How likely is a nuclear exchange between the US and Russia?

Luisa Rodriguez | June 2019 | Rethink Priorities

Summary
Project Overview
Would the US and Russia intentionally start a nuclear war?
Accidental nuclear war
Putting a number on the probability of a US-Russia nuclear war
   Historical evidence
   Evidence from expert surveys
   Evidence from Good Judgment Inc. Superforecasts
So, how likely is a nuclear war between the US and Russia?

Corrections
Credits
Bibliography
Summary

My previous posts address how bad a nuclear war is likely to be, conditional on there being a nuclear war (see this post on the deaths caused directly by a US-Russia nuclear exchange, and this post on the deaths caused by a nuclear famine), but they don’t consider the likelihood that we actually see a US-Russia nuclear exchange unfold in the first place. In this post, I get a rough sense of how probable a nuclear war might be by looking at historical evidence, the views of experts, and predictions made by forecasters. I find that, if we aggregate those perspectives, there’s about a 1.1% chance of nuclear war each year, and that the chances of a nuclear war between the US and Russia, in particular, are around 0.38% per year.

<table>
<thead>
<tr>
<th>Event</th>
<th>Data Type</th>
<th>Annualized Probability (median, where relevant)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear weapons used during war</td>
<td>Based on historical frequency</td>
<td>1.4%</td>
<td>Calculation</td>
</tr>
<tr>
<td>Nuclear attack</td>
<td>Expert survey</td>
<td>2.21%</td>
<td>Lugar, 2005</td>
</tr>
<tr>
<td>Nuclear war kills at least 1 million</td>
<td>Expert survey</td>
<td>0.39%</td>
<td>Sandberg &amp; Bostrom, 2008</td>
</tr>
<tr>
<td>Nuclear detonation by a state actor causing at least 1 fatality</td>
<td>Superforecaster prediction</td>
<td>0.40%</td>
<td>Good Judgment Inc., n.d.</td>
</tr>
<tr>
<td>Aggregate</td>
<td>Simple average</td>
<td>1.1%</td>
<td>Calculation</td>
</tr>
</tbody>
</table>

Project Overview

This is the fifth post in Rethink Priorities’ series on nuclear risks. In the first post, I look into which plausible nuclear exchange scenarios should worry us most, ranking them based on
their potential to cause harm. In the second post, I explore the make-up and survivability of the US and Russian nuclear arsenals. In the third post, I estimate the number of people that would die as a direct result of a nuclear exchange between NATO states and Russia. In the fourth post, I estimate the severity of the nuclear famine we might expect to result from a NATO-Russia nuclear war. In this post, I get a rough sense of the probability of nuclear war by looking at historical evidence, the views of experts, and predictions made by forecasters. Future work will explore scenarios for India and Pakistan, scenarios for China, the contradictory research around nuclear winter, the impact of several nuclear arms control treaties, and the case for and against funding particular organizations working on reducing nuclear risks.

Would the US and Russia intentionally start a nuclear war?

While a nuclear war between the US and Russia by no means seems imminent, relations between the two are very precarious, mainly as a result of conflict in Syria, the annexation of Crimea, and allegations of Russian interference in US politics. President Putin has made it clear that he is pessimistic about the relationship between the US and Russia. In an interview earlier this week, he stated that “[relations with the US] are going downhill, they are getting worse and worse” — a sentiment that has no doubt been exacerbated by US announcement that it will deploy 1,000 US troops to Poland (i.e. right next to Russia) (Reuters, 2019 June 13; see for example, BBC, 2019). Were any of these conflicts to worsen significantly, the tensions could escalate to the point of war. Whether a war between the US and Russia would necessarily become nuclear is unclear. But it’s certainly a risk.

On the other hand, nuclear deterrence is a compelling force. At least for now, both the US and Russia maintain nuclear forces that are ‘survivable,’ meaning that both countries have nuclear weapons that are well-hidden enough that the other couldn’t destroy them all in a first strike. This means that, if either country decided to initiate a first strike, it would almost certainly face nuclear retaliation. And, while both the US and Russia seem more inclined to target each others’ nuclear forces directly rather than target each other’s cities and industry, neither country could be sure that the other wouldn’t target its citizens and economy. Deterrence theorists conclude that, as long as this is true, the two countries will remain in a Nash equilibrium — neither country being willing to risk the worst consequences of a nuclear war.¹ The Brookings Institution puts it this way: “no sane adversary would believe that any political or military advantage would be worth a significant risk of the destruction of his own society” (Wirtz, 2000).

Unfortunately, it’s unclear how sane — or rather, rational — the world leaders with the power to launch nuclear weapons are. It’s easy to imagine that a world leader with few checks on their power would launch a nuclear attack without sufficiently weighing the

¹ See for example The Economist, 2016
gravity of the risks. And some leaders have more leeway to launch a nuclear first strike against an adversary than others. For example, neither President Trump nor Kim Jong Un would have to get approval from anyone else to launch a nuclear weapon. According to the president of the Ploughshares Fund, “[t]he president can order a nuclear strike in about the time it takes to write a tweet.”

But it’s really hard to know whether seemingly trigger-happy leaders would go through with launching a first strike. Donald Trump has a reputation for exaggeration and intimidation (for example, see tweet below), and military advisors around him have stated publicly that they would strongly advise him not to launch a nuclear attack if his decision seemed irrational (for example, see Diaz 2017). There’s also a strategic advantage gained by over-stating the extent to which one would seriously consider a nuclear first strike. So we can’t exactly take world leaders at their word.

Overall, my intuition is that deterrence is effective — that the promise of mutually assured destruction is a strong enough disincentive that the US and Russia would not risk their economy and tens of millions of citizens’ lives. But I could very well be underestimating the potential for leaders to misjudge the risks to their societies and themselves — or their potential to not consider the risks at all.

But let’s say world leaders are sufficiently deterred by the prospect of mutually assured destruction, and wouldn’t be willing under almost any circumstances to start a nuclear exchange. Might they start one by accident?

**Accidental nuclear war**

Since nuclear weapons were developed, there have been a number of accidents, equipment malfunctions, and human mistakes that brought the world close to nuclear war. Some of

---

2 (Ward, 2018)
these close calls were closer than others.

For example, in 1961, a B-52 bomber carrying two nuclear warheads broke apart in mid-air over Goldsboro, North Carolina (BBC, 2013). The breakup of the plane triggered the mechanism (a simple pulley system) meant to be used by the pilot to drop the bomb over a target, which started the arming process for both bombs. One of the bombs landed safely on the ground, fail-safe intact, after the successful deployment of its parachute.

The other plummeted to the ground after its parachute failed and completed three of its four arming steps. Had it detonated, the bomb — which had over 250 times the explosive yield of the nuclear bomb dropped on Hiroshima — could have killed millions of people across the East Coast (Mosher, 2017; Paoletti, 2017).³

³ “If the bombs had detonated, lethal fallout could have been deposited over Washington, Baltimore, Philadelphia, and even as far north as New York City, putting millions of lives at risk” (Paoletti, 2017).
According to a now-declassified report on the event, now known as the Goldsboro incident, “one simple, dynamo-technology, low voltage switch stood between the United States and a major catastrophe” (The Guardian, 2013, p.1). The report later says that the “bomb did not possess adequate safety for the airborne alert role in the B-52” (The Guardian, 2013, p. 2). In another recently declassified report, Secretary of Defense McNamara is quoted as saying that only “by the slightest margin of chance, literally the failure of two wires to cross, a nuclear explosion was averted” (McNamara (n. d.), p.1).

Other incidents like this are similarly alarming. In 1960, early warning systems reported that long-range Russian missiles had been launched with “99.9% certainty” (Baum, de Neufville, & Barrett, 2018, 30). It turned out it was just the moonrise, which officers eventually figured out after no other warning systems were sounding alarms.

Two decades later, a technological glitch caused early warning systems to go off at the US Strategic Air Command, alerting officials of an incoming attack — mistakenly (Baum, de Neufville, & Barrett, 2018). Strategic bomber pilots prepared to take off from their air bases, but the alert was found to be a false alarm when no other warning systems detected incoming missiles.

A list of eighteen close calls caused by accidents like this is available in Baum, de Neufville, and Barrett, (2018).

Given this history of near-miss events, it’s not hard to imagine that human or technological errors could lead us to nuclear war.

It’s tempting to imagine that the risk of accidental nuclear war has decreased as the technologies that have failed in the past, like early warning radar and fail-safes have likely improved. I found some evidence of this, though not as much as I expected. For example, the US's early warning systems that detect incoming missiles (which have been partially responsible for near-miss events in the past) have been upgraded with technology that is better at classifying and tracking projectiles (Owens, 2017). This likely reduces the risk that the US or Russia would mis-identify meteorological events or non-nuclear projectiles as an impending nuclear attack.

But the expanded use of technology in nuclear weapons systems may introduce _new_ risks. Unal and Lewis warn that, because nuclear weapons now rely more heavily on digital technology, they’ve become more vulnerable to cyber attack (Unal & Lewis, 2018).

It’s hard to guess the net effect of these developments, in part because we can’t be confident that we know about all of the close calls that have happened — especially in the recent past (recent enough that the details haven’t been de-classified) (Lewis et al., 2014).4

---

4 “It is impossible to say whether the risk of near misses has increased over time. This is primarily because it is not possible to have a complete sense of the number of near misses and therefore to determine if and when there was a greater concentration of them. While the consequences of a nuclear detonation have remained relatively consistent, levels of risk probability are difficult to estimate. It took decades to learn about the role of misperceptions in the Cuban missile crisis, and there is no reason to assume that the full picture has been drawn” (Lewis et al., 2014, p. 30)
Putting a number on the probability of a US-Russia nuclear war

To get a rough sense of how probable we should find an intentional or accidental nuclear war, I looked at historical evidence, the views of experts, and predictions made by forecasters.

**Historical evidence**

We can establish a base rate for the probability of nuclear war by looking at the number of times nuclear weapons have been used during a war: one time since they were developed 74 years ago. This could be interpreted to mean that the likelihood of nuclear war is around 1.4% per year.

But there are several reasons not to put much stock in this probability (Baum, de Neufville, & Barrett, 2018). The most important reason, in my view, is the fact that the one time that nuclear weapons were used during a war — by the US, during World War II — the US was the only country who had nuclear weapons. Because of this, they weren’t deterred by the threat of nuclear retaliation. Today, because countries considering using nuclear weapons risk a nuclear second strike, the threshold for using nuclear weapons is likely higher than it was when the US used nuclear bombs during WWII, thereby decreasing the probability of nuclear war somewhat.

Baum et al. (2018) also point out that a number of other historical circumstances have changed: the political leaders, the number of countries with nuclear weapons, and the relationships between countries with nuclear weapons, among other things. And again, they note that the probability might be somewhat higher than the base rate would suggest, given that there have been a number of geopolitical crises that almost escalated to the point of nuclear war but didn’t (near misses in grey box in the figure below):
To take these close calls into account, researchers Barrett, Baum and Hostetler (2013) took an alternative approach to estimating the baseline probability of nuclear war using historical frequencies (though they focus only on the risk of accidental nuclear war between the US and Russia). Rather than looking only at the instances where nuclear war actually happened, they looked back at the frequency of accidents that nearly led to nuclear war. They argued that, if one assumes that those near misses could have also ended in disaster, one can make inferences about the probability of nuclear war that are more informative than just looking at the base rate of nuclear war.

Based on this reasoning, Barrett et al. (2013) concluded that the median annual probability of inadvertent nuclear war between the US and Russia is about 0.9% (90% CI: 0.02% — 7%).

But there’s controversy over how to interpret those close calls. Should we consider them evidence that a nuclear war could easily have happened multiple times since WWII — that we’re just lucky they didn’t? Or should we think of them as evidence that, while near-misses are relatively common, it’s actually really hard for a close call to escalate to the point of a nuclear exchange?

Source: Adapted from Baum, de Neufville, and Barrett (2018)
It’s difficult to know for sure, but I’m struck by the fact that a number of the close calls caused by accidents reported by Baum et al. (2018) have similar-sounding endings — usually something like “because there was no other evidence of an attack…” X agency “determined it was a false alarm caused by” X malfunction (Baum et al., 2018, p. 30). This could be interpreted to mean that, even though human and technological error may lead to more close calls that we’d hope, the systems in place to identify mistakes before they escalate might just work well enough to keep nuclear war from happening by accident. But this is only very weak evidence, and there’s a lot of reason to be uncertain.

Given this uncertainty, it’s useful to look to other forms of evidence for how likely a US-Russia nuclear exchange — intentional or otherwise — might be.

**Evidence from expert surveys**

The 2008 Global Catastrophic Risk (GCR) survey asked experts to make predictions about nuclear war scenarios through to 2100. Specifically, the experts were asked to estimate the likelihood that nuclear wars kill a) at least one million people, b) at least one billion people, and c) enough people that humans become extinct (Sandberg & Bostrom, 2008).

<table>
<thead>
<tr>
<th>2008 Global Catastrophic Risk survey</th>
<th>At least 1 million dead</th>
<th>At least 1 billion dead</th>
<th>Human extinction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total killed in all nuclear wars by 2100 (median)</td>
<td>30%</td>
<td>10%</td>
<td>1%</td>
</tr>
</tbody>
</table>

From this we can glean that experts see the probability of nuclear war killing at least 1 million or 1 billion people by 2100 as reasonably small, but not insignificant — about 0.39% and 0.12% per year, respectively. They see the risk of extinction caused by nuclear war as much smaller, but again, it’s not insignificant at about 0.011% per year.

Respondents weren’t asked to consider the probability of specific nuclear exchange scenarios, so these figures can inform our understanding of the likelihood of a nuclear exchange between the US and Russia.

Another expert survey, the Lugar Survey On Proliferation Threats and Responses, asked experts all over the world to estimate the probability of nuclear attack, but over a shorter time span (Lugar, 2005).

The median view of experts estimating the probability of a nuclear attack within 5 years (from 2004-2009) was 10%, or 2.09% per year, and 20% over 10 years (from 2004-2014), or 2.21% per year.

---

5 Note, to estimate this annualized probability (and all other annualized probabilities in this post), I assume that each the probability of nuclear war happening in a given year is independent of the probability of nuclear happening in any other year, which isn’t actually the case.
Like the GCR survey, the Lugar Survey didn’t ask experts to consider specific geopolitical scenarios, so again, we can only learn a limited amount about a US-Russia exchange. Additionally, it’s worth noting that the five and ten-year time horizons reflected in these predictions have already passed and resolved in the negative.

I did identify one survey that asks explicitly about the probability of a nuclear exchange between the NATO states and Russia. The Project for the Study of the 21st Century (PS21) Great Power Conflict Report, released for publication in 2015, asked 50 national security experts from all over the world to estimate the probability of a variety of conflict scenarios that could plausibly occur in the next 20 years (Apps, 2015). The median view of the experts was that the probability of a nuclear exchange between the US and Russia in the next 20 years is around 4.72%, or 0.24% per year.

<table>
<thead>
<tr>
<th>PS21 Great Power Conflict Report</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>What would you estimate the probability of conflict between NATO States and Russia over the next 20 years (conventional, with or without nuclear exchange)</td>
<td>20%</td>
<td>22.64%</td>
</tr>
<tr>
<td>What would you estimate the probability of conflict between NATO States and Russia over the next 20 years (nuclear exchange)</td>
<td>4%</td>
<td>4.72%</td>
</tr>
</tbody>
</table>

**Evidence from Good Judgment Inc. Superforecasts**

Finally, I consider evidence from Good Judgment Inc. (GJI). Good Judgment Inc. uses insights learned from the Good Judgment Project, a research initiative that worked to understand how to make accurate predictions about the future by holding forecasting tournaments. During these tournaments, forecasters were asked to make predictions about the future, and GJP researchers worked to understand what made some people perform better than others.

In 2018, GJI had its superforecasters make a set of predictions about the probability of a nuclear exchange by the year 2021 (unpublished GJI data from Open Philanthropy Project).

---

6 I believe the Project for the Study of the 21st Century (PS21) Great Power Conflict Report has several typos (Apps, 2015). I present what I believe to be the correct values (and the values I use in my analysis) here.

7 Another typo in the PS21 Great Power Conflict Report makes it unclear whether the probabilities are over a 20-year time span or a 25-year time span. I assume the survey asked about the time-span noted in a table, not in the discussion of the results.
The GJI forecasts are likely to be less susceptible to the biases that lead experts to make somewhat worse predictions (AI Impacts, 2019).

The superforecasters made two relevant predictions:

1. The forecasters predicted that there is a 1% chance of a nuclear attack by a state actor causing at least one fatality before 1 January 2021 — equivalent to about 0.40% per year.
2. They then predicted that, assuming (1) proves true, there is a 3% chance that a nuclear attack by a state actor causes at least one fatality in Russia during the same time span.

<table>
<thead>
<tr>
<th>Good Judgment Inc 2018 forecasts</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) What is the probability that a nuclear attack by a state actor causes at least one fatality before 1 January 2021?</td>
<td>1%</td>
</tr>
<tr>
<td>(2) If question (1) results in the affirmative, what is the probability that a nuclear attack in Russia by a state actor causes at least one fatality before 1 January 2021?</td>
<td>3%</td>
</tr>
</tbody>
</table>

If most of the probability of a nuclear attack on Russia comes from a nuclear exchange between the US and Russia, then we can approximate the probability of a US-Russia nuclear exchange by calculating the probability that there is 1) a nuclear attack by a state actor and 2) a nuclear attack by a state actor in Russia, which is 0.03%, or 0.01% per year (unpublished GJI data from Open Philanthropy Project; Apps, 2015).

It should be noted that the GJI predictions were made with a very short time span and a specific set of political circumstances in mind; they can only really tell us about the probability of a US-Russia nuclear exchange before 2021.

So, how likely is a nuclear war between the US and Russia?

If we aggregate historical evidence, the views of experts and predictions made by forecasters, we can start to get a rough picture of how probable a nuclear war might be.⁸ We shouldn’t put too much weight on these estimates, as each of the data points feeding into those estimates come with serious limitations. But based on the evidence presented above, we might think that there’s about a 1.1% chance of nuclear war each year and that the chances of a US-Russia nuclear war may be in the ballpark of 0.38% per year (but note that

---

⁸ A better way to aggregate these figures would be to use formal Bayesian updating. Because of time constraints, I just take the simple average instead, which can still be somewhat informative.
these figures are deceptively precise).

<table>
<thead>
<tr>
<th>Event</th>
<th>Data Type</th>
<th>Annualized Probability (median, where relevant)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Probability of nuclear war</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear weapons used during war</td>
<td>Based on historical frequency</td>
<td>1.4%</td>
<td>Calculation</td>
</tr>
<tr>
<td>Nuclear attack</td>
<td>Expert survey</td>
<td>2.21%</td>
<td>Lugar, 2005[9]</td>
</tr>
<tr>
<td>Nuclear war kills at least 1 million[10]</td>
<td>Expert survey</td>
<td>0.39%</td>
<td>Sandberg &amp; Bostrom, 2008[13]</td>
</tr>
<tr>
<td>Nuclear detonation by a state actor causing at least 1 fatality</td>
<td>Superforecaster prediction</td>
<td>0.40%</td>
<td>Good Judgment Inc., n.d.</td>
</tr>
<tr>
<td>Aggregate</td>
<td>Simple average</td>
<td>1.1%</td>
<td>Calculation</td>
</tr>
<tr>
<td><strong>Probability of nuclear war between the US and Russia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accidental nuclear war between the US and Russia[13]</td>
<td>Fault tree analysis based on historical frequency</td>
<td>0.9%</td>
<td>Barrett et al., 2013[13]</td>
</tr>
<tr>
<td>US-Russia nuclear war</td>
<td>Expert survey</td>
<td>0.24%</td>
<td>Apps, 2015[14]</td>
</tr>
<tr>
<td>Nuclear detonation in Russia causing at least one fatality[15]</td>
<td>Superforecaster predictions</td>
<td>0.01%</td>
<td>Good Judgment Inc, n.d.</td>
</tr>
<tr>
<td>Aggregate</td>
<td>Simple average</td>
<td>0.38%</td>
<td>Calculation</td>
</tr>
</tbody>
</table>
(See notes on sources\textsuperscript{9, 10, 11, 12} and use of estimates\textsuperscript{13, 14, 15}).

**Corrections**

**June 12, 2019** — I included an incorrect table when I meant to present a table summarizing the results of the Lugar Survey On Proliferation Threats and Responses. I've replaced that table with the correct one.

**June 12, 2019** — I incorrectly attributed forecasts to Good Judgment Project superforecasters when they were actually made by Good Judgment Inc.

**June 12, 2019** — I originally reported that the average Good Judgment Inc. superforecaster prediction in answer to the question "What is the probability that a nuclear attack by a state actor causes at least one fatality before 1 January 2021" was 2%. When I re-calculated this figure independently using dis-aggregated data provided by Open Philanthropy Project, I found that the average GJI prediction as 1%. Accounting for this new estimate causes the average annualized probability of nuclear war to fall from 1.17% to 1.1%, and the annualized probability of a US-Russia nuclear war to fall from 0.39% to 0.38%.


\textsuperscript{10} See Sandberg & Bostrom (2008)

\textsuperscript{11} See Barrett et al. (2013)

\textsuperscript{12} See Apps (2015)

\textsuperscript{13} Because nuclear wars are presumably very unlikely to cause <1M fatalities, I assume this is equivalent to predicting the probability of any nuclear war.

\textsuperscript{14} This estimate is not directly comparable to the other estimates of the probability of a nuclear war between the US and Russia. While the others theoretically take into account the probability of accidental nuclear war and an intentional nuclear war, Barrett et al.'s 2013 estimate only consider accidental nuclear war. I include it in spite of this because I expect most of the probability of a US-Russia nuclear war comes from the potential for an accidental nuclear exchange. This makes me think that Barrett et al.'s 2013 estimate is not so different from an estimate of the probability of any US-Russia nuclear exchange.

\textsuperscript{15} Again, this assumes that the majority of the probability of a nuclear attack by a state actor causing at least one fatality in Russia comes from a nuclear exchange involving the US.
Credits

This essay is a project of Rethink Priorities. It was written by Luisa Rodriguez with contributions from Ida Sprengers. Thanks to Peter Hurford, Marinella Capriati, and Marcus A. Davis who provided valuable comments. Thanks also to Matt Gentzel and Seth Baum for providing guidance and feedback on the larger project, and to Carl Schulman for supplying some of the data reported in this post.

If you like our work, please consider subscribing to our newsletter. You can see all our work to date here.

Bibliography


80,000 Hours. (2018, December 27). We could feed all 8 billion people through a nuclear winter. Dr David Denkenberger is working to make it practical [Audio blog interview]. Retrieved from https://80000hours.org/podcast/episodes/david-dankenberger-allfed-and-feeding-everyone-no-matter-what/


