander, et al. "Learning from prepandemic data to forecast viral escape." *Nature* 622, no. 7984 (2023): 818–825. https://doi.org/10.1038/s41586-023-06617-0.

UNESCO Institute for Statistics. "UIS Data Browser." Accessed October 19, 2025. https://databrowser.uis.unesco.org/. UNESCO. (2021). *UNESCO science report: The race against time for smarter development*. https://www.unesco.org/reports/science/2021/en

United Nations. "UN Investigative Team Outlines Findings around ISIL Chemical Weapons Use." UN *News*, June 8, 2023. https://news.un.org/en/story/2023/06/1137492.

U.S. National Science and Technology Council. *Framework For Nucleic Acid Synthesis Screening*. September 2024. https://bidenwhitehouse.archives.gov/wp-content/uploads/2024/10/OSTP-Nucleic-Acid_Synthesis_Screening_Framework-Sep2024-Final.pdf

U.S. Department of Health & Human Services. *Screening Framework Guidance for Providers and Users of Synthetic Nucleic Acids*. October 2023. https://aspr.hhs.gov/S3/Documents/SynNA-Guidance-2023.pdf

U.S. Office of Technology Assessment. *Proliferation of Weapons of Mass Destruction: Assessing the Risks*. August 1993. https://ota.fas.org/reports/9341.pdf.

National Consortium for the Study of Terrorism and Responses to Terrorism (START). "Violent Non-State Actor CBRN Data Portal." Unconventional Weapons & Technology Division (UWT), University of Maryland, n.d. https://cbrn.umd.edu/.

Westwood, Sean J., Justin Grimmer, Matthew Tyler, and Clayton Nall. "Current Research Overstates American Support for Political Violence." *Proceedings of the National Academy of Sciences* 119, no. 12 (2022): e2116870119. https://doi.org/10.1073/pnas.2116870119.

Wheeler, Nicole E., Craig Bartling, Sarah R. Carter, et al. "Progress and Prospects for a Nucleic Acid Screening Test Set." *Applied Biosafety* 29, no. 3 (2024): 133–141. https://doi.org/10.1089/apb.2023.0033.

Wikipedia. "List of United States presidential assassination attempts and plots." https://en.wikipedia.org/wiki/List_of_United_States_presidential_assassination_attempts_and_plots.

Williams, Heather J., Nathan Chandler, and Eric Robinson, *Trends in the Draw of Americans to Foreign Terrorist Organizations from 9/11 to Today*. RAND Corporation, 2018. https://www.rand.org/pubs/research_reports/RR2545.html.

Williams, Bridget, Luca Righetti, Josh Rosenberg, et al. "Forecasting LLM-enabled Biorisk and the Efficacy of Safeguards," Forecasting Research Institute, July 1, 2025. https://static1.squarespace.com/static/635693acf15a3e2a14a56a4a/t/68812b62e85b2808f0366c41/1753295738891/ai-enabled-biorisk.pdf.

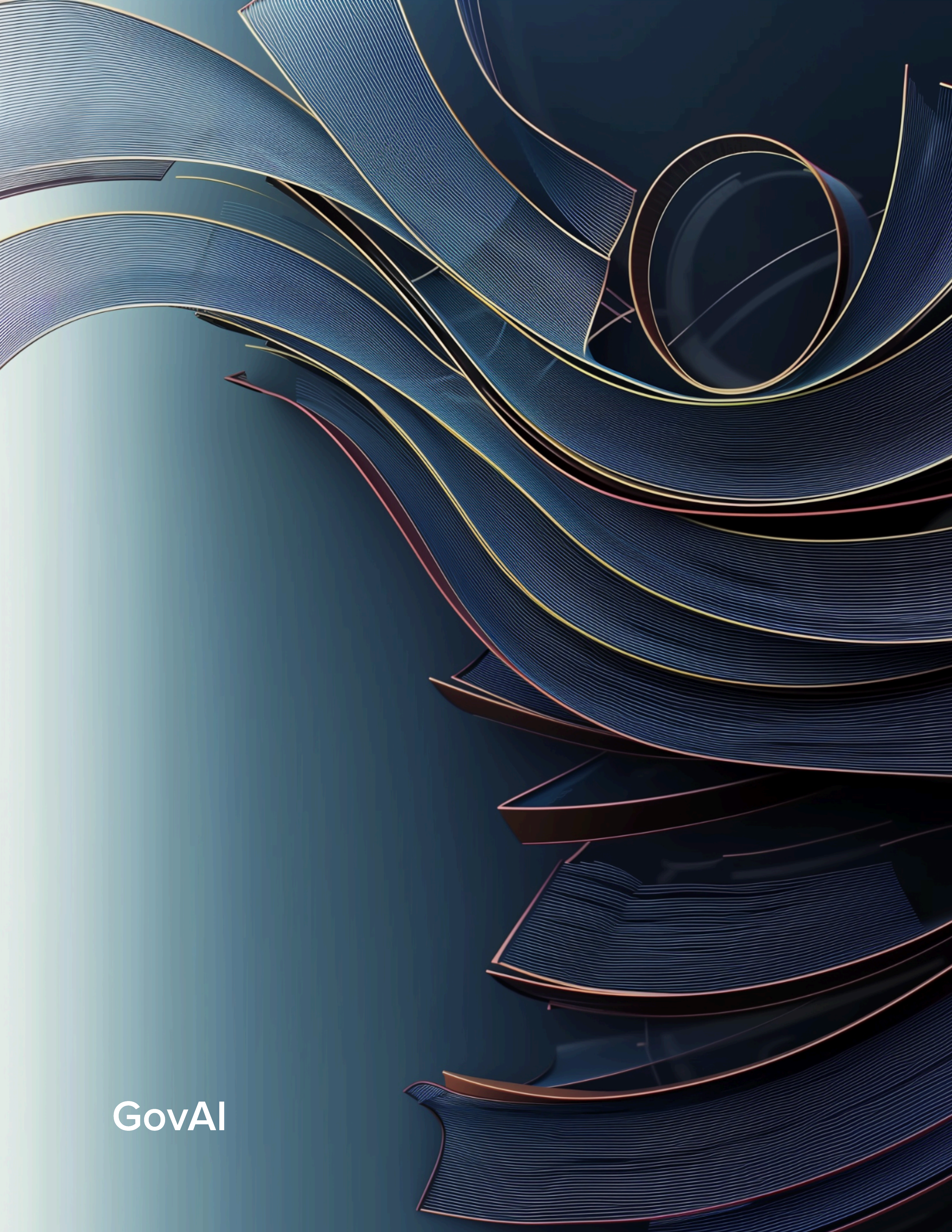
World Bank. "World development indicators." n.d. https://datatopics.worldbank.org/world-development-indicators/.

World Health Organization (WHO). "The Independent Advisory Group on Public Health Implications of Synthetic Biology Technology Related to Smallpox." Geneva: World Health Organization, 2004. https://iris.who.int/server/api/core/bitstreams/398b53c2-a72a-4258-970e-8331e003ef8d/content.

WHO. "Ensuring responsible use of life sciences research." n.d.https://www.who.int/activities/ensuring-responsible-use-of-life-sciences-research

WHO. "Global excess deaths associated with COVID-19 (modelled estimates)." Data set, last updated May 19, 2023. https://www.who.int/data/sets/global-excess-deaths-associated-with-covid-19-modelled-estimates.

Xie, Xuping, Antonio Muruato, Kumari G. Lokugamage, et al. "An Infectious cDNA Clone of SARS-CoV-2." *Cell Host & Microbe* 27, no. 5 (2020): 841–848.e3. https://doi.org/10.1016/j.chom.2020.04.004.



GovAI