

How many people would be killed as a direct result of a US-Russia nuclear exchange?

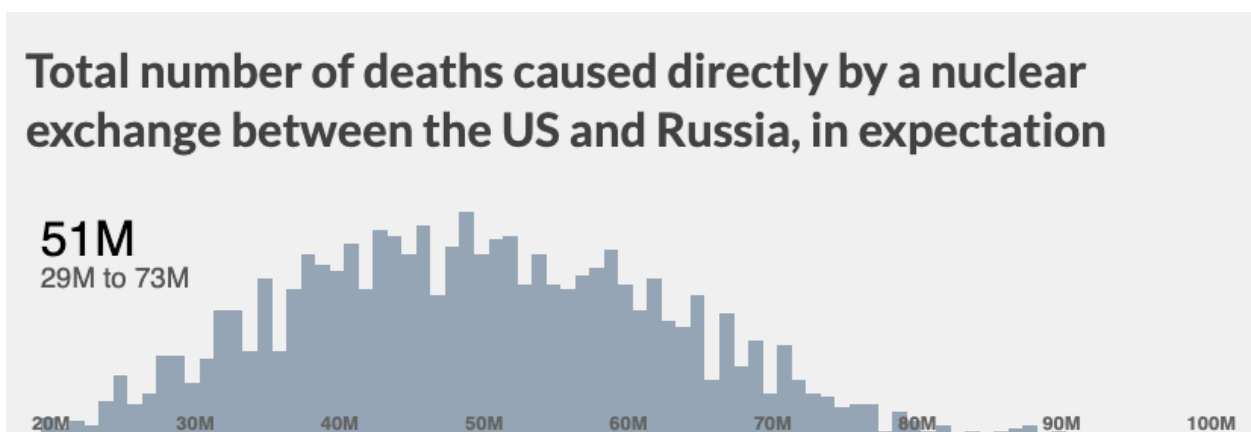
Summary

In this post, I estimate the number of fatalities caused directly by nuclear detonations in the US/NATO and Russia. I [model these effects in Guesstimate](#) using expert surveys and interviews, forecasts made by Good Judgment Project superforecasters, academic research, and media coverage of international relations, along with academic research into the effects of nuclear war and nuclear weapons policy.

There are many determinants that factor into the number of people that would die as a direct result of nuclear detonations during a US-Russia nuclear exchange. I consider the following six factors the most important. They make up the key parameters in my model:

1. The targeting strategy (i.e. what kinds of targets will each country attack?)
2. The number of military facilities each country might target
3. Whether each country would also target cities, in addition to military facilities
4. If they were to target cities, the number of cities each country might target
5. The sizes of the nuclear weapons in each country's nuclear arsenal
6. The population size of the cities that might be targeted during an exchange

When I take all of these factors into account, I expect that we'd see a total of 33 million deaths caused directly by nuclear detonations on military and civilian targets in NATO countries and Russia (90% confidence interval: 3022 million — 7548 million deaths).



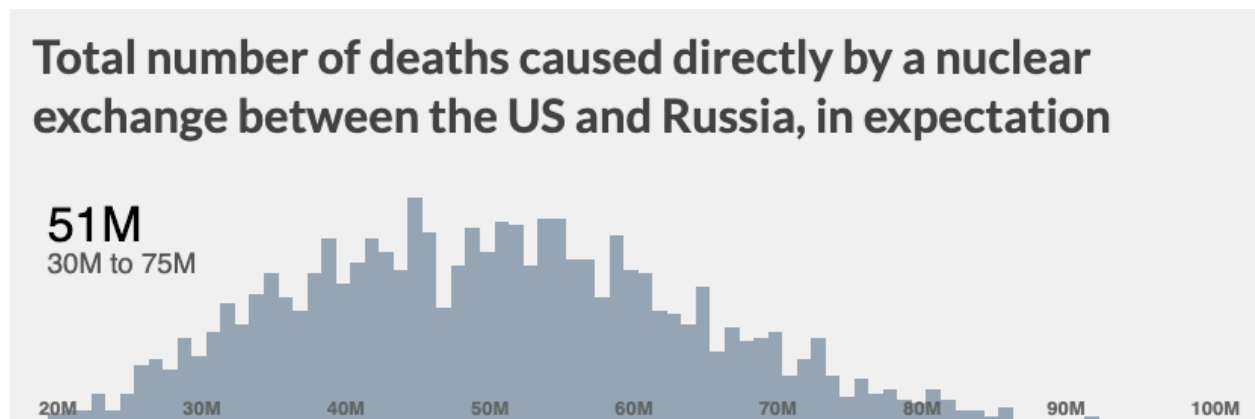
Date XXXXX 2019 Update

In light of feedback from Carl Schulman, Kit Harris, MichaelA, David Denkenberger, Topher Brennan, and others, I've made several revisions to this post that are now reflected in the text, figures, and estimates in the body of this post. The original post can still be found [here](#).

The changes that had the largest bearing on my results included:

- Changing the way I estimate the number of nuclear weapons that would be used in a countervalue nuclear exchange in expectation so that I don't accidentally truncate the tails of the distributions (details [here](#) and [here](#)).
- Generating a formula that can be directly entered into Guesstimate to estimate the number of deaths caused by a countervalue nuclear exchange rather than using a simplified formula to estimate the parameters for triangular distributions that are then entered into Guesstimate (details [here](#) and [here](#)).

After making these revisions, my estimate of the number of people that would be killed directly by nuclear detonations during a US-Russia nuclear exchange is about 51 million (90% confidence interval: 30 million — 75 million deaths) — ~43% more than my original estimate of 35 million (90% confidence interval: 23 million — 50 million deaths).



The impacts that each individual change had on my results can be seen [here](#).

I've also added a bit more discussion on the probability that a countervalue nuclear exchange would escalate, and sensitivity analysis so that people who disagree with my views on this can see how the results change under more pessimistic assumptions. My sensitivity analysis shows that, if you're more pessimistic than me about the probability of countervalue targeting and escalation, around 88 million people would be killed in expectation during a US-Russia nuclear exchange (details [here](#) and [here](#)).

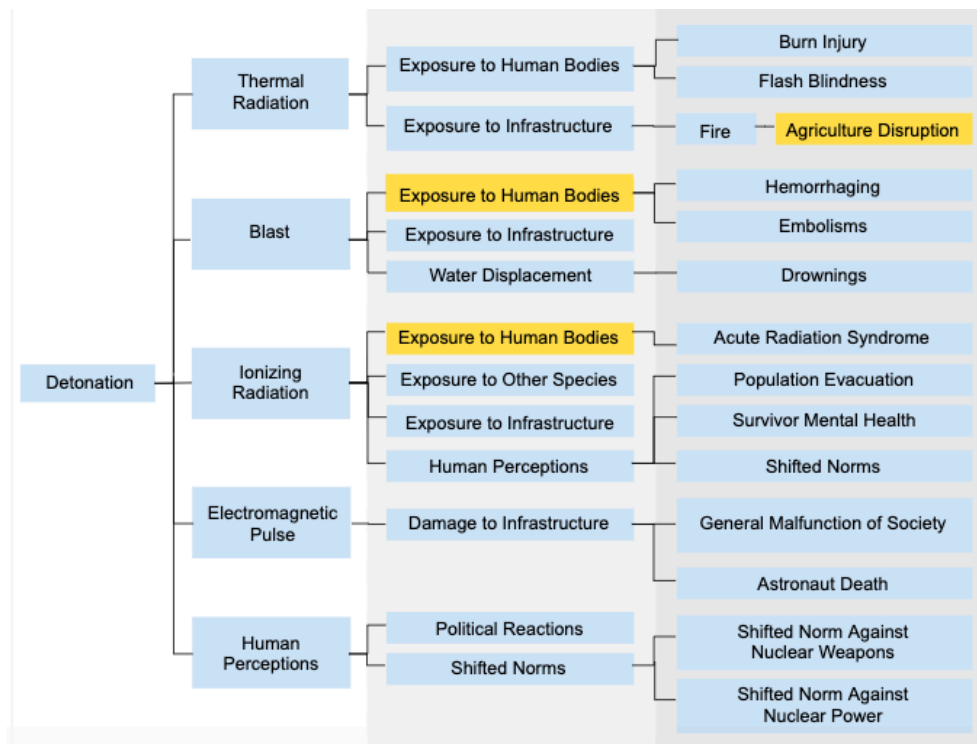
Thanks again to those who offered feedback, and also to Jaime Sevilla, Ozzie Gooen, Max Daniel, and Marinella Capriati for feedback and technical support implementing the revisions.

Project Overview

This is the third 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 this 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 the fifth 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. In the sixth and seventh posts, I estimate the direct and indirect effects of nuclear exchanges between (1) India and Pakistan and (2) China and its adversaries. Future work, to be published later in the summer, will explore 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.

Modeling the direct impacts of a nuclear war between the US and Russia i

In their paper on the impacts of nuclear war, Baum and Barrett ([2018](#)), researchers at the Global Catastrophic Risk Institute (GCRI) lay out the range of possible impacts that can result from a nuclear detonation. The impacts are wide-ranging and enormous in scale.



Source: Adapted from [Baum and Barrett \(2018\)](#)

Estimating the harm caused through *all* of the pathways described by Baum and Barrett (2018) is beyond the scope of this project. Instead, my focus in this post is to estimate the number of fatalities caused by the blast and radiation effects, as these would cause the majority of deaths caused directly by nuclear detonations (Toon et al., 2007). In a subsequent post, I'll also estimate the number of deaths caused by the famine that might result from the agricultural disruption that would likely follow a US-Russia nuclear exchange. In addition, I further limit my analysis to just human suffering, though many domesticated and wild animals would be affected as well.

I model these effects in Guesstimate using data from expert surveys and interviews, forecasts made by Good Judgment Project superforecasters, academic research and media coverage of international relations, and academic research into the effects and probability of nuclear war and nuclear weapons policy.

In the remainder of this post, I describe [my Guesstimate model of the direct effects of a US-Russia nuclear exchange](#), laying out the reasoning, data, and assumptions that inform the model's main parameters. I then summarize the model's results and itemize my key uncertainties.

The targeting strategy used by the US and Russia

There are two general approaches to deciding which targets to hit during a nuclear exchange: *countervalue* targeting and *counterforce* targeting, or some combination of both.

In [countervalue targeting](#), the primary goal is to kill civilians and damage economic infrastructure. As such, the countervalue targeting approach prioritizes large cities and industrial centers — targets that a country inherently *values*. Countervalue targeting is considered easy and cheap. Cities and factories are really hard to protect, easy to identify, and stationary, meaning the technical requirements of countervalue targeting are few, as are the number of nuclear weapons needed to be successful ([Doctrine for Joint Nuclear Operations, 1995, p. II-5](#)).¹ Because of this, countries interested in maintaining an effective nuclear deterrent that is minimally expensive commonly adopt what's called a Minimum Countervalue Deterrent (MCD) posture. Countries like China and North Korea have such deterrence policies, meaning that their deterrence rests explicitly on the idea that, if attacked, they would retaliate by targeting their enemy's biggest cities and industries, with the explicit goal of creating suffering and chaos.

The second strategy, [counterforce targeting](#), centers on damaging the enemy's nuclear forces to prevent a retaliatory nuclear second strike, removing the enemy's ability to engage in further counterforce or countervalue targeting. Under counterforce targeting, the prioritized targets would likely include, for example, nuclear weapons storage facilities, defense installations, and bomber bases. But counterforce targeting requires a huge and technologically sophisticated arsenal, which is expensive.² So while counterforce targeting is strategically superior to countervalue targeting, it is a costly approach to take.

During the Cold War, when much of the research into nuclear winter was being done, it wasn't possible to do counterforce targeting very well ([Lieber & Press, 2017](#)). Nuclear weapons just weren't sophisticated enough. But improvements in technology have made counterforce targeting feasible. For example, it used to be the case that a US nuclear missile fired from a submarine only had a 9% chance of destroying a nuclear silo. Now, it would have a 90% chance.

¹ "In general, weapons required to implement this strategy need not be as numerous or accurate as those required to implement a counterforce targeting strategy, because countervalue targets generally tend to be softer and unprotected in relation to counterforce targets." ([Doctrine for Joint Nuclear Operations, 1995, p. II-5](#))

² According to the 1995 Chairman of the Joint Chiefs of Staff, counterforce targeting is much more expensive and technically difficult than countervalue targeting for several reasons ([Doctrine for Joint Nuclear Operations, 1995, p. II-5](#)). First, the weapons systems have to be extremely accurate to be able to target specific missile silos, for example. By contrast, detonating a nuclear bomb somewhere in a large city takes a lot less accuracy. In addition, there may be as many as hundreds of thousands of attractive targets under counterforce targeting, meaning that you need many more nuclear weapons than would be necessary to target the enemy's ten most populated cities, for example. Further, unlike cities and industrial areas, counterforce targets are easier to protect, hide, and even move around. This again means that counterforce targeting requires many sophisticated nuclear weapons, but also means that the intelligence informing the targeting decisions has to be really high-quality and constantly updated.

My impression is that, once counterforce targeting became possible, US and Russian nuclear policy seemed to transition away from countervalue targeting strategies and toward counterforce targeting strategies.

As of 2010, the US says explicitly in its nuclear policy that it *won't* engage in countervalue targeting.³ Moreover, the US's nuclear forces *look* like the kind of forces you'd expect to see in a country that planned to implement counterforce targeting ([Kristensen & Norris, 2018](#)). The US arsenal is large, accurate, flexible, and relatively low-yield — all characteristics associated with counterforce targeting. Compare this to China's nuclear arsenal, which is small and crude — too low-tech for effective counterforce targeting, but more than sufficient to achieve Minimum Countervalue Deterrence. Given that China's posture can be achieved much more cheaply, I take this as *some* evidence that the US expects to rely on counterforce targeting in the event of a nuclear exchange.

As far as I can tell, Russia doesn't explicitly state that it would use a counterforce targeting strategy during a nuclear exchange. But earlier this year, Vladimir Putin talked openly about some of the locations Russia would target during a nuclear strike ([Mizokami, 2019](#)). While some of the targets seem to be relying on out-of-date intelligence about US military sites, all of the targets he listed were, at least at one point, of clear military significance.

And, like the US, Russia's nuclear arsenal is large and sophisticated (and expensive), characteristics which, again, are indicative of counterforce targeting ([Kristensen & Korda, 2019](#)).

The shift away from countervalue and toward counterforce targeting also makes strategic sense for the US for a bunch of reasons:

1. Countervalue targeting on the part of the US would risk countervalue retaliation by Russia. In other words, if the US decided to target Russian cities, it would be risking the lives of tens or hundreds of millions of its citizens *and* the destruction of its economy ([Wirtz, 2000](#)).
2. Counterforce targeting would better achieve the goal of winning a nuclear war. While a large portion of Russia's nuclear forces [would almost certainly survive a first strike](#), over time, concealed and relocatable nuclear weapons systems like bombers and even submarines would have to reveal themselves — for example, by landing to refuel. The US would likely target key sites, like airstrips and naval bases, and would probably be able to take out a large

³ In 2010, President Obama, updated the US's nuclear posture — only the third update made to US nuclear posture since the end of the Cold War. According to the Nuclear Employment Strategy ([2013](#)), a document summarizing the policy changes produced by the Department of Defense, the new guidance from President Obama “requires the United States to maintain significant counterforce capabilities against potential adversaries... The new guidance does not rely on a ‘counter-value’ or ‘minimum deterrence’ strategy.” The same report later says, point blank, “the United States will not intentionally target civilian populations or civilian objects.” ([Department of Defense, 2013, p.4](#)).

enough portion of Russia's nuclear forces to secure a military defeat eventually.

3. Irrespective of whether counterforce targeting would ensure a military defeat, counterforce targeting also better achieves the goal of minimizing civilian casualties. By eliminating some of Russia's nuclear arsenal, counterforce targeting would limit Russia's ability to mount a full-scale countervalue attack against the US.
4. Targeting cities just to kill civilians during a nuclear exchange is extremely and plainly illegal.⁴ International humanitarian law, which both the US and Russia try to at least *pretend* to abide by, says very clearly: states can't intentionally target civilians just to target civilians ([Wirtz, 2000](#); [ICRC a](#)).⁵ It also says that, if a military operation is going to result in civilian deaths, the number of deaths should be proportionate to the value of the military operation ([ICRC b](#)). The US certainly doesn't always follow international humanitarian law, but the fact that countervalue targeting would be so plainly and visibly in violation of the law may increase the barrier to pursuing it ([McCoy, 2015](#)).
5. Beyond the fact that counterforce is more aligned with humanitarian law, proponents also argue that, by emphasizing the destruction of military targets and minimizing civilian casualties, it is inherently more moral than a countervalue targeting strategy like Minimum Countervalue Deterrence ([Rudolf, 2018](#)).
6. Although there are lots of things left ambiguous in the US nuclear doctrine — in part because it strengthens deterrence, and in part because the US wants to keep certain options open — it might be politically costly for the US to break its own policies. The US has previously broken with precedent,⁶ so this doesn't entirely exclude the possibility of countervalue targeting. However, the fact that the US has stated quite clearly that it won't intentionally target cities might mean the cost of doing so is a bit higher than it otherwise would be.
7. Similarly, a countervalue strike would be generally bad for the US's credibility and reputation. The use of nuclear weapons — especially to destroy cities — is increasingly stigmatized by the international community ([Tannenwald, 1999](#)). The US cares at least a bit about preserving its international image as liberal and humanitarian, and it goes to great lengths to maintain a reputation of ethical superiority ([Wendt, 1992](#)). Dropping nuclear bombs on Russian cities would compromise that ([Tannenwald, 1999](#)). While this certainly wouldn't rule out the possibility of countervalue targeting, it probably raises the threshold for it.

⁴ Counterforce targeting is not necessarily in conflict with humanitarian law ([Rudolf, 2018, p. 10](#)). While counterforce targeting would cause civilian deaths, those deaths would not be the central aim of the targeting strategy, and would arguably be justified when compared to the benefit of eliminating an enemy's nuclear arsenal.

⁵ "Nevertheless, a countervalue doctrine is clearly at odds with established international humanitarian law, which requires that threats or uses of force must never be directed at civilians" ([Wirtz, 2000, p. 5](#))

⁶ See for example Watson, K ([2019](#)); Mehta, A. ([2017](#)); Cooper, H. ([2017](#))

Similarly, there are a number of reasons to think that counterforce targeting is strategically preferable from Russia's perspective as well:

1. Targeting US cities and industry would risk reciprocation by the US, threatening millions of Russian deaths. Even though the US has *said* it wouldn't target civilians, Russia would never be able to fully rule out the risk of countervalue retaliation that could destroy its society ([Wirtz, 2000](#)).
2. I've been told by experts that it is generally accepted that US nuclear forces would outmatch Russia's in a countervalue exchange. This makes me think that Russia wouldn't want to deliberately invite countervalue targeting early on by mounting an unprovoked countervalue strike.
3. Counterforce minimizes the potential for civilian casualties by limiting the number of surviving warheads that could then be pointed at Russian cities.
4. Again, countervalue targeting is illegal. While Russia has demonstrated its willingness to engage in illegal practices during war — for example, using landmines after they were banned, developing biological weapons in secret during the Cold War, probably cyberattacks — their denial of these practices demonstrate that they don't want the world to *think* that they do those things. The use of landmines, research into bio-weapons, and cyberattacks can all be plausibly denied, but detonating nuclear bombs in US cities can't be.

Together, these considerations make me think that the US and Russia have shifted away from countervalue targeting, and for good reason. Countervalue targeting is just too costly and risky, and there's a superior alternative in counterforce targeting.

But while I think it's exceedingly unlikely that the US or Russia would use countervalue targeting during a first strike, the chances would go up as a nuclear exchange continued (I'll discuss these circumstances a bit later). Given that, my model of the impacts of a nuclear exchange between the US and Russia takes into account the possibility of some amount of countervalue targeting by each country in addition to extensive counterforce targeting.

I start by looking at the impacts of counterforce targeting: how many targets would be attacked, with what kinds of nuclear weapons, and how many people would die in each country as a direct result?

Counterforce targeting by the US against Russia

A report by the Natural Resources Defense Council (NRDC) analyzed the impact of a US counterforce attack against Russia ([McKinzie et al., 2001](#)). The specific scenario they considered was based on the

counterforce targeting strategy known as Major Attack Option-1 (MAO-1) in the 1999 Single Integrated Operational Plan (SIOP-99) — the most recent nuclear targeting strategy document whose details have been ascertained by researchers. While the SIOP was replaced in 2003, I make the assumption that the details of the attack plan are similar enough to the type of targeting that might be executed today (I justify this below).

The NRDC researchers concluded 1,289 nuclear warheads — 500 W87 warheads, 220 W88 warheads, and 569 W76 warheads — would be used to detonate a combination of silo-based inter-continental ballistic missiles (ICBMs), road-mobile ICBMs, rail-mobile ICBMs,⁷ submarine-launched ballistic missile (SLBM) bases and facilities, long-range bomber bases and facilities, nuclear weapon storage sites, nuclear weapon design and production sites, and command/control/communications (C3) targets ([McKinzie et al., 2001](#)).⁸ While the size and make-up of the US nuclear arsenal has changed a bit since 2001 when this report was published, their MAO-1-like strike could be implemented today with just a few minor substitutions (they would just have to replace 300x 300-kiloton (kt) W87s with 300x 335kt W78s) ([Kristensen & Norris, 2018](#)).

They predicted that their MAO-1-like attack, which would take about 30 minutes from start to finish, would destroy 90% of Russia's silo-based ICBMs, destroy the garrisons where at least some of Russia's road-mobile ICBMs may be sheltered, destroy the major airfields and naval bases (limiting the places where bombers and subs can refuel), eliminate the nuclear weapons production complex, and majorly damaging country-wide communications systems ([McKinzie et al., 2001, p. 110](#)).

The human cost of this attack would be between 8 and 12 million lives ([McKinzie et al., 2001, p. 112](#)).⁹

⁷ As of 2005, Russia no longer uses rail-mobile missiles ([Starchak, 2017](#)).

⁸ While the specific targets selected for the four MAOs are still classified, the NRDC has been able to approximate the characteristics and effects of MAO-1 using target data from the USSTRATCOM National Target Base (NTB), academic literature, meteorological data, demographic data, and de-classified Department of Defense studies on the effects of nuclear weapons ([McKinzie et al., 2001](#)). This data was fed into a geographic information system (GIS) software, which was then able to generate 1) the optimal type, number, and delivery system for each nuclear weapon to maximize damage to the targets selected in their version of MAO-1, and 2) the number of casualties and fatalities associated with the attack.

⁹ A range is given because unknowable factors — for example, annual variation in wind patterns and the presence and use of residential nuclear sheltering — would impact the exact number of fatalities. ([McKinzie et al., 2001, p. 112](#)).

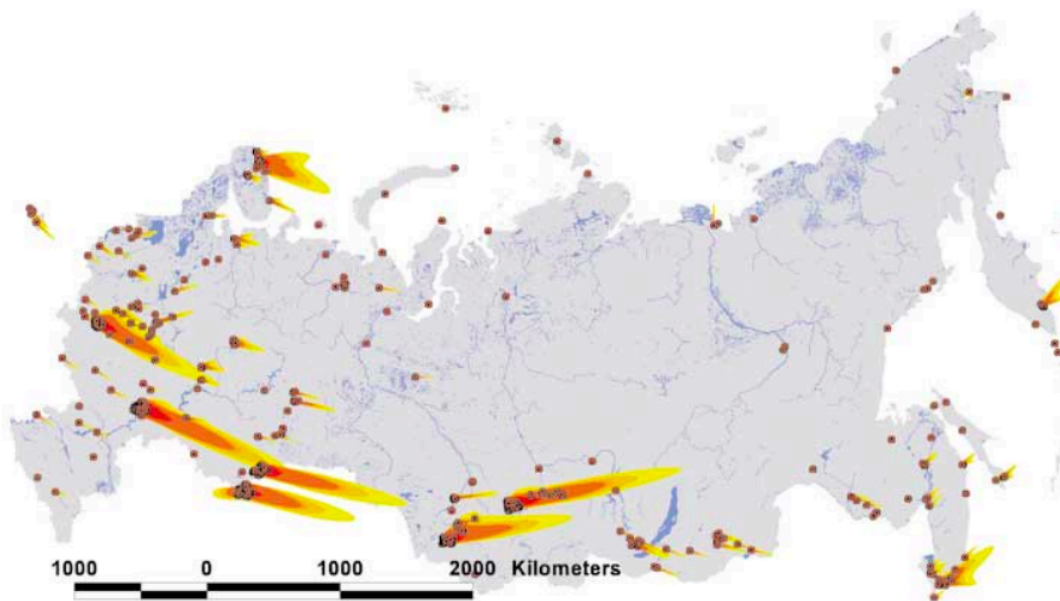


Figure 4.84: Fallout Patterns from the MAO-1 Across the Russian Landmass ([McKinzie et al., 2001, p. 111](#))

I think the MAO-1-like scenario modelled by the NRDC researchers is a very plausible first strike scenario. That said, it's possible that a first strike by the US might actually be more drastic or more limited in scale. The SIOPs included options for even larger-scale attacks (MAO-2,-3, and -4 get increasingly extensive), but also more limited nuclear attack plans.

I quantify this uncertainty by building a probability distribution that reflects the relative probability that counterforce targeting against Russia would be of varying scales. To do this, I used the probability distribution elicitation tool, [SHELF](#), which takes in a few “known” values in a distribution, and generates the parameters of the probability distribution that best fits the inputted values.

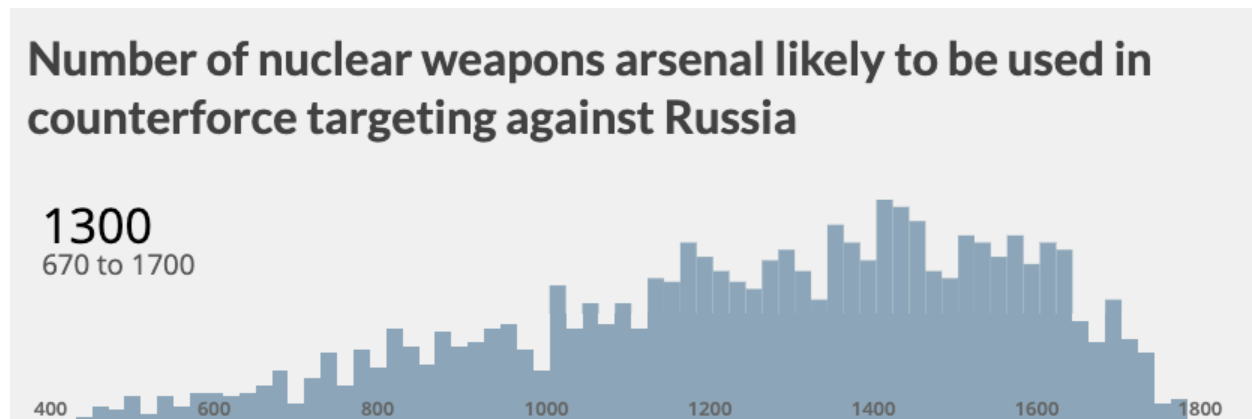
I build the distribution based on the incomplete information I have on plausible counterforce targeting scenarios, and my interpretation of that information. First, I assume that the maximally extensive counterforce targeting plan would involve, at most 1,800 nuclear warheads, as this is the number of nuclear warheads the US currently has deployed ([Kristensen & Norris, 2018](#)).

As far as I can tell, we don't know much about the more limited options named in past SIOPs (Limited Nuclear Options, Regional Nuclear Options, Directed Planning Options, and Adaptive Planning Options). To account for this uncertainty, I make the probability distribution very wide, by I assuming that a counterforce attack could plausibly involve very few nuclear weapons.

The goal of counterforce targeting is to reduce the risk and/or severity of retaliation. Given that, it seems *possible* that the US might execute a more limited first strike, but it seems most likely that a counterforce first strike against Russia would involve as many warheads as necessary to have a

meaningful impact on Russian nuclear forces. Based on this reasoning, I expect that most of the probability will be on the higher end of the plausible range — close to 1,289, based on the NRDC’s analysis of the number of warheads necessary to inflict a lot of damage (more detail in the footnote).¹⁰

The probability distribution I settled on looks like this:



To estimate the number of deaths that would be caused by an attack like this, I assume that the number of deaths is proportionate to the number of nuclear warheads detonated.¹¹ This assumption

¹⁰More details on this distribution: I wanted to generate a beta distribution to reflect the proportion of the maximum number of nuclear weapons that could plausibly be used in counterforce targeting that would actually be used in counterforce targeting against Russia. It could also be thought of as the probability that the US uses the maximum number of nuclear weapons in its counterforce targeting.

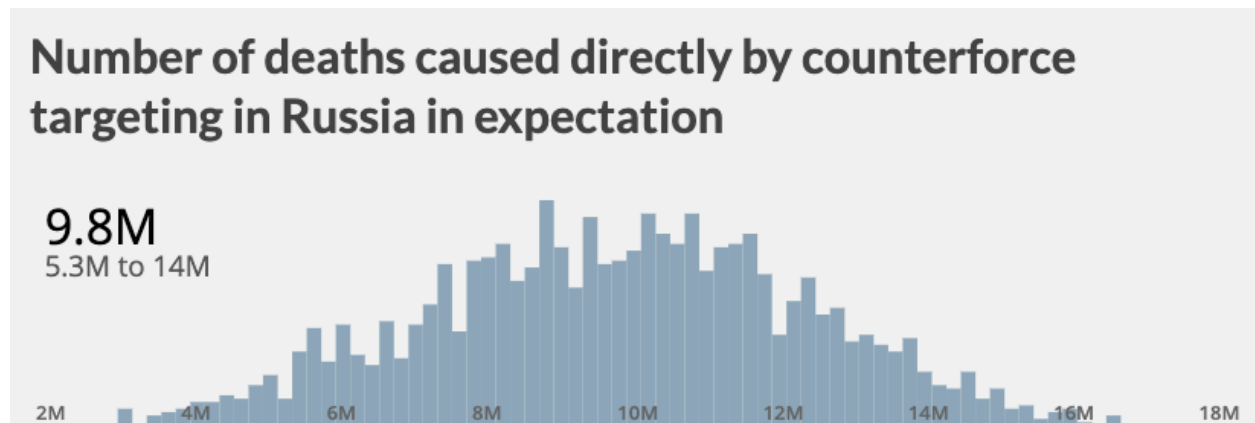
To generate the parameters of the beta distribution I think best fits what we know about US counterforce targeting, I enter what I believe to be the median value, the lower bound, the upper bound, the 0.05th percentile, and the 0.95th percentile. With help from [SHELE](#), I generated the beta distribution, $\text{beta}(12.8, 3.2)$, which approximates the values I’ve entered.

The benefit of the beta distribution is that, because non-probability values are standardized between 0 and 1, we can multiply the distribution by the maximum number of nuclear weapons to get the probability distribution of the number of nuclear weapons used. We can also multiply it by the maximum number of deaths caused by counterforce targeting to get a probability distribution for the number of deaths caused by counterforce deaths at the expected counterforce strike scales.

¹¹ In countervalue targeting, the assumption of linearity doesn’t make much sense because the first nuclear weapons used will be aimed at cities with extremely large populations, as these would be the highest value targets. Additional warheads would be used on increasingly small (lower-priority) cities, meaning that the relationship between the number of nuclear warheads detonated and the number of deaths is non-linear. Rather, number of deaths declines super-linearly with each additional nuclear warhead. In a counterforce strike, we should expect the population living around high-priority military targets and lower-priority military targets to be approximately the same on average. This means that each subsequent nuclear detonation in a counterforce strike should kill, on average, the same number of people. The assumption of linearity is therefore pretty reasonable.

allows me to use the same distribution shape to estimate the probability distribution representing the number of deaths that would be caused by the counterforce strike against Russia.¹²

I find that the probability distribution of the number of deaths likely to result from counterforce targeting in Russia during a US first strike looks like this:



According to my estimates, somewhere between 5.3 million and 14 million people would die as a result of counterforce targeting by the US during a first strike against Russia.¹³ 9.8 million in expectation!¹⁴ For context, that's equivalent to the entire population of Portugal or Sweden being wiped out in a matter of 30 minutes.

Counterforce targeting by Russia against the US

¹² I use the same beta distribution generated earlier to estimate the probability distribution of the number of fatalities I'd expect to see result from counterforce targeting against Russia. Just like above, I multiply the beta distribution by the parameter maximum — in this case, the maximum number of deaths we might see. This will generate a probability distribution representing the number of people we'd expect to die.

While I only know the median value of the deaths caused by the MAO-1-like strike, I can estimate the maximum number of deaths by multiplying the median number of deaths (between 8 million and 12 million) by the percent by which counterforce targeting of maximum scale (using 1800 nuclear weapons) would be deadlier than counterforce targeting of the median scale (using 1289 nuclear weapons): 142% (1800/1269). [Again, this rests on the assumption that the relationship between the number of nuclear weapons detonated and the number of deaths is linear.] This gives me a rough estimate of the maximum number of people that would die as a result of counterforce targeting: between 11 million to 17 million. When I multiply this range by the beta distribution, I get the probability distribution for the number of deaths described in the text.

¹³ I don't account for population growth here. I suspect it wouldn't make much of a difference because most detonations will be in remote areas where populations won't have grown by that much.

¹⁴ A report by the U.S. Office of Technology Assessment ([1979](#)) reached a similar conclusion, estimating that about 10 million would die during a counterforce attack on Russia

I use a pretty similar approach to understand the number of deaths we'd expect to see result from counterforce targeting by Russia during a first strike against the US, though there's more uncertainty in the type of counterforce targeting we would expect to see from Russia.

In 1975, the Department of Defense estimated that somewhere between 3.2 and 16.3 million Americans would die during a counterforce strike by Russia ([Daugherty, Levi, & Von Hippel, 1986](#)). A decade later, the Office of Technology Assessment published *The Effects of Nuclear War* ([1986](#)), which estimated that a Russian counterforce attack on the US would leave between 2 million and 20 million Americans dead. They argued that, under the most reasonable assumptions, there would likely be closer to 14 million deaths.¹⁵

In 1986, Daugherty, Levi, and Von Hippel ([1986](#)) sought to build on previous work to produce up-to-date estimates. They estimated what they considered to be the most likely counterforce strike by Russia — a strike using 2,839 nuclear warheads to attack 1,215 targets, including missile silos, strategic bombers, tanker bases, nuclear navy bases, nuclear weapons storage facilities, missile launch control facilities, national command posts, early-warning radars, and communications systems. They concluded that the counterforce scenario they modelled would result in the deaths of 13 million to 34 million Americans. I consider this the best estimate out there.

Unfortunately, none of the estimates I've found reflect the counterforce targeting scenario we would expect to see today. Russia had over 40,000 nuclear weapons in its arsenal in 1986, and some of those weapons were bigger than any weapon in any nuclear arsenal today. The scenario modelled by Daugherty, Levi, and Von Hippel ([1986](#)) used 2,839 nuclear warheads — Russia has only 1,600 deployed nuclear warheads today — and those nuclear weapons had a combined explosive yield that's about double what's in the modern Russian arsenal (including both deployed *and* un-deployed nuclear warheads)([Kristensen & Korda, 2019](#)).

Given this, I make some simplifying assumptions in order to generate probability distributions representing the number of nuclear weapons likely to be used during a modern-day counterforce attack as well as the number of deaths that that counterforce attack would cause.

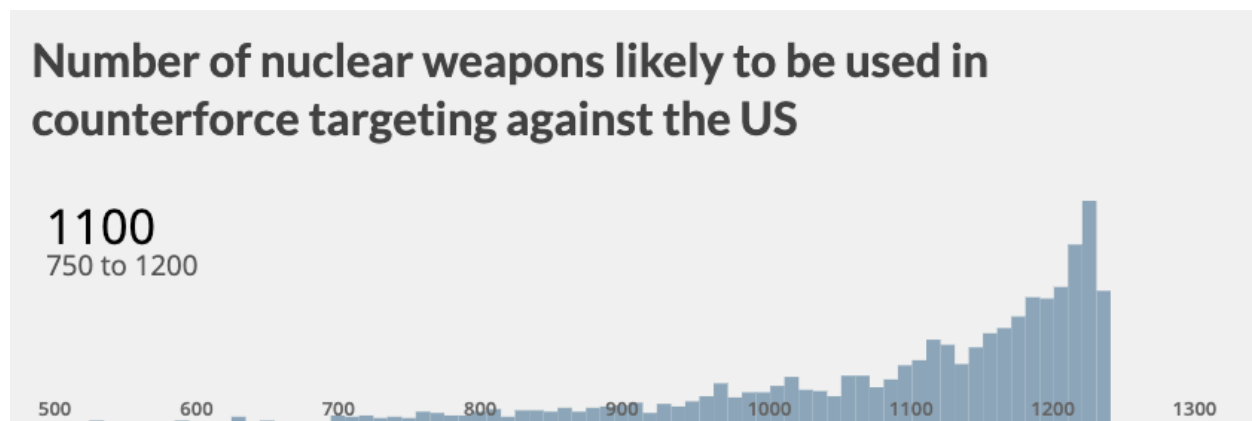
Like in the case of counterforce strike against Russia, I assume that a Russian counterforce strike against the US could plausibly involve as few as one warhead. However, rather than using an upper bound of 1,600, the number of nuclear warheads Russia currently has deployed, I use a maximum of 1,244 warheads — which accounts for proportional allocation of Russia's nuclear forces to the US's 1,800 deployed nuclear weapons as well as the 515 nuclear weapons maintained by NATO allies — France and the UK (more on France and the UK later) ([Kristensen & Norris, 2018](#); [Kristensen & Norris, 2017](#)).¹⁶

¹⁵ According to Daugherty, Levi, and Von Hippel ([1986](#)), as of the writing of their paper in 1986, the US government hadn't released updated estimates since those early estimates. (As far as I can tell, they still haven't).

¹⁶ I assume that, were Russia to use all of its 1,600 deployed nuclear warheads, it would allocate those weapons to the three NATO states with nuclear arsenals — the US, the UK, and France — in proportion to the number of

Other than the likely upper and lower bound, we know very little more about the specifics of a Russian counterforce attack on the US. With nothing else to go on, I assume the most probable counterforce would be on the high end of the plausible range, for the same reason I outlined above: if a country is going to mount a counterforce strike, it seems like they should try to do the maximum damage to their enemy's arsenal possible, in order to reduce the risk/severity of a retaliatory second strike.

Based on this reasoning, I expect that the number of warheads likely to be used in a counterforce first strike by Russia against the US is about 1,100:¹⁷



As I noted above, the Russian counterforce strike scenario modeled by Daugherty et al. ([1986](#)) — the model I consider the best one out there — involved more nuclear weapons than Russia currently has deployed. To account for this, I (again) assume that the number of deaths caused by a counterforce strike scales proportionally with the number of nuclear weapons detonated. More concretely, given that the number of people killed in a strike involving 2,839 nuclear warheads would kill between 13 million and 34 million people, I expect that a counterforce strike involving, at most, 1600 nuclear

nuclear warheads each country has: 1800 deployed by the US (including some hosted abroad), 215 maintained by the UK, and 300 maintained by France ([Kristensen & Norris, 2018](#); [Kristensen & Norris, 2017](#)).

¹⁷ More details on this distribution: I generated a beta distribution that would reflect the proportion of the maximum number of nuclear weapons that could plausibly be used in counterforce targeting that would *actually* be used in Russian counterforce targeting against the US. The distribution could also be thought of as the probability that Russia uses the maximum number of nuclear weapons in its counterforce targeting.

To generate the parameters of the beta distribution I think best fits what little we know about Russian counterforce targeting, I enter what I believe to be the median value, the lower bound, the upper bound, the 0.05th percentile, and the the 0.95th percentile. With help from [SHELF](#), I generated the beta distribution, $\text{beta}(5.3, 0.708)$, which approximates the values I've entered.

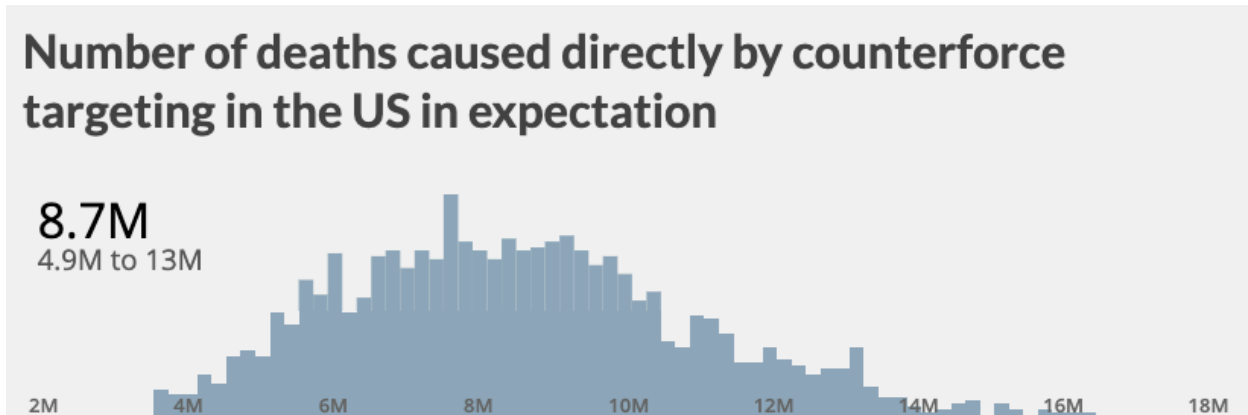
The benefit of the beta distribution is that, because non-probability values are standardized between 0 and 1, we can multiply the distribution by the maximum number of nuclear weapons to get the probability distribution of the number of nuclear weapons used. We can also multiply it by the maximum number of deaths caused by counterforce targeting to get a probability distribution for the number of deaths caused by counterforce deaths at the expected counterforce strike scales.

warheads would kill, at most, 44% (1244/2839) as many people, or between 6.6 million and 14 million people.¹⁸

This means that any estimate using the Daugherty et al.'s ([1986](#)) findings will be an overestimate. Unfortunately, it's the best I can do.

I use the same distribution shape that described the number of nuclear weapons liked to be used in the attack to estimate the probability distribution representing the number of deaths that would result.¹⁹ This again relies on the assumption that the number of people that would be killed during a counterforce strike is proportionate to the number of nuclear weapons used in the strike.

I find that somewhere between 4.9 million and 13 million people would die during a Russian counterforce first strike against the US.²⁰ That's 8.7 million in expectation — equivalent to *everyone* in Austria dying, and then some.



Worse still, Russia would probably target the nuclear forces of the UK and France as well, given their inclusion in NATO and likelihood that they would come to the aid of the US in a nuclear war. I didn't spend much time looking into this, but I think we can get a rough sense of how many people would be killed in those attacks if we make a few simplifying assumptions. I think we can assume that Russia

¹⁸ This calculation also implicitly assume that the nuclear warheads Russia currently has deployed are about the same size as those modelled by Daugherty et al. ([1986](#)). This is actually a pretty weak assumption. The scenario modelled by Daugherty et al. involved 1,342 megatons of explosive yield, while all of the nuclear weapons in Russia's current arsenal have a combined explosive yield of, at most, half that ([Kristensen & Norris, 2019](#)). I don't know how much explosive power there is in the 1600 nuclear warheads Russia has deployed, and the number of deaths caused by different sized nuclear weapons doesn't scale linearly, so it's pretty hard to adjust for this.

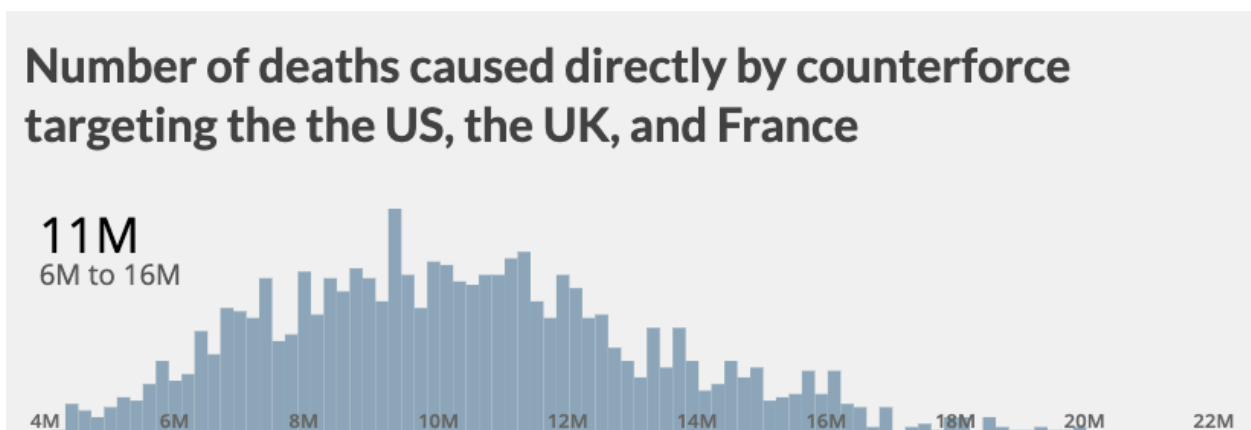
¹⁹ I use the same standardized beta distribution generated earlier to estimate the probability distribution of the number of fatalities I'd expect to see result from counterforce targeting against the US. Just like above, I multiply the beta distribution by the parameter maximum — in this case, the maximum number of deaths we might see (between 5.4 million and 14 million deaths). This generates a probability distribution representing the number of people we'd expect to die from the range of plausible counterforce strikes.

²⁰ Some of the US's deployed nuclear weapons are hosted by countries in Europe. My model implicitly assumes that the population sizes and densities around where those nuclear weapons are being hosted is comparable to those in the locations where US nuclear weapons are housed on US soil.

would employ similar, but proportional, counterforce targeting against the British and French nuclear forces. Given that the UK and France have population densities comparable to that of the US, we should be able to approximate the expected death toll by multiplying the death toll expected in the US by the relative size of the UK and French nuclear arsenals.²¹

As discussed briefly above, I expect that Russia would allocate its deployed nuclear weapons to the nuclear-armed NATO allies in proportion to the number of nuclear weapons each country maintains. This is because I believe the size of each country's nuclear arsenal is a rough proxy for the number of targets Russia would need to hit in order to achieve a successful counterforce strike. Based on this reasoning, I estimate that a Russian counterforce strike on the UK and France would cause an additional 22% of people to die, relative to the number that would die in the US.²²

When I multiply the number of people killed as a result of counterforce targeting in the US by 1.22, I find that between 6 million and 16 million people would die as a result of a Russian counterforce first strike. This range is quite speculative, as it relies on my (pretty baseless) intuitions about how Russia would allocate its deployed nuclear weapons. But it gives us a very rough idea of the death count we might expect to see.



²¹ This assumes that the US, France, and the UK all keep their nuclear weapons in places that have similar populations and population densities, which I expect would be the case (remote military facilities and silos), but may not be.

²² As of 2017, the UK's nuclear arsenal had about 215 nuclear warheads, and France's had about 300 ([Kristensen & Norris, 2017](#)). Again, the US has about 6,550 nuclear warheads, but only about 1,800 are deployed, bringing the combined NATO nuclear forces to about 2,315 nuclear weapons ([Kristensen & Norris, 2018](#)). I therefore expect 22% (515/2315) of Russia's deployed nuclear weapons would be allocated to France's and the UK's arsenals, causing an increase in the death toll of 22%.

Also note that, of the US's 1,800 deployed nuclear weapons, at least 150 of them are hosted by countries in Europe, so the actual death toll in Europe would be a bit higher and the actual death toll in the US would be a bit lower. The overall death toll should be about right, though, as the populations living around the nuclear weapons facilities are probably comparable.

The total number of deaths caused by counterforce targeting

So far, I've estimated the number of lives that would be lost during counterforce targeting in a first strike by the US against Russia, and a first strike by Russia against the US. There would be a first strike by just one of those countries, and the other would likely retaliate with a second strike that we might expect to look pretty different from the targeting plan it would have used during a first strike.

But if the US or Russia detected an imminent first strike — especially one large enough to take out a significant portion of their arsenal — they would almost certainly decide to retaliate in kind before their arsenal was destroyed (this incentive is called “Use it or Lose it”). To do this, each country would have to decide, in a very short amount of time, to “launch on warning,” (LOW) (as opposed to “launch under attack,” which basically just means waiting until the bombs start hitting before you retaliate). Given just a short amount of time, the country under attack wouldn't have much time to tailor its second strike based on factors like which ICBM silos are now probably empty. Given this, I expect that their second strike would probably look reasonably similar to a first strike.

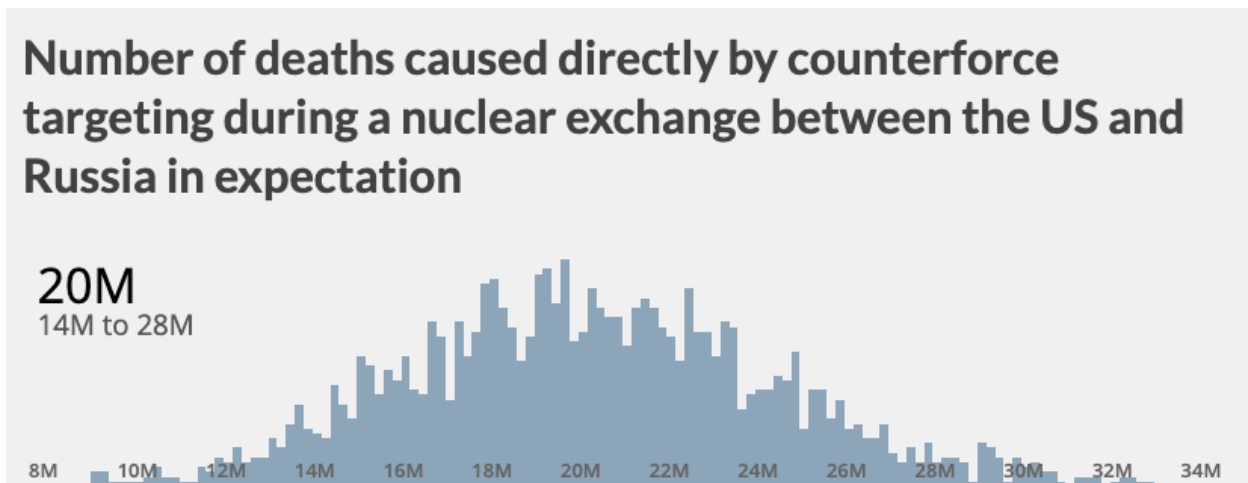
Moreover, to ensure the ‘survivability’ of their nuclear forces, both the US and Russia have divided their nuclear capabilities across three forms of deployment: land-based intercontinental ballistic missiles (ICBMs), sea-launched ballistic missiles (SLBMs), and air-based strategic bombers.²³ This is known as the *nuclear triad*.

The different deployment types make it possible to conceal and move nuclear weapons around, making them harder to detect and destroy, which in turn makes them more likely to survive a counterforce strike. I looked into [the survivability of the US and Russian nuclear arsenals](#) and found that somewhere between ~990 and ~1,660 of the US's nuclear warheads and ~450 and ~1,240 of Russia's nuclear warheads could plausibly survive a counterforce first strike. This makes me think that the US would have enough surviving warheads to execute a second strike that would look similar to a first strike even in the case that Russia *did* destroy its most vulnerable nuclear forces. Because Russia's nuclear arsenal is a bit more vulnerable, its second strike *might* be around the same size as its first strike, or it might be smaller — perhaps ~half the size of its first size. This means that if the US struck first, Russia didn't launch on warning, *and* some large portion of its arsenal were destroyed, simply adding up the deaths caused by a first strike-esque counterforce strike by each country would lead me to overestimate the number of deaths caused by counterforce targeting, though I suspect it wouldn't be by much.²⁴

²³ See for example the [NTI Glossary](#) or ([Congressional Research Service, 2018](#)).

²⁴ Whether a counterforce second strike by Russia would actually cause fewer deaths than a first strike is conditional on 1) the US striking first, 2) Russia choosing not to launch on warning, and 3) Russia being substantially under-prepared for a first strike. My best guess is that the probability of all three of these being the case is fairly low. If we naively assume that the probability that the US strikes first is 50%, the probability that Russia chooses not to launch on warning is also 50%, and that the US counterforce strike destroyed the ‘center value’ of the range for the number of nuclear weapons that might be destroyed (870), or 79% of the number of warheads I expect Russia would use against the US during a counterforce *first* strike (1,100), I would expect that

If I'm right about that, we can roughly estimate the total deaths that would be caused by counterforce targeting during a nuclear exchange between the US and Russia using simple addition. When I do this, I find that between 14 million and 28 million people would die as a result of counterforce targeting during a US-Russia nuclear exchange.



I suspect this estimate is biased in a few ways that have made it an overestimate on net, though there are a few counter-acting biases that likely reduce the extent to which the estimate is inflated. For example, I didn't adjust the fatality estimates from other sources for population growth since the year during which those sources completed their analysis. For those interested in other possible sources of bias, I itemized all the ones I could think of [at the end of the post](#).

Regardless, I think this estimate is unsettling. Counterforce targeting — which again, *doesn't conflict with the humanitarian laws governing war* — would have devastating consequences for the civilians that have the misfortune of living or working in areas where their governments store nuclear weapons. And as I'll discuss in my next post, the indirect impacts of counterforce targeting could be even more devastating.

First, though, I explore the potential for the US and Russia to engage in countervalue targeting — the deliberate targeting of population centers and industrial areas in order to cause an enemy as much pain and suffering as possible.

Would the US or Russia use countervalue targeting as a nuclear exchange escalated?

about 5% fewer deaths would be caused by a Russian second strike than by a Russian first strike ($0.5 \times 0.5 \times 0.21$). See [my post on the survivability of the US and Russian nuclear arsenals](#) for more details.

All wars have implicit and explicit limits — the products of laws, norms, and morals. Many of those limits are immutable over the span of a conflict. When a spoken or unspoken limit is breached, the conflict is said to have escalated ([Morgan et al., 2008](#)). And according to analysts at the RAND corporation, over the course of a conflict, *escalation thresholds* tend to get lower: “especially as losses mount, once-forbidding escalation thresholds often become easier to cross, as the associated costs and risks begin to pale in comparison to those already being incurred” ([Morgan et al., 2008, 35](#)). This means there’s an underlying force propelling conflicts toward further escalation. But some conflicts *don’t* escalate much or even at all. This tends to be the case for two reasons:

First, escalation is costly ([Morgan et al., 2008](#)). Deploying more military personnel may boost one’s chances of winning a war, but it’s also expensive and increases casualties. Similarly, the use of more devastating weapons makes victory more likely, but can lead to domestic and international backlash. For example, recall the international stigmatization of Russia following accusations that it was developing biological weapons during the Cold War.

The other reason escalation isn’t inevitable is that, while escalation offers the promise of success, it also risks further escalation by the other side ([Morgan et al., 2008](#)). This clearly motivated US restraint during the Vietnam War, as the US was extremely averse to escalating the conflict to a point that would risk Chinese or Soviet involvement ([Morgan et al., 2008](#)).

If the US and Russia *were* engaged in a nuclear exchange, it’s impossible to predict which of these competing drives would prevail. Nonetheless, I think there are a few key considerations that might inform our thinking on which is more likely.

Countervalue targeting by Russia in the US

As I discussed earlier, I think there are very [good reasons for Russia *not* to target US cities and industry](#). First and foremost, as the Brookings Institution put it, “retaliating against cities would be suicidal unless one’s own cities had already been destroyed.”

But there are a few compelling reasons why it might engage in countervalue targeting anyways:

1. While countervalue targeting is generally considered strategically inferior to counterforce targeting, the opposite may be true in certain circumstances.

Countervalue targeting could allow Russia to achieve ‘escalation dominance,’ which is when one side of a conflict escalates a conflict in a way that the other side can’t match (for example, because of a lack of military capabilities). Escalation dominance can give one side an upper hand in an otherwise evenly matched or even unfavorable arena.

According to the RAND Corporation report on escalation ([2008](#)), the US’s stated unwillingness

to target civilians could be an opportunity for Russia to gain escalation dominance.²⁵ If Russia could target US cities knowing that the US wouldn't reciprocate, it could inflict enormous damage on the US without risking proportionate retaliation.

2. While I still suspect that the reasons not to engage in countervalue targeting outweigh this potential advantage, Russia might consider countervalue targeting if it felt there was no other way to de-escalate the conflict. For example, Russia might consider countervalue targeting as a final resort if it found that the US nuclear forces were so survivable that counterforce targeting failed completely and seemed likely to continue to fail.
3. There are probably circumstances where Russia might respond to a first strike with countervalue targeting for reasons not related to strategy. For example, I can imagine Russia pursuing a countervalue second strike out of fear, rage, and grief. It's hard to guess how likely these circumstances are, and the likelihood probably fluctuates over time depending who's in power along with a host of other factors.
4. Finally, the US is very unlikely to use all of its nuclear weapons in a first strike, so there is still enormous strategic benefit toward eliminating as much of the remaining arsenal to minimize the severity of the rest of the exchange. And because Russia couldn't be sure which nuclear weapons the US launched, it would have to consider launching a full-scale counterforce attack back to make sure it destroyed as many as possible.

Nonetheless, if Russia were retaliating against the US after a first strike, it would have less reason to attack the US' nuclear forces as, presumably, the US would have already used much of its nuclear arsenal during the first strike.

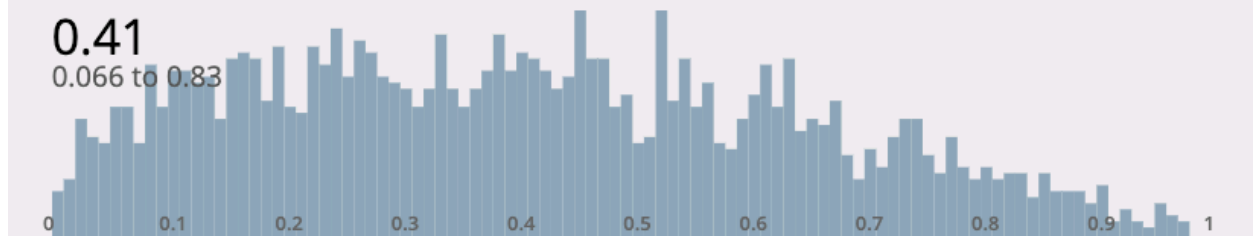
Weighing what I know, I expect that the risks associated with targeting US cities would greatly outweigh the benefits, so I think it's a bit less likely to do so than not.^{26,27} But I'm very uncertain about, so I keep my probability distribution of the probability that Russia would target US cities fairly wide. My subjective view is that the probability that Russia uses countervalue targeting after a first strike is somewhere between 7% to 83%.

²⁵ "...escalation dominance often has more to do with exploiting the enemy's asymmetric vulnerabilities than with developing unique means of attack: For example, U.S. aversion to killing large numbers of civilians often gives an escalatory advantage to insurgent enemies that has nothing to do with a lack of U.S. combat capability." ([Morgan et al., 2008, p. 17](#))

²⁶ To generate the parameters of the beta distribution I think best fits what we know about whether Russia executes countervalue targeting against the US, I enter what I believe to be the median value, the lower bound, the upper bound, the 0.05th percentile, and the the 0.95th percentile. With help from [SHELF](#), I generated the beta distribution, $\text{beta}(1.43, 2.04)$, which approximates the values I've entered.

²⁷For tractability reasons, I'm henceforth assuming that Russia wouldn't use countervalue targeting against other NATO states. This will lead me to underestimate the harm caused by Russian countervalue targeting.

Probability of any countervalue targeting by Russia against the US



What would it look like if Russia *did* decide to target US cities and industries?

A single countervalue strike, or series of strikes, on US cities could range from quite small to huge. For simplicity, I imagine three types of countervalue targeting scenarios:

1. A limited strike, defined, somewhat arbitrarily, as involving between 1 and 20 nuclear detonations.²⁸
2. A moderate strike, or series of strikes, involving between 21 and 50 nuclear detonations.
3. A full-scale countervalue strike, or series of strikes, involving as few as 51 nuclear detonations, and as many as 1,240 ([the maximum number of nuclear warheads that would have a decent chance of surviving a US counterforce first strike](#)).²⁹

I then tried to assess the probability that Russian countervalue targeting against the US would fall within each range.

I'm inclined to think that Russia wouldn't target US cities, if it *did*, it would only target a relatively small number of them. The biggest reason for this is that there are steeply declining "returns" to countervalue targeting in countries with urbanized populations like the US. The first nuclear bombs dropped on the first 10 US cities would kill almost 9 million people, while nuclear detonations on the 51st – 60th targets would kill just under 2 million. Once you hit the 241st – 250th targets, an additional 10 nuclear detonations would kill only another 11,000 people — a loss to be sure, but not nearly as devastating as the first 9 million. This makes me think that, at some point, dropping additional warheads on civilians wouldn't offer much strategic benefit. I expect a similar principle applies to the targeting of industrial zones.

²⁸ As I discuss more later, I assume that all of the bombs detonated would have the explosive yield of the median-sized bomb in the attacking country's arsenal: 300 kilotons in the US arsenal and 500 kt in the Russian arsenal ([Kristensen & Norris, 2018](#); [Kristensen & Korda, 2019](#)). This will probably underestimate the impacts on large and important targets, like huge cities, and overestimate the impacts on industrial targets, and smaller, less important cities. Which of these effects is bigger depends on the specifics of the targeting strategy and the scale of the exchange.

²⁹ Within each scenario, the number of nuclear weapons used is assumed to be uniformly distributed.

Surveys of experts and superforecaster predictions offer additional support for the limited-escalation hypothesis.

In 2018, the Good Judgment Project asked its superforecasters to make predictions about the number of nuclear weapons that would be detonated in the event of a state-sanctioned nuclear attack occurring before 2021. The forecasters predicted that a nuclear exchange taking place in the next few years would be much more likely to involve between 1 and 9 detonations (84%) than 10 or more nuclear detonations (16%) (unpublished data from Open Philanthropy Project). When asked to explain their predictions, the superforecasters cited “Nuclear doctrines of the major arsenal countries” as a key factor, with comments like: “This is effectively a question about which state actors will be involved and whether they will act with a substantial fraction of their arsenal. The only way you get to 1000+ is countervalue attacks by Russia and the US. That is actually against doctrine” (unpublished data from Open Philanthropy Project).

Survey results from the Global Catastrophic Risk (GCR) survey paint a similar picture ([Sandberg & Bostrom, 2008](#)). Respondents believed that the probability that nuclear wars would cause between 1 million and 1 billion deaths by 2100 was 20%, while they estimated the probability that nuclear war would cause over a billion deaths by 2100 was half that — just 10%. This suggests that experts might think it’s more likely that nuclear wars would stay relatively small rather than escalate (though there could be lots of things at play here — see the footnote for more details).³⁰

But some experts disagree. In an [interview with the Future of Life Institute](#), [Robert de Neufville](#) — a Good Judgment Project superforecaster and Director of Communication at the Global Catastrophic Risk Institute (GCRI) — argued that there *could* be a nuclear war involving just a few nuclear weapons but that he’d only expect to see this between countries with small nuclear arsenals (so, not the US and Russia). [Alan Robock](#) — one of the climate scientists who conducted much of the nuclear winter research — argued in a [separate interview with the Future of Life Institute](#) that it’s unlikely that two countries with massive nuclear capabilities would stop after just a few bombs — particularly when facing the threat of many nuclear detonations from their enemy, and even more so in the fog of war where communication is limited. [Seth Baum](#), the executive director at GCRI noted that, despite the decreasing returns to larger scale countervalue targeting, countries might choose to target additional cities, even if they were very small, for signaling reasons — similar to what the US did in Japan.

Weighing the somewhat limited evidence, I put a *bit* more weight on the probability that countervalue targeting by Russia would remain small, using somewhere between one and twenty nuclear

³⁰It’s difficult to draw robust conclusions about the probability of escalation from the GCR Risk survey as there are likely several factors at play. For example, it’s impossible to know whether the respondents’ answers were skewed toward a smaller number of deaths because they believe a) only countries with relatively small populations would be involved (e.g. North Korea), b) only countries with limited nuclear arsenals would be involved (e.g. India and Pakistan, North Korea), c) any nuclear war is unlikely to escalate to the point of using hundreds or thousands of nuclear weapons, d) a nuclear war would involve mostly counterforce targeting rather than countervalue targeting, e) there will be fewer large nuclear wars by 2100 rather than many smaller nuclear wars by 2100, or f) some combination of a-e.

detonations. I also put a fair amount of weight on the probability that countervalue targeting would escalate. If this happened, I expect it would escalate to full-scale countervalue targeting — to the point of using hundreds of nuclear weapons to target US cities and industry — rather than stay moderate in scale. Represented quantitatively, my views look like this, though, importantly, reasonable people would disagree with me here:

Countervalue targeting scenarios	Number of nuclear weapons used	Probability that countervalue targeting stays at this scale
<i>Limited countervalue targeting</i>		
Average	~7	0.5
95% confidence interval	~1 – 23	0.14 – 0.86
<i>Moderate countervalue targeting</i>		
Average	33	0.14
95% confidence interval	20 – 50	0 – 0.47
<i>Full-scale countervalue targeting</i>		
Average	400	0.36
95% confidence interval	150 – 940 ³¹	0.12 – 0.67
Note: The number of nuclear weapons likely to be used in the limited and moderate scenarios is assumed to be uniformly distributed. The number of nuclear weapons likely to be used in the full-scale scenario is assumed to be lognormally distributed.		

I then aggregate these three scenarios into a single nuclear exchange scenario, which reflects the number of nuclear weapons I'd expect to be detonated in a countervalue attack against the US *in expectation* (so taking into account the probability that an exchange stays limited, escalates a moderate amount, or escalates to a full-scale nuclear war).³²

When I do this, I find that Russia might use about 230 nuclear weapons in a countervalue attack against the US in expectation (90% confidence interval: 3 — 950 nuclear weapons).

³¹ To generate the parameters of the lognormal distribution I think best fits the number of nuclear weapons that might be used in a full-scale countervalue attack (/series of attacks) on the US, I enter what I believe to be the median value, the lower bound, the upper bound, the 0.05th percentile, and the 0.95th percentile. With help from [SHELF](#), I generated the lognormal distribution, lognormal(5.85,0.542), which approximates the values I've entered.

³² I do this aggregation by having Guesstimate sample the number of nuclear weapons likely to be used from the three escalation scenarios (limited, moderate, and full-scale) in proportion to the range of probabilities of each scenario.

230
3.3 to 950

The number of deaths caused by nuclear detonations in civilian targets

The graph displays the number of fatalities for 13 countries across 49 targets. The Y-axis represents fatalities from 0 to 800,000. The X-axis represents targets from 1 to 49. China starts with the highest fatalities (~760,000) and drops sharply. Iran and Egypt also start high (~600,000). Most countries show a general downward trend, with some fluctuations. The United States and United Kingdom show the lowest fatalities throughout the targets.

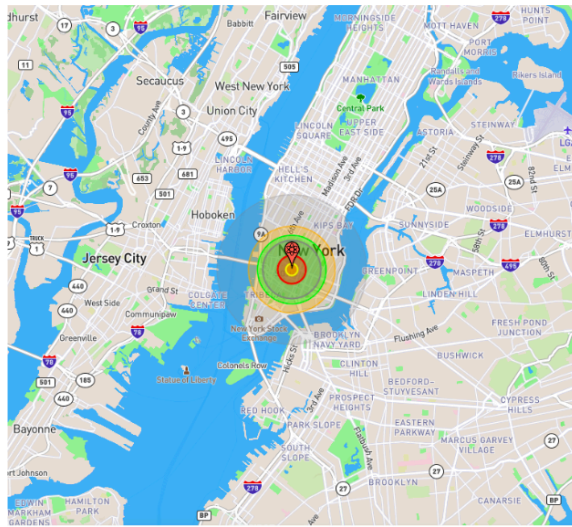
Country	Target 1	Target 5	Target 10	Target 15	Target 20	Target 25	Target 30	Target 35	Target 40	Target 45	Target 49
Argentina	150,000	120,000	110,000	100,000	80,000	70,000	60,000	50,000	40,000	30,000	20,000
Brazil	380,000	200,000	180,000	150,000	120,000	100,000	80,000	60,000	50,000	40,000	30,000
China	760,000	550,000	510,000	590,000	450,000	420,000	380,000	350,000	320,000	280,000	210,000
Egypt	600,000	500,000	330,000	310,000	280,000	250,000	220,000	190,000	160,000	130,000	100,000
France	250,000	180,000	150,000	120,000	100,000	80,000	60,000	50,000	40,000	30,000	20,000
India	460,000	300,000	350,000	330,000	280,000	260,000	230,000	200,000	180,000	150,000	120,000
Iran	600,000	450,000	350,000	320,000	280,000	250,000	220,000	190,000	160,000	130,000	100,000
Israel	220,000	150,000	120,000	100,000	80,000	60,000	40,000	30,000	20,000	10,000	5,000
Japan	120,000	100,000	80,000	60,000	50,000	40,000	30,000	20,000	10,000	5,000	2,000
Pakistan	280,000	200,000	180,000	150,000	120,000	100,000	80,000	60,000	50,000	40,000	30,000
Russia	280,000	200,000	180,000	150,000	120,000	100,000	80,000	60,000	50,000	40,000	30,000
United Kingdom	120,000	100,000	80,000	60,000	50,000	40,000	30,000	20,000	10,000	5,000	2,000
United States	220,000	150,000	120,000	100,000	80,000	60,000	40,000	30,000	20,000	10,000	5,000

Fig. 6. Potential fatalities caused by airbursts of 15-kt yield on each of 50 targets in the countries listed ([Toon et al., 2007](#))

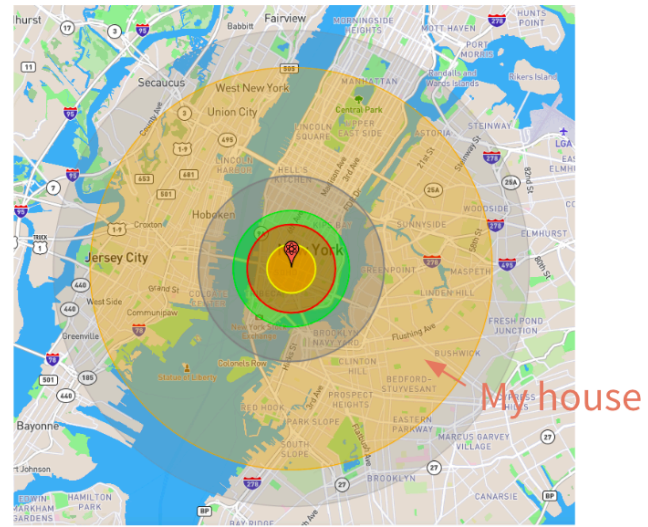
While the authors didn't publish the raw data from all 50 targets for each country, they *did* publish a figure from which the data can be derived and re-fitted. Using their data, I generated a function that describes the relationship between the number of nuclear bombs that might be detonated on US cities and the number of civilian fatalities we'd expect to see as a result (see [Appendix B](#) for details).

$$US \text{ fatalities} = \sum_{n=1}^x (156,573 * e^{-0.0272n}) * 6.4$$

This function takes into account the difference between the bomb yield analyzed in Toon et al. ([2007](#)) and the yield likely to be used by Russia during modern-day exchange:



Little Boy - 15 kilotons
(nuclear bomb dropped on Hiroshima)



SS-18 M6 'Satan' - 500 kilotons
(median sized weapon in the Russian arsenal)

I then transform that equation into one that can be used to estimate the number of deaths that would result from a nuclear exchange using x nuclear weapons (rather than the number of deaths caused by the n^{th} nuclear detonation). Because the equation above is a geometric series, it can be expressed as the following closed form function:

$$US \text{ fatalities} = \frac{e^{(x+1) * -0.0218} - e^{-0.0218}}{e^{-0.0218} - 1}$$

. From there, I calculate the number of people that would die as a result of countervalue targeting in expectation, by plugging in the number of nuclear weapons that would be detonated in expectation in the US in the event of countervalue targeting by Russia. If Russia were *definitely* to execute

countervalue targeting against the US, and I assume that no city would be hit with more than one nuclear weapon, I expect somewhere between 36 million and 37 million people to die.

However, it seems likely that, if Russia *did* decide to target US cities, they would probably choose to drop more than one nuclear bomb on some cities — probably big cities and/or economically important cities — in an effort to maximize casualties and economic disruption. To try to account for this, I looked at the population size for all of the US cities I expect Russia would target (again, I assume cities would be prioritized in order of population size), and calculated the proportion of the population that would be killed by a single nuclear detonation for each city.

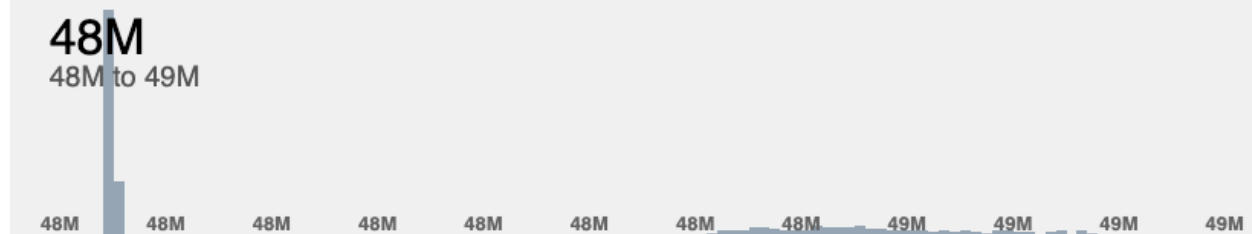
I found that, on average, 75% of the population in the 50 largest cities in the US would be killed as a result of the first detonation, which makes me think that second and third nuclear detonations wouldn't kill sufficiently many people to be strategically advantageous in most cities. That said, a substantial proportion of the population (>50%) could be expected to survive a single nuclear detonation in the four largest US cities. I expect that Russia would see value in detonating multiple nuclear weapons in those cities.

To estimate how many nuclear weapons would be detonated in those four cities, I make two assumptions: first, I assume that second, third, etc., nuclear weapons would cause the same number of deaths as the first. Second, I assume that Russia would allocate additional nuclear weapons to a given city until the point where an additional nuclear bomb would kill fewer people than the former (in effect, the point at which the surviving population would be smaller than the maximum potential deadliness of another nuclear weapon). Neither of these assumptions are great — especially the assumption that subsequent detonations have the same impact as the first — but they likely bias the estimate in opposite directions, which means some of the bias will net itself out.

By making these assumptions, I can roughly estimate the number of nuclear weapons that would be dropped on the four largest US cities. Using the formula above, I can then calculate the number of additional deaths that would be caused by the detonation of multiple bombs on particularly big cities. When I did this, I found that multiple detonations on large cities would probably lead to approximately 11 million additional deaths.

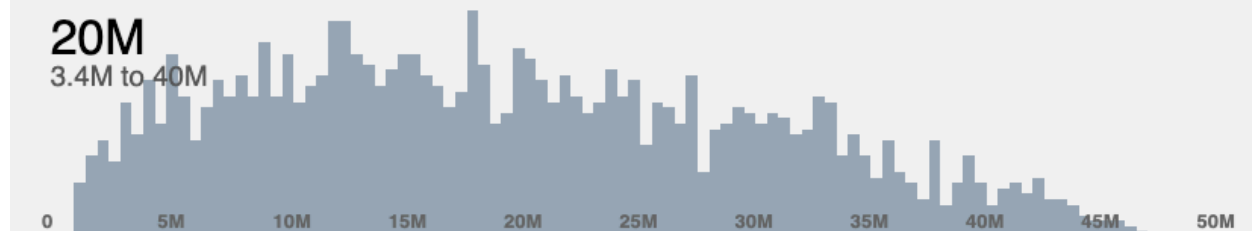
When I add this to the estimate above, I find that countervalue targeting in the US would lead to the deaths of about 48 million people (90% confidence interval: 48 million – 49 million deaths).

48M
48M to 49M



As a final step, I account for the probability that Russia engages in countervalue targeting in the first place. To do this, I multiply the number of people who might die as a result of countervalue targeting by the probability that Russia executes *any* countervalue targeting against the US. I find that, in expectation, between 3.4 million and 40 million people would be killed.

20M
3.4M to 40M



Countervalue targeting by the US against Russia

Next, I explore whether the US would consider countervalue targeting against Russia, and how bad it would be if it did.

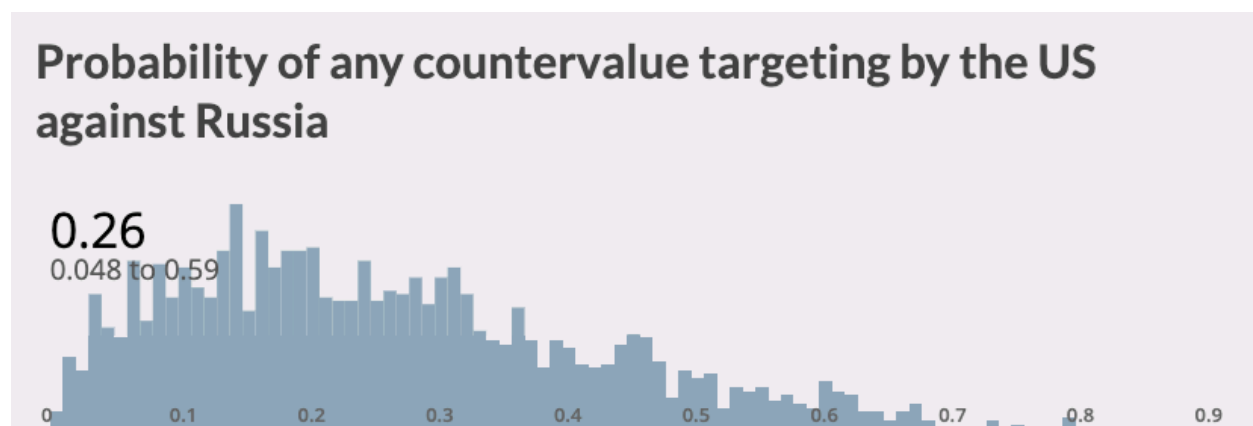
Again, I think there are [very good reasons to think that the US wouldn't target Russian cities and industry](#). That said, I think there are a few compelling reasons why the US would consider countervalue targeting — especially if a nuclear exchange were dragging on:

1. According to subject-matter experts, Russia is outmatched by the US in terms of its ability to successfully target its enemies' cities. Countervalue targeting could therefore be a way for the US to gain a decisive advantage during a nuclear exchange.

Based on the US's nuclear posture, which again states explicitly that it won't target civilians deliberately, I don't think the US plans to exploit this advantage. However, like Russia, the US may eventually perceive escalation to countervalue targeting as its only option for ending a nuclear exchange if, for example, counterforce targeting were failing or communications were down. I suspect this wouldn't be the case, as I believe counterforce targeting could eventually lead to a clear military victory, but I'm pretty uncertain.

2. Like in Russia's case, I am less confident that the US would use counterforce targeting in response to a counterforce first strike, as targeting Russian nuclear forces may be seen as less advantageous after a portion of its nuclear forces will have already been used. But again, I think there are still good reasons to expect that the US would mount a counterforce second strike in response to a counterforce first strike. First and foremost, Russia would likely save a portion of its nuclear arsenal for later strikes, so it would still in the US's interest to eliminate the unused nuclear weapons to make sure they couldn't be used later on.
3. Finally, it seems quite possible that countervalue targeting would be used to respond to a counterforce first strike for reasons that weren't strategic or rational.

Given this, I put a fairly wide range on the probability that the US would use countervalue targeting — albeit less wide than the probability that Russia uses countervalue targeting, as the US's doctrine is more explicit about its targeting policies. My best guess is that the probability is somewhere between 5% and 60%, with more weight on the lower end of the range.³³



³³ To generate the parameters of the beta distribution I think best fits what we know about whether the US would execute countervalue targeting against Russia, I enter what I believe to be the median value, the lower bound, the upper bound, the 0.05th percentile, and the 0.95th percentile. With help from [SHELF](#), I generated the beta distribution, $\text{beta}(1.71, 4.7)$, which best approximates the values I've entered.

If the US *did* decide to pursue countervalue targeting against Russia, they could choose to drop just a few nuclear weapons on a couple of cities, or hundreds of nuclear weapons in cities and towns across Russia. Like in the case of Russian countervalue targeting against the US, I simplify this range of possibilities by considering three countervalue scenarios representing different levels of escalation:

1. A limited strike involving between 1 and 20 nuclear detonations.
2. A moderate strike/series of strikes involving between 21 and 50 nuclear detonations.
3. A full-scale countervalue strike/series of strikes involving as few as 51 nuclear detonations, and as many as 1,660 ([the maximum number of nuclear warheads would plausibly survive a Russian first strike](#)).³⁴

I decided to use the same probabilities here as I did [in the case of Russian countervalue targeting](#), as most of the factors that would lead Russia to either show restraint or escalate to larger scales of countervalue targeting would be similar in both countries (i.e. not different enough for me to make meaningful changes to my probability estimates).

Countervalue targeting scenarios	Number of nuclear weapons used	Probability that countervalue targeting stays at this scale
<i>Limited countervalue targeting</i>		
Average	~7	0.5
95% confidence interval	~1 – 23	0.14 – 0.86
<i>Moderate countervalue targeting</i>		
Average	33	0.14
95% confidence interval	20 – 50	0 – 0.47
<i>Full-scale countervalue targeting</i>		

³⁴ Within each scenario, the number of nuclear weapons used is assumed to be uniformly distributed.

Average	500	0.36
95% confidence interval	210 – 1,000 ³⁵	0.12 – 0.67
Note: The number of nuclear weapons likely to be used in the limited and moderate scenarios is assumed to be uniformly distributed. The number of nuclear weapons likely to be used in the full-scale scenario is assumed to be lognormally distributed.		

As before, I aggregate these scenarios into a single probability distribution representing the number of nuclear weapons likely to be used (in expectation) during a countervalue strike by the US against Russia — about 280 nuclear weapons (90% confidence interval: 4 – 1,100 nuclear weapons):³⁶



From there, I use the [same approach as that outlined above](#) (and detailed in [Appendix B](#)) to understand the relationship between the number of nuclear bombs detonated in a countervalue attack and the number of people that would die as a result.

Using the modelling done by Toon et al. (2007), I estimate the following equation, which can be used to calculate the number of Russian fatalities that would be caused by countervalue targeting by the US (see [Appendix B](#) for details):³⁷

$$\text{Total number of Russian civilian fatalities} = \sum_{n=1}^x (122,033 - 40,239 * e^{-0.017n}) * 5.5$$

³⁵ To generate the parameters of the lognormal distribution I think best fits the number of nuclear weapons that might be used in a full-scale countervalue attack (/series of attacks) on Russia, I enter what I believe to be the median value, the lower bound, the upper bound, the 0.05th percentile, and the 0.95th percentile. With help from [SHELF](#), I generated the lognormal distribution, lognormal(6.1,0.473), which approximates the values I've entered.

³⁶ Again, I do this aggregation by having Guesstimate sample the number of nuclear weapons likely to be used from the three escalation scenarios (limited, moderate, and full-scale) in proportion to the range of probabilities of each scenario.

³⁷ In this case, I transform the equations derived from Toon et al. (2007) using data from [NUKEMAP](#) to account for the fact that the median nuclear bomb in the US arsenal has an explosive yield of about 300 kt, not 15 kt ([Kristensen & Norris, 2018](#)).

This equation can be expressed as the following closed form function, which can then be used to estimate the number of people that would die as a result of the detonation of x nuclear weapons:

$$\text{Russian fatalities} = \frac{e^{(x+1) * -0.0170} - e^{-0.0170}}{e^{-0.0170} - 1}$$

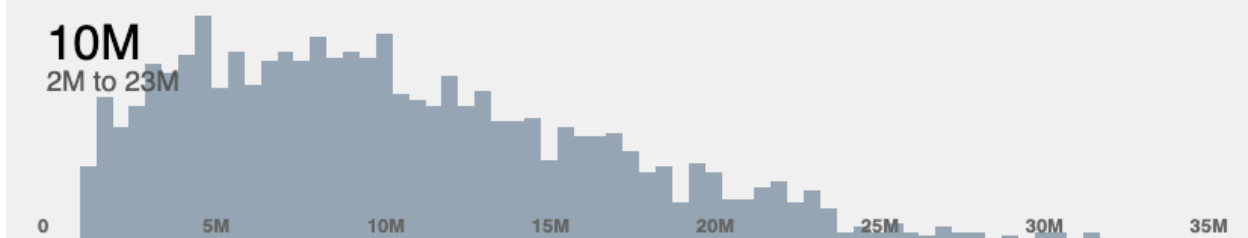
I then solve the equations using the expected number of nuclear weapons that would be detonated during a countervalue strike against Russia. Conditional on the US deciding to execute countervalue targeting against Russia, I expect that somewhere between 39 million and 40 million people would die).

As in the case with countervalue targeting against the US, I expect that a countervalue strike by the US against Russia would involve the detonation of multiple nuclear weapons in larger cities. Using the same strategy and assumptions as above, I conclude that that the US would likely find it worthwhile to drop additional nuclear warheads on the five largest Russian cities. I estimate that these detonations would cause a total of 16 million additional deaths. When I add these deaths to those caused by the first nuclear detonations in each city targeted, I conclude that, altogether, countervalue targeting against Russia would likely lead to about 56 million deaths (90% confidence interval: 55 million – 56 million deaths).



Finally, I multiply the number of people who would die as a result of countervalue targeting by the probability that the US *actually* pursues countervalue targeting against Russia. This gives me the number of people that would die, in expectation. I find that between 2 million and 23 million people would be killed as a result of US countervalue targeting against Russia (10 million in expectation).

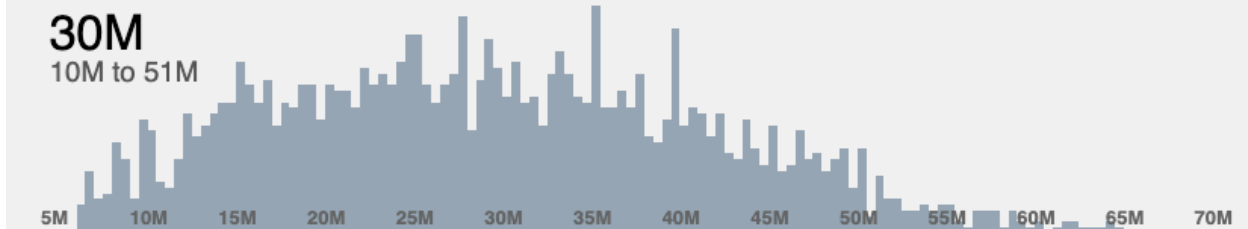
Number of deaths caused directly by countervalue targeting against Russia, in expectation



The total number of deaths caused by countervalue targeting

When I add up the deaths that would be caused by countervalue targeting in Russia and the US, I find that between 10 million and 51 million people would be killed during a US-Russia nuclear exchange in expectation:

Number of deaths caused directly by countervalue targeting during a nuclear exchange between the US and Russia in expectation

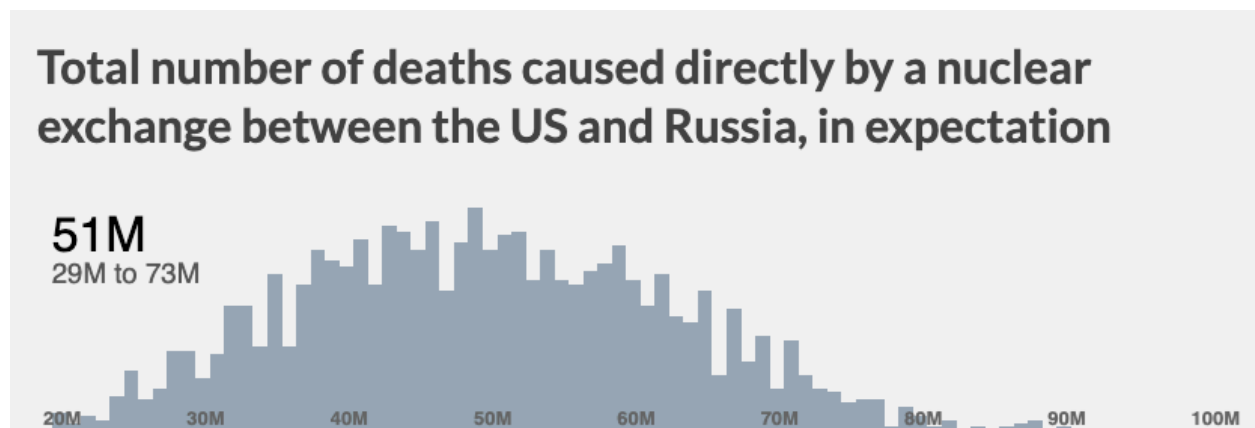


The total number of deaths caused by both counterforce and countervalue targeting

There are a few problems with adding the number of deaths caused by counterforce and countervalue targeting together. First and foremost, we'd have to assume that the ~14 million people expected to be killed during US and Russian countervalue targeting weren't *already* killed during the counterforce

first strike and second strike. This assumption isn't completely unreasonable as the people killed during counterforce and countervalue targeting likely live in pretty different places (almost by definition). The people killed during counterforce targeting would be mostly military personnel, along with the few civilians who live in rural areas deemed remote enough to build missile silos. By contrast, the people killed during countervalue targeting would mostly be people in large cities or hubs of industry — so almost entirely in urban areas.

Given this, simple addition of the two estimates can give us a rough but imperfect idea of just how deadly a US-Russia nuclear exchange would be:



I expect that a nuclear war between Russia and NATO would claim between 29 million and 73 million lives — 51 million in expectation. That's equivalent to *everyone* in Australia, Netherlands, and Hungary dying. And that's just the number of people that would die as a *direct* result of the nuclear blasts and fallout.

In my next post, I'll consider the *indirect* impacts of a US-Russia nuclear exchange, which are even more troubling. Specifically, I estimate the number of people that would die of starvation during the famine that might be caused by a [nuclear winter](#).

Appendix A: Simplifying assumptions and how they might bias my model

I want to be transparent about the fact that many aspects of my model, especially the probabilities of countervalue targeting and escalation, are very speculative and involved a lot of subjective judgment. Below, I summarize 1) all of the simplifying assumptions I made, 2) the ways they might bias my model, 3) how much those biases might affect my results, and 4) how hard it would have been to replace that assumption with actual data:

Assumption	If I were wrong about this, my estimate would be an:	Magnitude of bias	How hard it would be to address
I only consider human suffering	Unknown ³⁸ (0)	Unknown (0)	Very hard
The US and Russia would likely only use counterforce targeting during a first strike	Underestimate (-1)	Moderate (3)	Very hard
The attack plan described by McKinzie et al., (2001), which is based on the SIOP-99, would be reasonably similar to the one that would be used today.	Unknown	Small-Moderate (1.5)	Very hard
The US and Russia would both target most of their enemy's nuclear forces during a counterforce strike to reduce the risk and severity of retaliation (rather than execute a more limited counterforce strike)	Overestimate (1)	Moderate (2)	Pretty hard
I assume zero 'overkill,' which is where multiple nuclear weapons are detonated on the same site to make sure a target is destroyed	Overestimate (1)	Small (1)	Kind of hard
Only deployed nuclear weapons would be used in a US-Russia nuclear exchange	Underestimate (-1)	Large (3)	Very hard
The number of deaths caused by a counterforce strike scales linearly with the number of nuclear warheads in the attack	Probably overestimate (0.5)	Small (1)	Very hard
Russia would allocate its deployed nuclear weapons to targeting NATO allies in proportion to the number of nuclear weapons each country maintains	Unknown (0)	Small (1)	Very hard
The nuclear warheads Russia currently has deployed are about the same size as those modelled by Daugherty et al. (1986)	Overestimate (1)	Moderate (2)	Pretty hard
The US, France, and the UK all keep their nuclear weapons in places that have similar populations and population densities, which I expect would be the case (remote military facilities)	Probably underestimate (-0.5)	Likely small (3)	Not that hard
Some of the US's deployed nuclear weapons are hosted by countries in Europe. The population sizes and densities around where those nuclear weapons are being hosted is comparable to those in the locations where US nuclear weapons are housed on US soil.	Unknown (0)	Likely small (1)	Not that hard
The US and Russia haven't deployed any secret technologies allowing them to track each other's SLBMs, road-mobile ICBMs, or strategic bombers	Overestimate (1)	Large (3)	Very hard
All of the US's SLBMs and strategic bombers would	Overestimate (1)	Large (3)	Very hard

³⁸ I consider the effect on wild and domestic animals ambiguous. This is because I'm not sure if most sentient animals are living net positive or net negative lives, so I'm not sure whether it's a good or a bad thing for them to die en masse.

survive a first strike			
All of Russia's SLBMs, strategic bombers, and road-mobile ICBMs would survive a first strike	Overestimate (1)	Large (3)	Very hard
In many cases, a second strike would look pretty similar to a first strike	Overestimate (1)	Moderate-Large (2.5)	Very hard
A second strike would be proportionate to the first strike	Unknown (0)	Moderate (2)	Pretty hard
A nuclear war would escalate (or not) in the same way and for the same reasons conventional wars escalate	Probably underestimate (-0.5)	Moderate- Large (2.5)	Very hard
In many cases, the costs of escalation to countervalue targeting would outweigh the benefits	Underestimate (-1)	Large (3)	Pretty hard
In the context of retaliation and escalation, revenge would usually be a weaker motivation than strategy	Underestimate (-1)	Moderate-Large (2.5)	Pretty hard
Russia would only execute countervalue targeting against the US — not other NATO countries	Underestimate (-1)	Moderate (2)	Not hard but time-consuming
Within each countervalue targeting scenario, the number of nuclear weapons used is assumed to be uniformly distributed	Likely underestimate (-0.5)	Small-Moderate (1.5)	Pretty hard
In countervalue targeting where multiple nuclear weapons are dropped on one city, the second, third, etc., nuclear weapons detonated would cause the same number of deaths as the first	Overestimate (1)	Moderate (2)	Pretty hard
Both Russia and the US would allocate additional nuclear weapons to a given city until the point where an additional nuclear bomb would kill fewer people than the former (in effect, the point at which the surviving population would be smaller than the maximum potential deadliness of another nuclear weapon)	Underestimate (-1)	Small-Moderate (1.5)	Kind of hard
Civilian targets would be selected on the basis of population and would be prioritized in descending order from most to least populated. This wouldn't hold if industrial areas were targeted instead, or if countervalue targeting was used for signaling purposes, like it (arguably) was in Japan.	Overestimate (1)	Moderate (2)	Pretty hard
Nuclear bombs detonated on military targets would be detonated at ground level (surface/ground bursts)	Underestimate (-1)	Small (1)	Pretty hard
Nuclear bombs detonated on civilian targets would be detonated in the air (airbursts)	Overestimate (1)	Small (1)	Pretty hard
The US and Russia would use the median-sized nuclear bomb in their arsenals during a countervalue strike (or the bombs they actually use would roughly equal out to the median).	Unknown (0)	Unknown (0)	Pretty hard

The relationship between target number and fatalities can be extrapolated out beyond 50 nuclear detonations.	Unknown (0)	Probably small (1)	Pretty hard
The people killed during US and Russian countervalue targeting wouldn't have already killed during counterforce targeting.	Overestimate (1)	Probably small (1)	Pretty hard
Russia and the US wouldn't attack non-nuclear weapons-possessing countries with nuclear weapons.	Underestimate (-1)	Moderate-Large (2.5)	Pretty hard
No other nuclear weapons possessing countries would get involved in a US-Russia nuclear exchange.	Underestimate (-1)	Large (3)	Very hard

I can aggregate these crudely to try to get a general sense of the way my assumptions might bias my model on net:

To do this I first assign a value of -1 to an assumption that would lead my model to produce underestimates, +1 to an assumption that would lead to an overestimate, 0 to an assumption with unknown implications. I then multiply each of those directional values by 1, 2, or 3, depending on the magnitude of the bias (i.e. a small bias would be multiplied by 1, a large bias by 3). I can then add the values up to get a rough indication of the overall direction and magnitude of the bias in my model.

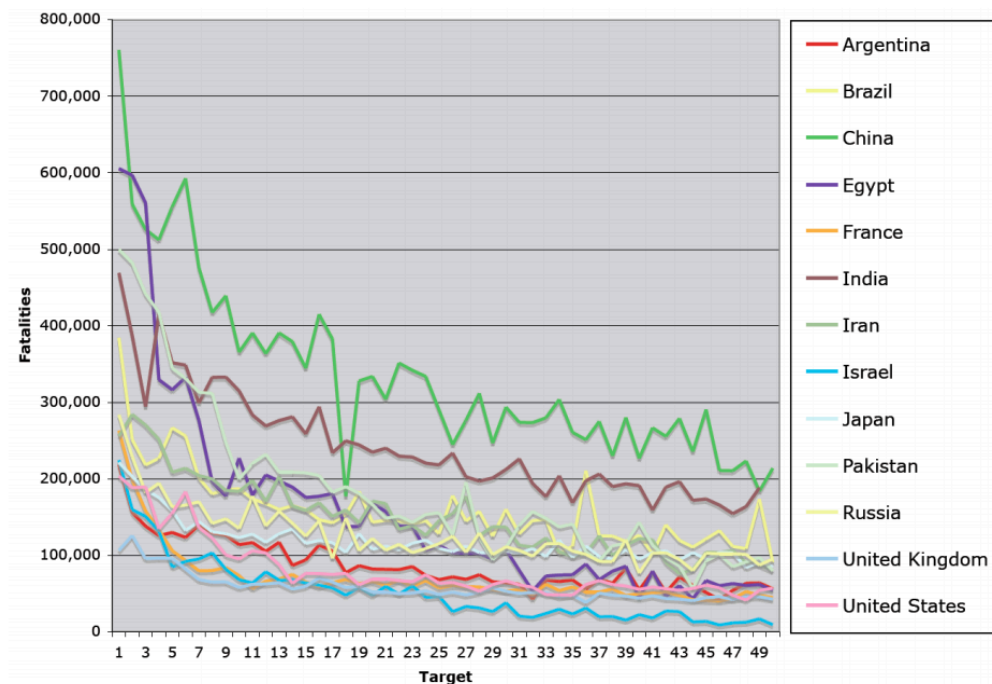
A score of zero would indicate that there are no biases in my model, or equivalently, that all of the biases in my model ~cancel out. Because there are 31 assumptions in my table, a score (in this case) of 93 would indicate that my estimate is likely enormously inflated. In this case, I get a score of 1, which tells me that my estimate may be a bit of an overestimate.

Appendix B: Estimating the number of fatalities caused by countervalue targeting

Researchers have quantified the number of fatalities caused by a nuclear detonation of a specific size, type, and in a particular place. But the number of fatalities varies a lot with the size of the nuclear weapon, the way it's detonated (details below), and the locations targeted. Because of this, the results of most studies can't be generalized to nuclear exchange scenarios not explored specifically in a given analysis.

The nuclear detonation simulator, [NUKEMAP](#), allows a user to input a specific target and bomb size, and reports the approximate number of casualties you'd expect to see result from that detonation scenario. But [NUKEMAP](#) can't realistically be used to quantify the fatalities expected in a nuclear exchange between the United States and Russia, because we just don't know enough about the number of targets, the locations that would be targeted, or the size of the bomb that would be used in each detonation.³⁹

Facing these limitations, I draw on a study by [Toon et al. \(2007\)](#), in which the authors use data from the nuclear bomb dropped on Hiroshima to estimate the number of fatalities we'd expect to see in the event that 50 15 kiloton (kt) bombs were detonated in thirteen key countries.



³⁹ And on a practical note, it would take a really long time to input parameters for nuclear detonations in as many as 1,000 targets in the United States alone.

“Fig. 6. Potential fatalities caused by airbursts of 15-kt yield on each of 50 targets in the countries listed.”

Source: Toon et al. (2007)

These observations can be used to estimate the relationship between the number of targets and the number of casualties we might expect to see. For example, Toon et al. (2007) report that the following function can be used to predict the number of fatalities expected to result from the detonation of a 15 kt bomb in the n^{th} target city in the United States under certain assumptions (discussed below):

$$US\ civilian\ fatalities = 208,940 - 43,055 * \ln(target_number)$$

In an ideal world, I'd be able to use the exact functions published in the Toon et al. (2007) paper to estimate the number of civilian fatalities we'd expect given the number of locations that would be targeted in a US/Russia nuclear exchange. However, Toon et al. (2007) report that the data are best-described by logarithmic decay. And while the logarithmic functions they report fit their observations extremely well, they're less useful when trying to extrapolate beyond 50 targets. This is because, when extrapolated, the logarithmic functions end up predicting a negative number of fatalities (which, of course, makes no sense).

To account for this, I decided to use the raw data from Toon et al. (2007) to estimate my own country-specific equations, but I assume that the number of fatalities per subsequent target decays exponentially rather than logarithmically. This eliminates the possibility of getting nonsensical results like negative fatalities as functions that decay exponentially decay to zero, no lower.

Unfortunately, Toon et al. (2007) didn't publish the raw data from all 50 targets for each country. They *did* however publish a figure from which the data can be derived (see my work [here](#) and [here](#)).

When I do this with data for the United States, I find that the number of fatalities expected to result from the detonation of a 15 kiloton bomb in the n^{th} target city in the US can be estimated using the following equation:

$$Total\ number\ of\ US\ civilian\ fatalities = 156,573 * e^{-0.0272n}$$

If we assume that only one nuclear weapon would be used per countervalue target, we can then use our understanding of the number of nuclear bombs we might expect to be used in countervalue strike by Russia, x , in the equation below. This gives us a rough estimate of the total number of civilian fatalities expected during the countervalue strike against the US:

$$Total\ number\ of\ US\ civilian\ fatalities = \sum_{n=1}^x 156,573 * e^{-0.0272n}$$

For example, we can calculate the number of fatalities we'd expect to be caused by a countervalue attack involving 50 15 kiloton nuclear bombs detonated in US cities:

$$\begin{aligned} \text{Total number of US civilian fatalities} &= \sum_{n=1}^{50} 156,573 * e^{-0.0272n} = \\ & (122,033 * e^{-0.0272*1}) + (122,033 * e^{-0.0272*2}) \dots + (122,033 * e^{-0.0272*50}) \end{aligned}$$

When we solve the equation, we get ~4,075,571 fatalities. We can compare this to the number Toon et al. (2007) get when they actually model the number of fatalities in the 50 densest US cities: 4,056,000. This comparison suggests that the function fits the values derived from more complicated modelling reasonably well.

1. *The targets would have to be selected on the basis of population and would be prioritized in descending order from most to least populated.*

If the US and Russia were choosing to detonate nuclear weapons on civilian targets, I think it's reasonable to assume that a key aim would be to kill as many people as possible by destroying key urban areas. This aim would be best met by prioritizing targets on the basis of population size, making this, in my view a reasonable assumption. On the other hand, if the countervalue targeting centered more around crippling the US economy by targeting industrial areas, extrapolating from Toon et al.'s (2007) data would lead to overestimated fatality estimates.

2. *The bombs would need to be mostly detonated in the air (airbursts), which tend to cause more fatalities than the other main detonation strategy — detonating the bomb on the ground (this is known as a ground burst or surface burst).*

Research has shown that airburst nuclear detonations cause more fatalities than ground bursts, making them more strategic in the targeting of civilian targets.⁴⁰ Ground bursts, also known as surface bursts, are [more effective](#) at destroying underground targets, so they're primarily used to destroy bunkers and underground missile silos. As a result, I expect the vast majority of civilian targets are likely to be targeted with airbursts rather than ground bursts, making this another fair assumption.

3. *The bombs detonated in the exchange would have to have an average yield of 15 kilotons — the size of the bomb dropped on Hiroshima during World War II.*

Russia has in its nuclear arsenal weapons ranging in size from smaller than 1 kt to 800 kt. The median is probably around 500 kt (though it's hard to know for sure) ([Kristensen and Korda, 2019](#)). This would have a large bearing on the number of people that would be killed by a modern Russian nuclear attack relative to the scenarios modelled in Toon et al. (2007).

⁴⁰ For example, see ([Toon et al., 2007](#)).

I therefore conclude that results generated in Toon et al. (2007) shouldn't be *directly* extrapolated to an exchange between Russia and the US.

As a work around for this problem, I transform the equations derived from Toon et al. (2007) using data from [NUKEMAP](#) to account for the fact that the median nuclear bomb in the US arsenal has an explosive yield of about 300 kt, not 15 kt ([Kristensen & Norris, 2018](#)).⁴¹ Similarly the median bomb in the Russian arsenal has an explosive yield of about 500 kt ([Kristensen & Korda, 2019](#)).

I do this by multiplying the results from the Toon et al. (2007) fatality estimates by the factor by which a 300 kt/500 kt nuclear bomb would be deadlier than a 15 kt bomb. To estimate that factor, I use [NUKEMAP](#) to predict the number of fatalities expected from the detonation of a single 15 kt nuclear bomb in 20 cities in the US and Russia (10 in each), and compare that estimate to estimates of the fatalities that would be caused by the detonation of a 300 kt/500 kt bomb in the same location. In the equations below, I refer to this as the Yield Factor (YF).

I picked cities that were large enough to be plausible targets (population > 50,000), but with a range of population densities and sizes to improve my ability to account for the fact that the number of additional deaths caused by larger nuclear weapons is affected by both factors.

From there, I generate a function that describes the relationship between the number of detonations in the US and the number of civilian fatalities we'd expect to see. This function takes into account the difference between the bomb yield analyzed in [Toon et al. \(2007\)](#) and the yield we're likely to see in a modern-day exchange:

$$\text{Total number of US civilian fatalities} = \sum_{n=1}^x (156,573 * e^{-0.0272n}) * YF_{US}$$

For another example, we can calculate the number of fatalities we'd expect to be caused by a limited nuclear attack, where just two 500 kiloton nuclear bombs were detonated in the two densest US cities:

$$\begin{aligned} \text{Total number of US civilian fatalities} &= \sum_{n=1}^x (156,573 * e^{-0.0272n}) * 6.4 = \\ &= \{(\sum_{n=1}^x 156,573 * e^{-0.0272*1}) * 6.4\} + \{(\sum_{n=1}^x 156,573 * e^{-0.0272*2}) * 6.4\} \end{aligned}$$

When we solve this equation, we get an estimate of about 1,924,197 fatalities. As a quick sanity check, we can compare this to the number generated by [NUKEMAP](#), which predicts that the detonation of two 500 kt bombs in the two densest cities in the US would cause 1,184,330 deaths in New York, and 323,340 deaths in Los Angeles, for a combined total of 1,507,670 fatalities. The estimate generated by

⁴¹ NUKEMAP's methodology [is admittedly cruder](#) than other ways of modeling nuclear detonation fatalities. I use NUKEMAP anyways because it allows me to easily vary the size of the nuclear weapons, while most other papers to be too rigid to be extrapolated from.

[NUKEMAP](#) differs somewhat from that generated by my adapted equation, but the two estimates are reasonably close.

Using all of the same reasoning as above, I come to the following formula, which can be used to calculate the number of Russian fatalities that would be expected to be caused by countervalue targeting by the US using x nuclear weapons (see my work [here](#) and [here](#)).

$$\text{Total number of Russian civilian fatalities} = \sum_{n=1}^x (122,033 - 40,239 * e^{-0.017n}) * 5.5$$

Appendix C: What if you think countervalue targeting and escalation are more likely than I do?

The number of people killed during a US-Russia nuclear exchange is very sensitive to parameters that well-informed experts disagree on. In particular, whether the US and Russia would target each others' cities makes a big difference to the number of people that would die. The results are also sensitive to the probability that countervalue targeting would escalate, eventually reaching the point where hundreds of nuclear bombs were dropped on US and Russian cities and industry.

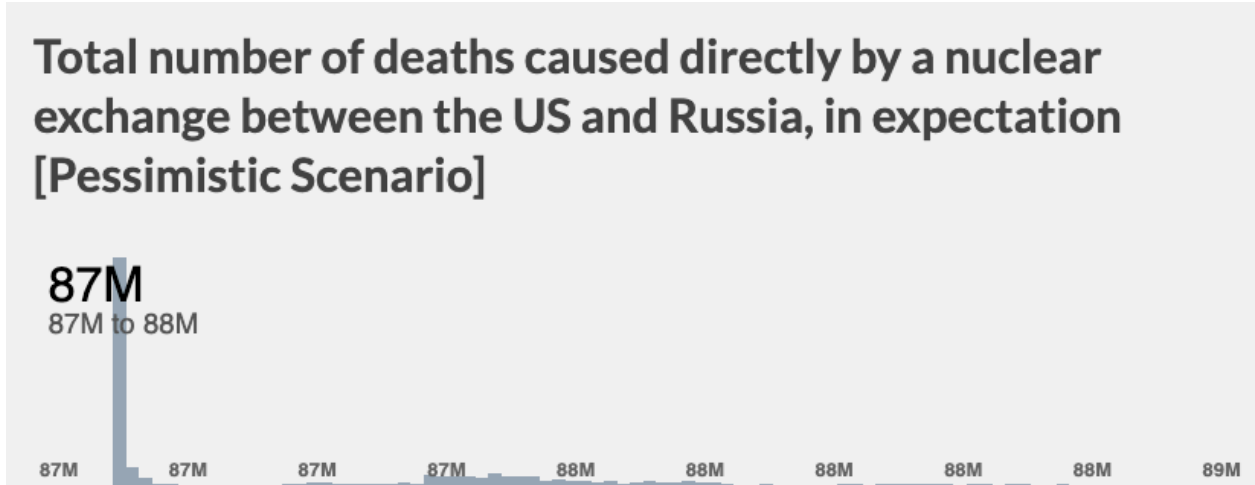
Below, I explore an alternative scenario that illustrates how my results would change if you held more pessimistic views than I do about the likelihood of countervalue targeting and the likelihood that countervalue targeting would escalate.

In this scenario, I assume that neither the US or Russia would use any counterforce targeting — instead focusing exclusively on a countervalue targeting strategy. I also assume that a countervalue nuclear exchange would be less likely to stay limited or escalate a moderate amount, and much more likely to escalate to a full scale countervalue nuclear war.

Pessimistic Scenario: Countervalue targeting used instead of counterforce targeting, and countervalue targeting is more likely to escalate than not				
US-Russia Nuclear Exchange Scenarios	Targeting by the US against Russia		Targeting by Russia against the US	
	Base Case	Scenario 2	Base Case	Scenario 2
<i>Probability of counterforce targeting</i>				
Average	1	0	1	0
95% confidence interval	–	–	–	–
<i>Probability of countervalue targeting</i>				
Average	0.27	1	0.41	1
95% confidence interval	0.05 – 0.58	–	0.07 – 0.81	–
<i>Probability of full-scale countervalue targeting, conditional on any countervalue targeting</i>				

Average	0.36	0.75 ⁴²	0.36	0.75 ⁴³
95% confidence interval	0.11 – 0.67	0.36 – 0.98	0.11 – 0.67	0.36 – 0.98

In this more pessimistic scenario, I estimate that around 88 million people would die in expectation (90% confidence interval: 87 million to 88 million people) — up from 51 million in the base case.



Appendix D: How do my countervalue estimates compare to estimates in the literature?

I make a number of assumptions in my estimation of the number of people that would be killed in a countervalue exchange between the US and Russia — so many assumptions, that it’s hard to know how much to trust the actual model output. To do some basic sanity checking around how trustworthy my estimates might be, I compared my estimates to similar estimates in the literature. As a reminder, by my estimation, about 48 million people would die if only Russia implemented countervalue targeting against the US, 56 million people would die if only the US implemented countervalue targeting against Russia, and 104 million people would die if both countries implemented countervalue against the other:

Countervalue scenarios	My estimate
------------------------	-------------

⁴² This parameter is represented by the beta distribution, beta(3.2,1.09). Because I increased the probability of full-scale countervalue targeting, I had to also decrease the probability of limited and moderate countervalue targeting. I decreased the probabilities of each in proportion to their initial values. After doing so, the probability of limited countervalue targeting became 0.2 (90% CI: 0.02 – 0.53), represented by the beta distribution, beta(1.2,4.73). The probability of moderate countervalue targeting became 0.05 (90% CI: 0.0 – 0.28), represented by the beta distribution, beta(0.343,6.49).

⁴³ See footnote 45

Number of people that would die as a result of countervalue targeting in the US, conditional on Russia pursuing countervalue targeting at all	48 million
Number of people that would die as a result of countervalue targeting in Russia, conditional on the US pursuing countervalue targeting at all	56 million
Number of people that would die as a result of countervalue targeting in the US and Russia, conditional on the US and Russia pursuing countervalue targeting at all	104 million

You can compare these estimates to those published in the literature:

Source	Country	Fatality Estimate	How my estimate compares
Harwell, 1984	US	50 million – 100 million	0.48X – 0.96X
U.S. Office of Technology Assessment 1979	USSR	27.5 million – 55.2 million	1.01X – 2.04X
U.S. Office of Technology Assessment 1979	US	78 million – 173 million	0.28X – 0.62X
McKinzie et al., 2001	Russia, Scenario 1	35 million – 54 million	1.04X – 1.60X
McKinzie et al., 2001	Russia, Scenario 2	31 million – 56 million	1.00X – 1.81X
Daugherty, Levi, & Von Hippel, 1986	US	32 million – 61 million	0.79X – 1.50X
Martin, 1982	US, USSR, Europe, China, and Japan	400 million – 500 million	0.21X – 0.26X
Average factor by which my estimate differs from the estimates in the literature			0.69X – 1.25X

I find that my estimate is about 0.97X the average value in the literature — so, relatively similar to previous estimates on average.

Edits and Corrections

July 18, 2019 - I incorrectly assumed that some portion of French and British nuclear weapons are stored in silos. In fact, the French and British nuclear forces are mostly deployed on submarines, with a few on bombers. I plan to explore how this affects my estimate of the number of people that would die following a counterforce attack by Russia against NATO in the next few weeks.

September 24, 2019 — I originally assumed that only one bomb would be dropped on each target during a countervalue attack. In reality, a countervalue strike would almost certainly involve multiple nuclear detonations within the same city — especially in big cities — using multiple independent reentry vehicles (MIRVs) ([Kristensen & Norris, 2018](#); [Kristensen & Korda, 2019](#)). Because I didn't take this into account, my estimate of the number of deaths caused by US and Russian countervalue

targeting may severely understate the actual number of deaths that would result. In the next few weeks, I'll be revising my model to account for this.

Credits

This essay is a project of [Rethink Priorities](<http://www.rethinkpriorities.org/>). It was written by Luisa Rodriguez. Thanks to Peter Hurford, Marinella Capriati, Ida Sprengers, Marcus A. Davis, Jason Schucraft, and Neil Dullaghan for their 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](<https://www.rethinkpriorities.org/newsletter>). You can [see all our work to date here](<https://www.rethinkpriorities.org/publications>).

Bibliography

- Baum, S. & Barrett, A. (2018). A model for the impacts of nuclear war. Global Catastrophic Risk Institute Working Paper 18-2. *Global Catastrophic Risk Institute*.
<http://dx.doi.org/10.2139/ssrn.3155983>
- Congress of the United States Office of Technology Assessment (1979). The effects of nuclear war. OTA. Retrieved from <https://ota.fas.org/reports/7906.pdf>
- Cooper, H. (2017, April 5). Trump gives military new freedom. But with that comes danger. *New York Times*. Retrieved from
<https://www.nytimes.com/2017/04/05/us/politics/rules-of-engagement-military-force-mattis.html>
- Department of Defense. (2010). Nuclear Posture Review. Retrieved from
https://dod.defense.gov/Portals/1/features/defenseReviews/NPR/2010_Nuclear_Posture_Review_Report.pdf
- Department of Defense. (2013). Report on nuclear employment strategy of the United States specified in section 491 of 10 U.S.C. Retrieved from:
<https://www.globalsecurity.org/wmd/library/policy/dod/us-nuclear-employment-strategy.pdf>

Department of Defense. (2018). Nuclear Posture Review. Retrieved from <https://media.defense.gov/2018/Feb/02/2001872886/-1/-1/1/2018-NUCLEAR-POSTURE-REVIEW-FINAL-REPORT.PDF>

Doctrine for Joint Nuclear Operations. (1995). Retrieved from http://www.nukestrat.com/us/jcs/JCS_JP3-12_95.pdf

Dougherty, W., Levi, B., & Von Hippel, F. (1986). Casualties due to the blast, heat, and radioactive fallout from various hypothetical nuclear attacks on the United States. In F. Solomon & R. Q. Marston (Eds.), *The Medical Implications of Nuclear War*. Washington, D.C.: National Academies Press. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK219165/>

The Editors of Encyclopaedia Britannica. (n. d.). "Nuclear Winter". *Britannica*. Retrieved from <https://www.britannica.com/science/nuclear-winter>

[Good Judgment Project Forecasts]. (n.d.). Unpublished raw data.

Kristensen, H. M., & Korda, M. (2019). Russian nuclear forces, 2019. *Bulletin of the Atomic Scientists*. <https://doi.org/10.1080/00963402.2019.1580891>

Kristensen, H. M., & Norris, R. S. (2017). Worldwide deployments of nuclear weapons, 2017. *Bulletin of the Atomic Scientists*. <https://doi.org/10.1080/00963402.2017.1363995>

Kristensen, H. M., & Norris, R. S. (2018). United States nuclear forces, 2018. *Bulletin of the Atomic Scientists*. <https://doi.org/10.1080/00963402.2018.1438219>

International Committee of the Red Cross (a). (n. d.). IHL Database: Customary IHL: Rule 1. *ICRC*. Retrieved from https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule1

International Committee of the Red Cross (b). (n. d.). IHL Database: Customary IHL: Rule 14. *ICRC*. Retrieved from https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule14

Lieber, K. A., & Press, D. G. (2017). Nuclear deterrence in the computer age: The erosion of stalemate. *Belfer Center for Science and International Affairs*. Retrieved from <https://www.belfercenter.org/sites/default/files/files/publication/Computer%20Nukes%20-%20final.pdf>

McCoy, A. W. (2015, February 24). You Must Follow International Law (Unless You're America). *The Nation*. Retrieved from <https://www.thenation.com/article/you-must-follow-international-law-unless-youre-america/>

McKinzie, M. G. & Cochran, T. B., Norris, R.S., Arkin, W. M. (2001, June). The U.S. nuclear war plan: A time for change. *Natural Resources Defense Council*. Retrieved from <https://www.nrdc.org/sites/default/files/us-nuclear-war-plan-report.pdf>

- Mehta, A. (2017, October 3). Mattis reveals new rules of engagement. *Military Times*. Retrieved from <https://www.militarytimes.com/flashpoints/2017/10/03/mattis-reveals-new-rules-of-engagement/>
- Mizokami, K. (2019, February 26). Russian state TV lists U.S. targets for Putin's new nukes--But one target doesn't make sense. *Popular Mechanics*. Retrieved from <https://www.popularmechanics.com/military/weapons/a26536357/putin-new-nukes-targets/>
- Morgan, F. E., Mueller, K. P., Medeiros, E. S., Pollpeter, K. L., Cliff, R. (2008). Dangerous thresholds: Managing escalation in the 21st Century. *RAND Corporation*. Retrieved from <https://www.rand.org/pubs/monographs/MG614.html>
- NTI Glossary. *Nuclear Threat Initiative*. Retrieved on 2019, May 24 from <https://www.nti.org/learn/glossary>
- Nuclear winter with Alan Robock and Brian Toon [Audio blog interview]. (2016, October 31). *Future of Life Institute*. Retrieved May 19, 2019, from <https://futureoflife.org/2016/10/31/nuclear-winter-robstock-toon-podcast/>
- NUKEMAP (n.d.). Created by Alex Wellerstein, 2012-2019. Retrieved from <https://nuclearsecrecy.com/nukemap/>
- Rudolf, P. (2018). US Nuclear deterrence policy and its problems. Berlin: *Stiftung Wissenschaft und Politik*. Retrieved from: https://www.swp-berlin.org/fileadmin/contents/products/research_papers/2018RP10_rdf.pdf
- Sandberg, A., & Bostrom, N. (2008). Global Catastrophic Risks Survey. *Future of Humanity Institute, Oxford University*. Retrieved from <http://www.global-catastrophic-risks.com/docs/2008-1.pdf>
- Starchak, M. (2017, December 13). Russia terminates development of new rail-mobile ballistic missile. *Eurasia Daily Monitor*. Retrieved from <https://jamestown.org/program/russia-terminates-development-new-rail-mobile-ballistic-missile/>
- Tannenwald, N. (1999). The nuclear taboo: The United States and the normative basis of nuclear non-use. *International Organization*. <https://doi.org/10.1162/002081899550959>
- Toon, O. B., Turco, R. P., Robock, A., Bardeen, C., Oman, L., & Stenchikov, G. L. (2007). Atmospheric effects and societal consequences of regional scale nuclear conflicts and acts of individual nuclear terrorism. *Atmospheric Chemistry and Physics*. <https://doi.org/10.5194/acp-7-1973-2007>
- Watson, K. (2019). Trump nixes public report on civilians killed by drone strikes. *CBS News*. Retrieved from

<https://www.cbsnews.com/news/trump-signs-executive-order-canceling-public-reports-on-civilian-drone-strike-deaths/>

Wirtz, J. J. (2000). The nuclear turning point: A blueprint for deep cuts and de-alerting of nuclear weapons. Edited by Harold A. Feiveson. *Political Science Quarterly*.
<https://doi.org/10.2307/2658039>

Wendt, A. (1992). Anarchy is what states make of it. *International Security*. Retrieved from
http://acme.highpoint.edu/~msetzler/IntlSec/IntlSecReads/Wendt_1992.pdf

What are the odds of nuclear war? A conversation with Seth Baum and Robert de Neufville [Audio blog interview]. (2018, April 27). *Future of Life Institute*. Retrieved May 19, 2019, from
<https://futureoflife.org/2018/04/27/podcast-what-are-the-odds-of-nuclear-war-a-conversation-with-seth-baum-and-robert-deneufville/>

Woolf, A. F. (2009). U. S. strategic nuclear forces: Background, developments, and issues. *Congressional Research Service*. Retrieved from <https://fas.org/sgp/crs/nuke/RL33640.pdf>

Woolf, A. (2018). Defense primer: strategic nuclear forces. *Congressional Research Service*. Retrieved from <https://fas.org/sgp/crs/natsec/IF10519.pdf>