Rootclaim covid-19 origins debate: Final decision

Eric Stansifer

2023 December 18

Abstract

In 2023 November, Rootclaim organized a three part, 20 hour debate with written supplements on the origins of the covid-19 pandemic. Rootclaim, represented by Saar Wilf, claimed that sars-cov-2 was the artificial consequence of gain-of-function laboratory research; challenger Peter Miller contended that sars-cov-2 transferred from an animal host without laboratory involvement.

On the basis of the evidence and arguments presented during the debate, I find with high confidence that zoonotic spillover is the more likely origin of sars-cov-2. The most important basis for this decision is the relative epidemiological proximity of the earliest indicators of covid to a plausible animal source rather than a potential laboratory source.

While Rootclaim identified unusual characteristics of the genetics of sars-cov-2 and the potential for laboratory work to create a similar virus, I found this insufficient to overcome the proximity evidence and the prior against a laboratory origin.

This decision was made independently of any decision reached by the other judge; the debate outcome will be determined by the consensus of both judges.

Contents

1 Introduction

When the "novel coronavirus" started dominating headlines in mid-January 2020, I was checking the news every day: counting cases, tracking international transmissions, wondering when it would reach my city, and preparing for quarantine.[1]

I know now in retrospect that instead of planning for the coming pandemic I should have spent that time reading about covid origins theories, and kept at it for these past four years uninterrupted, as only then might I have been adequately prepared for the deluge of detailed claims and arguments presented during the debate. Before participating in this debate I had no realization of the vast depth of evidence and analysis encompassed in the gamut of "was covid a lab leak?".

It is not possible to answer that question without addressing the depth of evidence head-on, and the length of this document reflects that duty. Even still many of the claims brought up during the debate pass unchallenged here.

However the reader should not be intimidated by its length; the argument is quite simple at its heart, and I have striven to make my discussion approachable.[2]

In this section, we will introduce the necessarily context to understand the differing arguments about the origin of covid: this covers both the spread of the earliest known covid cases, and the details of how the sars-cov-2 virus works and how it relates to other viruses.

In section 2 I will explain and demonstrate how Bayesian updating can be used to estimate the probability of a hypothesis conditioned on observations.

In section 3 I demonstrate how Bayesian inference can be done incorrectly, and I assess Rootclaim's two Bayesian calculations.

Section 4 concerns the most critical question of the debate: how significant was the covid outbreak in the Huanan Seafood Market? This includes an extended discussion of Rootclaim's reasoning for expecting a covid outbreak there under the lab leak hypothesis.

Section 5 covers a miscellanea of topics related to the genetics of sars-cov-2, including the Project DEFUSE proposal that is the centerpiece of the lab leak hypothesis.

In section 6 we have a brief interlude to ask why there has been no whistleblower for the origin of covid.

Finally in section 7, having assessed others' arguments at lengths, I present my own best interpretation of the zoonotic and lab leak hypotheses and their relative likelihood.

Readers may also find the glossary of technical terms in the appendix useful.

In the following, I will be writing at length about my disagreements with the argument presented by Rootclaim, which may create the illusion that I think they were wrong about

[1] which never happened, presumably warded off by my pasta stockpile like rain from an umbrella

[2] I am assuming most readers will be skipping over the equations or more math-y parts, which are mostly non-essential to the text; if you plan on verifying my calculations you will have some work ahead of you, as I usually only gave the minimum information needed to infer how each calculation was performed.

everything (or conversely that I agree with Peter's claims, by absence of discussion of them). This is not true; rather, this document is plenty long even without any discussion of points of agreement. Much of my analysis and reasoning is condensed or omitted.

I have grounded my analysis in the facts and evidence, but it was performed under acute time pressure on a subject which I am not an expert in, did not have a strong pre-existing opinion, and (per the rules of the debate) I could not independently research. Much of the analysis involves conflicting tertiary sources, so that I must critically assess claims made by people with greater expertise or experience than I, and with access to primary or secondary sources that I lack. I have done my best; may the reader be understanding of any errors in fact or judgement that follow.

While I do ultimately compute a numerical confidence level in the competing hypotheses, the resulting number should not be taken in earnest. To get to this quantitative conclusion required assigning numbers to many speculative or unknowable processes for which there are a wide range of believable values. I expect the reader to favor different numbers or choices than I made.

1.1 Background: what happened in Wuhan?

A butterfly with the flap of its wings is said to change the path of a hurricane on the other side of the Earth.[3] There is no better exemplar of the butterfly effect than the beginnings of a pandemic: a single cough, or sneeze, a decision to go to work or see a friend, and the lives of 10 million people swing in the balance.

Inevitably the earlier one tries to trace the origins of covid-19, the narrower the trail, until it is reduced to the mundane daily activities of a handful of individuals whose identities may be lost to time.

The farthest back we have been able to follow the trail is to the city of Wuhan, a massive metropolis of 12 million people[4] in the Chinese province Hubei. On 2019 December 26 or 27, at the height of flu season,[5] Dr Zhang Jixian[6] examined two patients whose unusual CT scan reminded her of her experience with the SARS-1 outbreak of 2002-2003; she is generally recognized as the discoverer of covid-19, thanks to her prompt and appropriate response. Near this time, other healthcare workers began to notice a pattern of pneumonia cases and likewise leapt to action.

On December 30, an emergency notice circulated amongst Wuhan hospitals regarding 27 hospitalized patients with pneumonia of unknown cause, at least some of whom were associ-

[3]This is true, if you mean a hurricane at least two weeks in the future, which is approximately the time scale over which the weather is chaotic.

[4]bigger than any city in the US or west Europe, by some definitions

[5]The average US adult has a 10% chance of catching the flu each winter, and catches 0.5 to 2 colds a year, so one figures that concurrent with the beginnings of covid there were on the order of one million people in Wuhan with the flu, of which ten thousand hospitalized.

[6]at the Hubei Provincial Hospital of Integrated Chinese and Western Medicine, a grade 3A tertiary hospital

ated with the Huanan[7] Seafood Wholesale Market (hereafter "HSM") in downtown Wuhan. Around this time government officials began to become involved, in many ways interfering with the actions of healthcare professionals. HSM was sanitized and closed the morning of January 1, and samples were collected on and after January 1 [22, 12].

On 2020 January 20, the Chinese National Health Commission announced confirmation of human-to-human transmission of covid-19 to multiple healthcare workers; significant evidence of human-to-human transmission existed before this time. On January 23 Wuhan entered lockdown and China invested heavily in tracking and curtailing covid, successfully containing it by April. Covid would remain under control with increasing difficulty until omicron made China's "zero covid" policy impossible to maintain.[8]

1.1.1 Huanan Seafood Market

So; what is the significance of HSM? HSM is a large urban market with at least 1200 employees[9] working in 680 shops, located on the first story of a multistory building; while each avenue of stalls is roofed, the ends are not, and a major road divides the market into west and east halves. The market sold a variety of goods, some of which included wild or exotic live animals. At least some of the trade in animals was likely illegal. While sars-cov-2 originated in bats, it is not known that any bats were sold in the market; any intermediate species between bats and humans are unknown. The market did have live animals that were susceptible to sars-like coronaviruses and were imported from south China, where the progenitor of sars-cov-2 is almost certainly found [12].

Thus when multiple people who were employed at HSM became sick with a virus resembling sars-cov-1, it was natural to suspect the possibility of an animal host at the market being responsible; indeed in early January, before there existed adequate capacity for testing for covid, having an epidemiological link to the market was one component of a diagnosis of covid.

In the time from 2019 December 10, when the earliest known covid case, Wei Guixian, became ill, and the market closure on January 1, approximately 200 people are known[10] to

[7] ie, "South China"

[8] Being off-topic, I elsewhere refrain from giving my *extremely vehement* opinions about the extraordinary incompetence collectively demonstrated by industrialized western countries in the face of the initial covid outbreak, made all the more embarrassing by the fact that the WHO gave advance warning with explicit action items, and China was successful even without those. I think the US should take a hard look at how other countries handle certain issues and have the ambition to at least *try* to be a world leader; compare how the US has 80 km of high-speed rail (the Acela corridor) and plans to build another 192 km by 2033 (in California), whereas China has built 42000 km since 2008. While disease control is easier in an autocracy, I hasten to point out that New Zealand and Taiwan both successfully eliminated covid until omicron. Had the US the same death rate as New Zealand, Taiwan, Australia, Japan, or South Korea, about 1m of the estimated 1.2m US covid deaths would have been averted. There needs to be an accounting for the US's actions that asks why we let 1 million people die.

[9] There are 1162 vendors according to [32] annex page 182, and presumably a variety of other full time employees as well.

[10] You'd think that although we don't know how many people had covid in 2019 December, surely we'd know how many we *know* of. Alas not. This is the elusive "unknown known" of the Rumsfeld quadrafecta.

Figure 1: Two photographs from the west half of HSM, showing the varying air circulation conditions within the market. The first image looks south along the main corridor from alley 13, the second is located within alley 9. These images were taken from `http://babarlelephant.free-hoster.net/visiting-the-wuhan-seafood-market/`; click on the yellow dot above the 13 and two to the left of the 9.

have had onset of covid, of which 62 or 63 were vendors at the market and others had known epidemiological connections to it.

1.2 Background: what is in the sars-cov-2 genome?

Coronaviruses are large viruses whose genome consists of a single-strand of RNA with approximately 30000 nucleotides. The RNA is entangled with many copies of the viral nucleocapsid protein, and when outside of cells is protected by an envelope consisting of a membrane taken from the host cell and studded with viral proteins.

Viral RNA is translated[11] by ribosomes in the host cell to form proteins. This is done in groups of 3 nucleotides at a time, called "codons". As there are 4 possible nucleotides at each position, there are 64 possible codons, each of which yields a particular amino acid; the chain of amino acids constitutes a protein, with the particular chemical properties of the amino acids causing it to fold into a shape suitable for that protein's specific function. The assignment of an amino acid to each possible codon is called the *genetic code*. Humans have 20 different amino acids, so there exist multiple codons that yield the same amino acid, and *synonymous mutations* which change the nucleotide sequence without changing the amino acid sequence are possible. The genetic code is not random: mutations tend to be synonymous, and nonsynonymous mutations tend to yield amino acids with similar chemical

[11]The coronavirus genome is *positive sense*, meaning that first the genomic RNA is transcribed to make negative sense RNA, which is then translated.

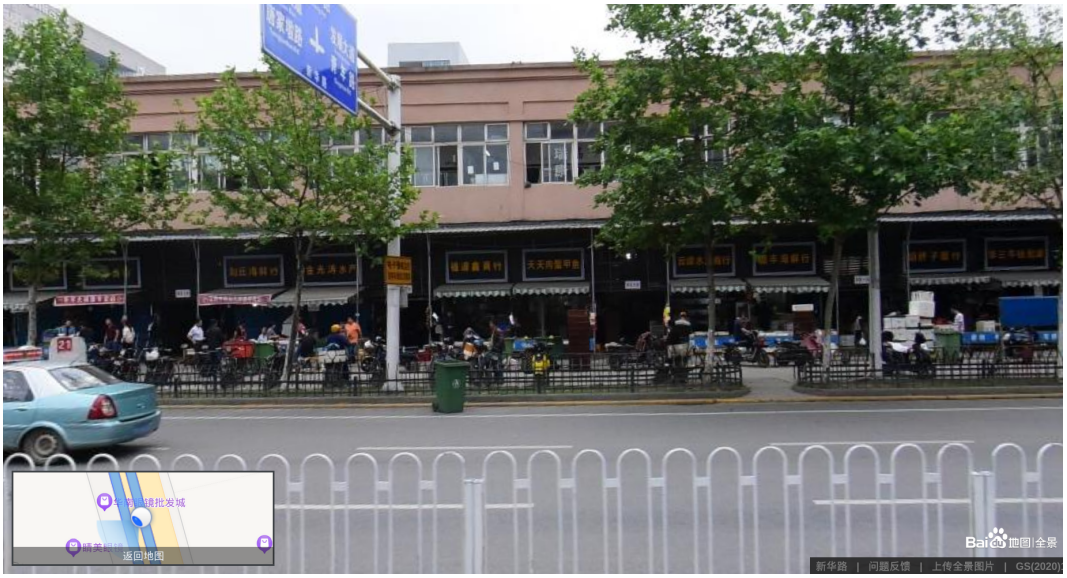

Figure 2: Photograph of the west half of HSM, looking west from Xinhua road, with the east half of HSM behind the viewer. Notice that the end of each alley opens up directly to the air, with two storefronts between each alley. The busy street would have slowed foot traffic between west and east HSM. I believe this image was captured after HSM was closed down but before the current barriers were installed. Photograph from Baidu street view.

properties, or at least more often than one would expect from a randomly designed genetic code.

As host cells do not contain the cellular machinery to replicate RNA, the first two-thirds of the viral genome, called the "open reading frame 1ab" (ORF1ab), encodes 16 different proteins involved in the viral replication and assembly process. The rest of the protein encodes the four structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as accessory proteins. The envelope and membrane proteins are components of the viral envelope, while the large spike protein protrudes outwards from the envelope, making it the first point of contact with other objects. The bumpy texture of the spike proteins resembles a "crown", giving coronaviruses their name.

The spike protein plays an important role in the infection process, as it binds to a matching receptor in the membrane of the target cells. In sars-cov-2, the spike binds to the ACE2 receptor found throughout the human body, including lungs and intestines; after binding, the spike is then cleaved into two pieces, which for reasons inscrutable to me facilitates merging the virus envelope with the cellular membrane and so delivering the genome into the cell. This cleaving process will become relevant after we have introduced the phylogenetics of coronaviruses.

1.2.1 Phylogeny of coronaviruses

As viruses evolve over time, they can be organized into a *phylogenetic tree* showing how closely related one virus is to another; for example, coronaviruses are just one branch

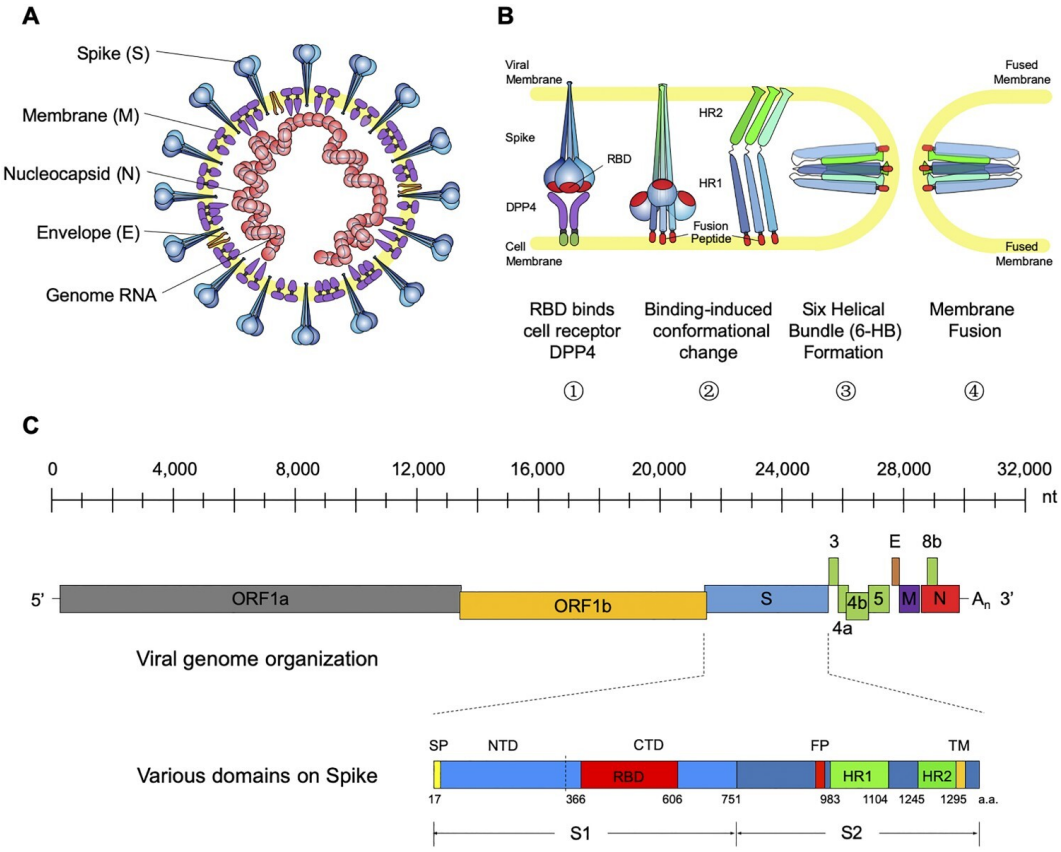

Figure 3: The structure of a coronavirus, how it enters target cells, and its genome. Depicted is MERS, a close relative to sars-cov-2; note that sars-cov-2 binds to ACE2 instead of DPP4. The spike protein consists of two subdomains, S1 and S2, which are separated as part of the cell entry process. Diagram is figure 2 from [35].

of nidoviruses. Coronaviruses are divided into four subgroups, named alphacoronaviruses through deltacoronaviruses, not to be confused with the variants of sars-cov-2 also named after greek letters. β-covs are divided further into four subgroups labeled lineages A through D, not to be confused with lineages A and B of sars-cov-2. Lineage B, the sarbecoviruses ("sars-like betacoronaviruses") includes both sars-cov-1, the virus which caused the 2002-2003 SARS outbreak, and sars-cov-2, which causes covid-19. The genomes of sars-cov-1 and sars-cov-2 are 80% identical.

In total there are 7 known coronaviruses that cause disease in humans: two mild α-covs HCoV-229E and HCoV-NL63, two mild β-covs of lineage A ("embecoviruses") HCoV-OC43 and HCoV-HKU1, the two human sarbecoviruses mentioned above, and the deadly MERS virus of lineage C ("merbecoviruses"). While MERS is only weakly transmissible between humans and most MERS causes are in Saudi Arabia, where it is believed to have transmitted many times to humans from camels, extended human-to-human transmission chains have been observed extending into other countries.

The wild virus most closely related to sars-cov-2 is banal-20-52 [30], a sample collected from

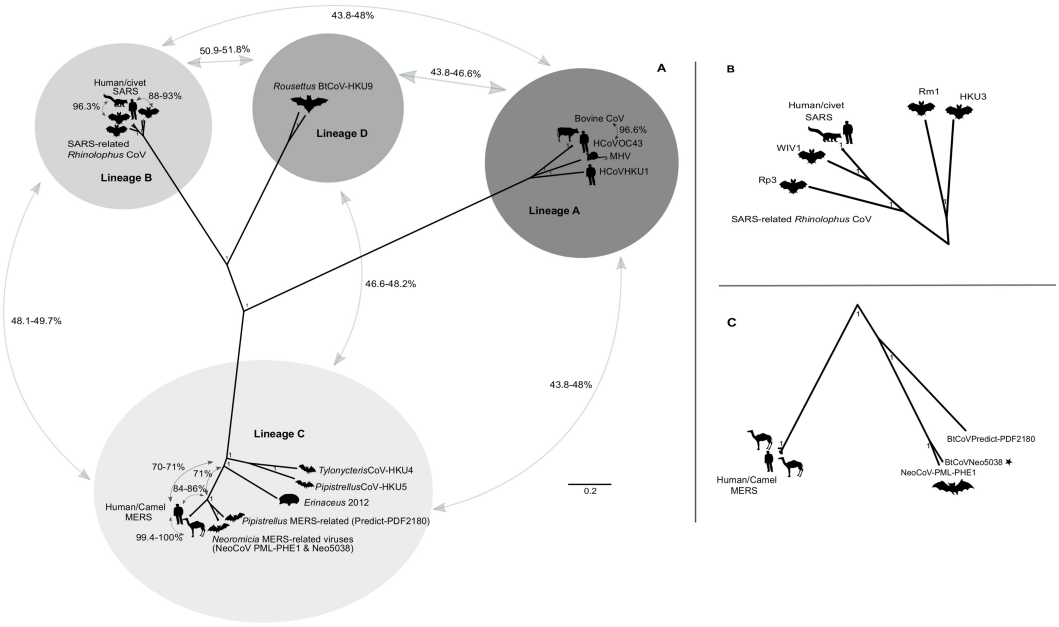

Figure 4: Phylogenetic tree of selected betacoronaviruses. Inserts at right show Lineage B (sarbecoviruses) and Lineage C (merbecoviruses); the former includes both sars-cov-1 and sars-cov-2, but this diagram was made in 2018. Diagram is figure 9 from [17].

a bat (dropping) in Laos with 96.8% nucleotide identity, or a disagreement of about 1000 nucleotides. As many nucleotide substitutions are synonymous, there is about 99% amino acid identity, and the time to most recent common ancestor (tMRCA) might be about 20 to 30 years ago [7]. However, different parts of the sars-cov-2 genome are closest to different wild viruses, and recombination appears to play a role in the genetic history of sarbecoviruses, complicating efforts to reconstruct a phylogeny [30].

Sars-cov-2 itself divides into many subgroups; like the flu virus, one often finds that newer variants displace older variants, causing notational difficulties as the phylogenetic tree divides further and further. It was for this reason that the PANGO notation was developed [29]; for example, the pango notation "B.1.1.529" refers to the root of the omicron variant. The very earliest division of the sars-cov-2 phylogeny is into two parts, called lineages A and B in the PANGO notation.[12] Lineage B would contain the highly successful D614G mutation, meaning that the aspartic acid (D) at location 614 of the spike protein (70 nucleotides from the FCS) is replaced by glycine (G). The D614G mutation first became prominent in Italy in 2020 March, and I believe is directly related to the worldwide March-May peak many countries observed.[13] Lineage B has now completely displaced lineage A.

[12]In Nextstrain, lineage A is called 19B, as it was the second branch to be identified in 2019, and the root of lineage B is called 19A; the root of omicron is called 21M. In GISAID, lineage A is called S clade. Nextstrain and GISAID provide freely available tools for viewing phylogenetic trees, tracking variants, and downloading genomes.

[13]In summer of 2020 I invested some time into tracking the progression of D614G, but never finished my work as it was eventually obviated by the second wave that came that fall.

Due to the high mutation rate of viruses, a phylogenetic tree can be created with some reliability down to the level of individual illnesses. Samples of sars-cov-2 collected in 2019 December and 2020 January have been sequenced and some are available in GISAID, facilitating assembling them into a phylogenetic tree of the very earliest covid cases, though the results are not without dispute. Lineages A and B each clearly form a tree with a single unambiguous root from which there are many branches, called a "polytomy". These two roots differ from each other in exactly two nucleotides. The root of lineage A has the sequence T8782, C28144 (abbreviated T/C) and the root of lineage B has the sequence C8782, T28144 (abbreviated C/T); note that the mutation at 8782 is synonymous. Several early samples have been collected that may exhibit the C/C or T/T pattern. The phylogenetic relationship between the roots of the A and B lineages is not immediately clear without further analysis.

We can summarize the above with a phylogenetic tree:

- coronaviruses
  - betacoronaviruses
    - embecoviruses (aka Lineage A)
    - sarbecoviruses (aka Lineage B)
      - sars-cov-1
      - banal-20-52
      - sars-cov-2
        - Lineage A (now extinct)
        - Lineage B
          - alpha, beta, ..., omicron
    - merbecoviruses (aka Lineage C)
      - MERS

1.2.2 Furin cleavage site

While most of the disagreements between sars-cov-2 and banal-20-52 and other bat coronaviruses are point mutations, there is one important larger difference. Recall that when the spike protein S binds to the ACE2 receptor on a human target cell, to facilitate fusion with the cell the spike protein must be cut into two subregions, called S1 and S2. In sars-cov-2 infections in humans, this cutting process is performed by the furin enzyme.

Furin recognizes proteins with an amino acid sequence of the form `RxxR` or especially `RRxR` (here, "x" stands for any amino acid and `R` stands for arginine), cutting them; thus any such amino acid sequence is referred to as a furin cleavage site (FCS). The proper functioning of sars-cov-2 therefore necessitates that the boundary between S1 and S2 is an FCS. Indeed:

```
CAG ACT CAG ACT AAT TCT CCT CGG CGG GCA CGT AGT GTA GCT AGT
```

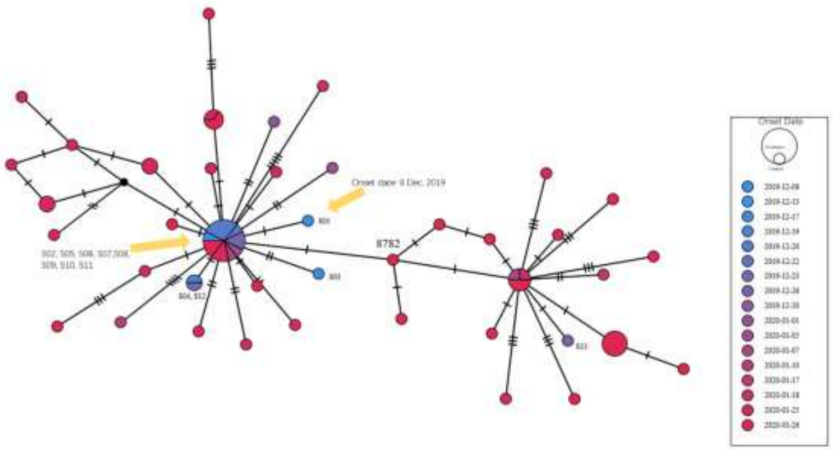

Fig. 7. Haplotype network of early sequences of Wuhan. One viral genome that carried a C>T variant at site 8782 (compared to the reference genome) connected the S/A and L/B major lineages, and this genome was sampled from Wuhan in late January 2020.

Figure 5: Unrooted phylogenetic "tree" of early sars-cov-2 samples collected in Wuhan. Each circle represents a collection of samples with identical genome; the size of the circle is the number of samples, and the color(s) of the circle is the date of illness onset for those people. Bluer circles are earlier illnesses, and redder circles are later. The number of tick marks on a line indicates the number of nucleotides difference between two genomes. The left cluster are lineage B cases, and the right cluster are lineage A cases, with a few ambiguous cases in between. The middle arrow identifies the first known case, a Mr Chen with onset 2019 December 8, although he may have become ill on December 12. (Technically this is a haplotype network, not a phylogenetic tree, in that it shows only the genetic similarity of different cases without any attempt to identify which are ancestral/related to each other. It is possible for distantly related samples to have similar genomes by chance, as the same mutations can arise more than once independently.) Figure from [32], page 79. See also page 69 for a larger tree.

```
Q   T   Q   T   N   S   P   R   R   A   R   S   V   A   S
```

Shown is an excerpt from the nucleotide sequence and corresponding amino acid sequence for Wuhan-Hu-1, the reference genome for sars-cov-2, collected in 2019 December [25]. The `RRAR` subsequence constitutes an FCS.

While sars-cov-2 passes between humans primarily through respiratory droplets, this is not the case for bat coronaviruses, which are spread through droppings and therefore must survive the gastrointestinal tract. This creates a different life cycle for virus replication, and consequentially an FCS may be evolutionarily harmful. We can compare sars-cov-2 to the closest known wild virus banal-20-52 at this location ([3] at 35:30):

```
CAG ACT CAa ACT AAT TCA CGT AGT GTg GCa AGT
 Q   T   Q   T   N   S   R   S   V   A   S
```

I have marked the three point mutations in this section with lower-case letters; each is syn-

onymous. (There are about 1000 point mutations in the 30000 nucleotides.) More notably, however, is that banal-20-52 is 12 nucleotides shorter here, as it is missing the sequence `T CCT CGG CGG GC`.[14]

1.3 Format of the debate

Rootclaim was created to demonstrate the usefulness of a particular methodology for analyzing complicated questions, where there exists abundant lines of circumstantial evidence both for and against any particular hypothesis. After systematically evaluating each piece of evidence, they can be integrated into a gestalt that gives a final level of confidence we can have in each hypothesis.

I refer to this technique as *Bayesian reasoning* or *Bayesian computation*, which will be discussed in the next section; however Rootclaim are far from the only ones to use Bayesian reasoning, and I believe they claim to distinguish themselves from others in the details of how the evaluate the evidence. In any case I agree with their critiques of other methodologies (or, frequently, lack thereof) and that most people who get the right answer do so more by chance than not.

This debate is the first organized by Rootclaim, of hopefully multiple (on different topics, to be clear!), as a series of 3 spoken debate sessions totalling 20 hours [1, 2, 3, 4, 5, 6] and various written supplements [7, 8]. I am one of two judges, deciding independently, with the outcome of the debate determined by the average of our decisions.

I assess the debate structure designed by Rootclaim to be very fair and not create any advantage for them, and it was clear that Saar consciously acted in good faith throughout. To me, the debate felt more like a collaboration of four people working together to seek the truth than an adversarial contest; it is rare to see contentious issues worked out in such a productive manner between people acting in good faith.

As I find that both parties argued in good faith and there were no significant violations of (the spirit of) the debate rules, my evaluation is based solely on the arguments as presented at an object level, or my best interpretation thereof. Peter identified and criticized a diversity of lab leak theories put forth by other parties, but these had no bearing on Rootclaim's theory. Statements were evaluated independently of the credibility of other parties who assert them. I did not attempt to "keep score" of which side made more factual errors or changed their argument more often; each claim was given independent credulity. Neither side was laden with the burden of proof. While in other contexts it might be appropriate and fair to dismiss arguments a priori for rhetorical reasons (eg, as a defense against techniques like sea lioning or gish gollop), here my sole concern is whose theory is best supported by the evidence.

Per the rules of the debate, I am to evaluate the relative likelihood of the two hypotheses advanced: Rootclaim's claim that the sars-cov-2 outbreak was due to gain-of-function laboratory research, and Peter's challenge that it was the result of natural spillover from wildlife without laboratory involvement. Only the information provided through the debate, debate

[14]Note that alternatively Rootclaim suggests that the insert is `CCT CGG CGG GCA`, and there was a fourth point mutation from `TCT` to `TCA` immediately before it.

material, and references therein may be considered; no prior beliefs or private research of my own can be used, as this would deny the participants opportunity to identify flaws and rebut such arguments. In so far as was reasonable within the constraints of time I attempted to challenge each participant on any claims they made that appeared suspect, so that they had opportunity to bolster their argument or substitute it with another. Unfortunately time is limited and only so many rounds of argument and counterargument can be sustained; eventually one must reach a decision.

I have laid out here in some detail how I reached the decision I did, though in view of the *considerable* volume already written on this subject I do not imagine this will be sufficient to compel anyone to change their mind. Doubtlessly the participants of the debate will have rebuttals to what I write that disagrees with them, and perhaps I may wish to revise this in the future, but at least for the purposes of the debate it is finished.

1.3.1 About this document

My foremost goal in this document is to give the best possible account for the origins of covid as I understand them, and only incidentally explain how I reached my decision. Therefore, to avoid leaving glaring holes in explanatory passages, I have occasionally looked up minor details, eg to fill out the background section or section B; in so far as possible I have restricted myself to sources directly cited by the debate parties, though I may have inadvertently used a few indirect citations. Most of this information had no bearing on the decision; the one piece of information I found that did was relegated to appendix C.

In section 7 I present an uncomfortably quantitative analysis of the two conflicting theories and my affirmative conclusion for which is more likely based on the information from the debate. However it felt unsatisfying to say why I believed the arguments *for* one theory without also saying why I failed to be persuaded by the arguments against it. The majority of this document, particularly sections 3.1, 3.2, 4.5, and A, contains detailed criticisms of arguments presented by Rootclaim, identifying where I believe those arguments to be flawed. These refutational passages are not part of my affirmative conclusion, so I have taken greater liberty in the use of complicated models or analysis. For the affirmative analysis I have eschewed such complexities: best to keep it simple and pay a small price in accuracy rather than risk large model error.

2 Introduction to Bayesian computation

Suppose you have two exclusive events[15] A and $B = \neg A$ you wish to distinguish. Unfortunately we cannot observe them directly; instead we make some sequence of other observations $O_1, O_2, \ldots, O_n$.

If one of our observations O_1 were, say, incompatible with A, then we would be done; by simple deductive reasoning we could say "A implies $\neg O_1$; O_1; therefore $\neg A$". Instead each

[15] In probability theory, an *event* is a statement which has a probability of being true. We assume events follow certain rules, such as if X and Y are events then there is an event called $X \cap Y$, representing that both X and Y are true, satisfying $P(X \cap Y) \leq P(X)$.

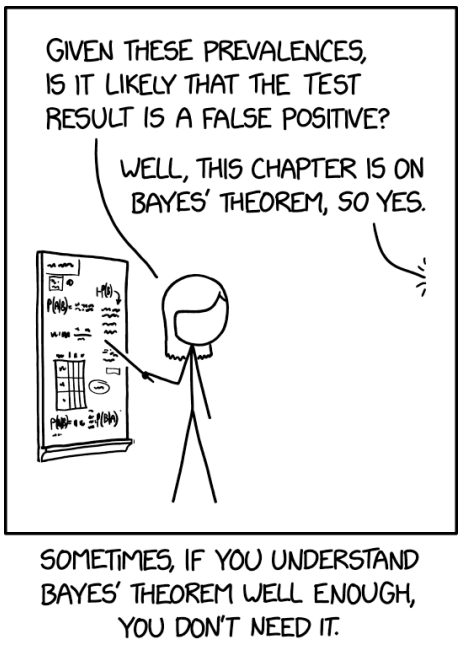

Figure 6: XKCD 2545

of our observations are consistent with either A or B – but, critically, not *equally* so. We measure the degree of consistency with the conditional probabilities[16] $P(O_i|A)$ and $P(O_i|B)$. From this information it is a simple application of Bayes' Theorem to compute $P(A|O_1)$, and repeated application to get $P(A|O_1 \cap O_2 \cap \cdots \cap O_n)$.

There is nothing deeper to performing Bayesian computation than repeated use of Bayes' Theorem, but with a little organization we can gain better understanding of the process and be less likely to make mistakes. First, we can build a table of the information we know:

prior	$P(A)$	$P(B)$
observation 1	$P(O_1\|A)$	$P(O_1\|B)$
observation 2	$P(O_2\|A \cap O_1)$	$P(O_2\|B \cap O_1)$
observation 3	$P(O_3\|A \cap O_1 \cap O_2)$	$P(O_3\|B \cap O_1 \cap O_2)$

Because we will be interested in the event that all of the observations are jointly true $O_1 \cap O_2 \cap O_3$, rather than only one or another of them being true, we have to use the probability of each observation conditioned on the previous ones also happening. If our observations are independent of each other then this is unnecessary, as $P(O_3|A \cap O_1 \cap O_2) = P(O_3|A)$ in this case.

Next, let us use the definition of conditional probability:

[16]The symbol $P(X|Y)$, read "(the conditional) probability of X given Y" is defined as the ratio $P(X \cap Y)/P(Y)$.

$$
\begin{aligned}
P(A) &= P(A) \\
P(A \cap O_1) &= P(A) \cdot P(O_1|A) \\
P(A \cap O_1 \cap O_2) &= P(A \cap O_1) \cdot P(O_2|A \cap O_1) \\
&= P(A) \cdot P(O_1|A) \cdot P(O_2|A \cap O_1) \\
P(A \cap O_1 \cap O_2 \cap O_3) &= P(A \cap O_1 \cap O_2) \cdot P(O_3|A \cap O_1 \cap O_2) \\
&= P(A) \cdot P(O_1|A) \cdot P(O_2|A \cap O_1) \cdot P(O_3|A \cap O_1 \cap O_2)
\end{aligned}
$$

Thus, the cumulative products we get by just multiplying down the columns of the previous table are the joint probabilities:

-	1	1
prior	$P(A)$	$P(B)$
observation 1	$P(A \cap O_1)$	$P(B \cap O_1)$
observation 2	$P(A \cap O_1 \cap O_2)$	$P(B \cap O_1 \cap O_2)$
observation 3	$P(A \cap O_1 \cap O_2 \cap O_3)$	$P(B \cap O_1 \cap O_2 \cap O_3)$

(Note that mathematically there is nothing distinguishing the prior probability from any of our observations; the choice of what information to call "prior" versus "observation" is just a matter of convention. You can think of prior as the "null" or trivial observation: how much you should update your probabilities based on not making any observations at all. In the calculations by Rootclaim, they treat the prior slightly differently from the other observations, resulting in needing to re-normalize the probabilities at the end; of course this does not change the conclusions, only the order things are calculated in.)

We should interpret each row as telling us the *relative* probability of either of the hypotheticals A, B jointly with the cumulative observations. Initially, before making any observations, the prior probabilities $P(A)$ and $P(B)$ sum to 1 (because $B = \neg A$) but as we apply successive observations the sum of each row will decrease and become smaller than 1. This residue probability (ie, the amount by which each row sums to less than 1) represents the chance of some hypothetical alternative in which at least one of the observations *didn't* happen.

Ok, what good has this done us? Well, our goal is to find the conditional probability $P(A|O_1 \cap \cdots \cap O_n)$. This is just equal to the fraction of the nth row that is in the first column:

$$P(A|O) = \frac{P(A \cap O)}{P(O)} = \frac{P(A \cap O)}{P(A \cap O) + P(B \cap O)}.$$

Indeed, all that matters is the *ratio* of the two columns:

$$P(A|O) = \frac{1}{1 + P(B \cap O)/P(A \cap O)}$$

Since the columns of the second table were found by just multiplying the columns of the first table, the *ratio* of the columns of the second table are just the product of the *ratio* of the columns of the first table:

$$\frac{P(A \cap O_1 \cap O_2 \cap O_3)}{P(B \cap O_1 \cap O_2 \cap O_3)} = \frac{P(A)}{P(B)} \cdot \frac{P(O_1|A)}{P(O_2|B)} \cdot \frac{P(O_2|A \cap O_1)}{P(O_2|B \cap O_1)} \cdot \frac{P(O_3|A \cap O_1 \cap O_2)}{P(O_3|B \cap O_1 \cap O_2)}$$

Thus we began with eight[17] pieces of data in the first table but it turns out that all we needed was four pieces of data, supposing we can directly measure these ratios.

Indeed frequently the ratio $P(O_1|A)/P(O_1|B)$, called the *Bayes factor*, is easier to measure than either $P(O_1|A)$ or $P(O_1|B)$ separately; in messy, real-world scenarios the probability of the event O_1 might be wrapped up with many uncertain factors that have nothing to do with either A or B. Estimating $P(O_1|A)$ requires assessing these irrelevant factors, but estimating the Bayes factor does not.

In extremes where probabilities get very close to 0 or 1 we can simplify matters further by taking logarithms everywhere – why multiply when instead you can add? We then have logarithmic Bayes factors

$$I_1 = \log \frac{P(O_1|A)}{P(O_1|B)}$$

where I_1 is the *information* contained in the first observation, with positive numbers informing us in favor of event A, and negative numbers informing us against it. If the logarithm is base 2, then I_1 has units of bits. Adding up each of our pieces of information gives

$$\log \frac{P(A \cap O_1 \cap \cdots \cap O_n)}{P(B \cap O_1 \cap \cdots \cap O_n)} = I_1 + \cdots + I_n$$

(Given a new piece of information I_{n+1}, we can simply add it to the sum we have so far; this is called *Bayesian updating*.)

This result is also called the (conditional) *log-odds* of A.[18] Log-odds can in some situations be more intuitive than ordinary probability, especially for extreme probabilities. A log-odds of 0 means a probability of 50%, and positive log-odds means an event that is more likely to happen than not. From the (conditional) log-odds of A we can compute the ordinary (conditional) probability:

$$P(A|O_1 \cap \cdots \cap O_n) = \frac{1}{1 + \exp(-(I_1 + \cdots + I_n))}.$$

How does this work in practice? Suppose some event of interest A has some probability of being true; start by computing the unconditional (ie, prior) log-odds of A. We make a series of observations, and assess for each observation how much information it provides in

[17] Actually seven, because we knew $P(A) + P(B) = 1$, so the first row only had one data point.

[18] The log-odds of A is defined as $\log(P(A)/P(\neg A))$, but here we have conditioned on the observations $O_1 \cap \cdots \cap O_n$.

favor of A versus against it. We update our log-odds of A by adding to it the information (positive or negative) from each observation. The result is the updated (ie, conditional on the observations) log-odds for A. Alternatively, if we don't want to work with logarithms, instead of adding up log-odds we can directly multiply probabilities.

Let us work an example. Suppose you are hiring an engineer, and you want to know their competency A at a particular skill. You make three independent observations: they did adequately on an interview assessing that skill, they have an obscure certificate for that skill, and they went to University of Example which has a good engineering program. Most candidates, whether competent or not, do not have that certificate nor went to that university, so it is hard to assess the probability of those observations; but the *relative* probabilities, or Bayes factors, are easier to guess at. We have

-	$P(O_i\|A)$	$P(O_i\|\neg A)$	Bayes	log-odds
prior	0.1	0.9	0.111	-2.2
interview	0.4	0.1	4	1.39
certificate	-	-	1.2	0.18
UoE grad	-	-	2	0.69
total	-	-	1.0667	0.065

Adding up the last column we get a log-odds of 0.065, or a probability of 51.6%, just barely above even odds that the candidate has this particular skill.

This example also illustrates a second important principle to understand with Bayesian reasoning: it can applied to any situation, and always gives an answer, regardless of how appropriate the technique is for the application. I certainly hope no one involved in hiring candidates is using a calculation of this nature to help make that decision, or at least not with the same lack of care as I did above. One must pay close attention to potential problems: did you account for all the available evidence? how accurate are your probabilities? how robust is your result to changes in the data? The messier and more "real-world" your situation, the easier it is to run afoul.

2.1 Hypothesis testing and p-values

Bayes factors have a simple and direct relationship with p-values; understanding this relationship may be beneficial, as the reader may already be familiar with the latter due to their wide use in hypothesis testing, or conversely the reader may find Bayes factors more intuitive and use it to gain comprehension for p-values.[19]

In hypothesis testing, one considers a *null hypothesis* A (or, more commonly, H_0, but we will mimic the notation from above) and the *alternate hypothesis* $B = \neg A$ (usually denoted H_a). We ask, what is the probability of an observation (or collection of observations) O, conditional on A? Except, now O is not a binary event, which happens or not, but some kind of numeric data.

[19]Someday I will more clearly explain the relationship between Bayesian computation and hypothesis testing than I did here, so the reader may want to skip this section for now.

Suppose, if A is true, that O is a real number drawn from some known probability distribution (ie this is called a *random variable*). We will then, for every possible value x of O, assign a corresponding value p which lies in the range $[0, 1]$, called the p-value. How to do this is the main source of confusion in hypothesis testing, as the process is subjective. A common choice is the map:

$$x \mapsto P(O \leq x)$$

In words, this says for one possible value of x we assign it the p-value that is the probability O would have been smaller than or equal to x. Say, for example, O is the number of heads in 100 coin flips, and our null hypothesis is that the coin is fair. If we observe $x = 40$ coin flips, then the probability of observing 40 or fewer heads would have been 0.02844, which would be the p-value for the above choice of map.

So how do we choose the map? What we do is *rank* the possible values of O by how anomalous they are under the null hypothesis A; or equivalently, by how well they support the alternative hypothesis B over A. For the 100 coin flips, the further the result is from 50, the more anomalous; this is an example of a *two-tailed* test, because the two ends of the hypothesis are the most notable and have a small p-value, while the middle of the distribution is expected and has a large p-value. The mapping given above is a one-tailed test; one end of the range of values of O (here, the low end) is interpreted as being the most anomalous, and the other end is interpreted as being the most expected.

For our coin flip test, a two-tailed[20] test would have been more appropriate; 60 or more heads would have been just as anomalous as 40 or fewer heads, so that $x = 40$ would have had a p-value of $P(O \leq 40 \text{ or } O \geq 60) = 0.05689$. A one-tailed test would have been appropriate if we had, prior to collecting observations, expected or been interested in a coin that was biased for one side in particular. If we were trying to find coins that were biased in favor of heads, say, we would use the p-value $P(O \geq x)$ and the observation $x = 40$ heads would not have been anomalous at all ($p = 0.97156$) in favor of the alternative hypothesis over the null hypothesis.

In summary, the two things we need the mapping to satisfy are as follows:

- Assuming the null hypothesis, the result p should be a random variable uniformly distributed in the range $[0, 1]$.
- Assuming the alternate hypothesis, we want p to be as small as possible.

If we observe $p = 0.001$ then this is very surprising under the null hypothesis – a 1 in 1000 coincidence! The second property of the mapping gives us the best chance of observing these surprisingly small values of p if, in fact, the null hypothesis isn't true: this increases the "statistical power" of the test.

Our reasoning process therefore looks as follows. If the null hypothesis is true, then a very small value of p is unlikely, so if we observe a small value of p we "reject" the null hypothesis. If we do not observe a small value of p, we fail to reject the null hypothesis, and do not reach any particular conclusion.

[20] pun not intended

There are two key differences between hypothesis testing and Bayesian reasoning: with hypothesis testing, we do not assume any prior probabilities for the two hypotheses – they are not treated as random events at all, but as simply unknown prepositions; and we do not calculate a distribution for the observable O conditioned on the alternative hypothesis B. If we have access to those two additional pieces of data, then we should perform a Bayesian calculation, and we would get a stronger conclusion of the form of the probability of A or B given O, instead of a weaker conclusion of "reject" or "fail to reject" A with some confidence level p.

Hypothesis testing is used when these extra pieces of data are unavailable or highly speculative. For our coin flips, we don't have any prior expectation of the probability of a coin being biased, and we can't calculate the probability of observing a certain number of heads conditioned on the coin being unfair (since the alternative hypothesis is for *any* bias, not a particular amount of bias). The p-value obtained from hypothesis testing cannot be turned into a probability for a hypothesis without adding this information.

Frequently, when working with complicated systems, it makes sense to focus on a single observation and yield just a p-value $P(O_i|A)$ or Bayes factor $P(O_i|A)/P(O_i|\neg A)$ for that observation, and leave the calculation of $P(A|O_1 \cap \cdots \cap O_n)$ to secondary analysis. Otherwise, when combining observations from multiple primary analyses it would be necessary to undo all their calculations to sort out conflicting priors and other choices. For this reason primary research usually stops at computing a p-value.

3 Lies, damned lies, and how to get large numbers by taking powers of 2

> *In the space of one hundred and seventy-six years the Lower Mississippi has shortened itself two hundred and forty-two miles. That is an average of a trifle over one mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oolitic Silurian Period, just a million years ago next November, the Lower Mississippi River was upwards of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing-rod. And by the same token any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three-quarters long, and Cairo and New Orleans will have joined their streets together, and be plodding comfortably along under a single mayor and a mutual board of aldermen. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.* –Mark Twain [31]

Suppose you suspect something about your senator is quite odd, and perhaps they are not human at all, but rather secretly a lizard from space. To test this, you collect 100 pieces of data that inform the question. For many of these data points you know what to expect for humans: how tall are they, how much do they weigh, shapes of facial features, thickness of

fingers. Of course they would not make their lizardness so obvious, so you collect data on vocal patterns, word choices, oddities in their education history, interviews with supposed personal friends. You collect reports on UFOs near their presumed arrival time, look for military flight patterns, cryptic comments made by public officials around that time. Soon, you have 100 independent observations that bear on the question.

Predictably, even if you have been completely fair in your data collection – all 100 data points are accurate, and you did not specifically seek out information that you know will be odd[21] – if you are naive with your analysis in certain ways you will invariably conclude that they are a lizard.

How? Well, if you use hypothesis testing, you might calculate a p-value for each of your 100 observations. Shockingly, you discover their ring finger is extremely long – at $p = 0.01$ it is highly anomalous, a 1 in 100 coincidence. Clearly a lizard! This is a common trap (usually inadvertent) called *p-hacking*.

Bayesian reasoning is capable of exactly the same error. For each of these observations you compute a Bayes factor. Half of them are less anomalous than average (Bayes factor 1) and half of them are more anomalous than average. They are only weakly anomalous – a 1 in 2 coincidence each, so Bayes factor 2 – but with 50 such pieces of weak evidence your total Bayes factor is 2^{50}, totally overwhelming any prior you had against secret space lizards.

However let us be a little less reckless; after all, all of those less anomalous observations should have had a Bayes factor smaller than 1 because they are evidence for humanity. Recall[22] we can convert Bayes factors to probabilities as:

$$p = \frac{1}{1+b}$$

so for a p-value p we can correspond the Bayes factor $b = 1/p - 1$. If our observations are chosen so that p is uniformly distributed in the range $[0, 1]$, then b will be distributed $[0, \infty]$.

As we calculate the total log-odds by adding up each of the 100 random values of $\log b$, we undergo a random walk. The median step size $|S| = |\log b|$ of this random walk is $\pm \log 3 = \pm 1.1$; the average is $\pm \log 4 = \pm 1.39$; and the variance σ^2 is

$$\mathrm{Var}(S) = E(S^2) - E(S)^2 = E(S^2) = \int_0^1 (\log b)^2 \, dp = \frac{\pi^2}{3} = 3.28987$$

so the standard deviation is $\sigma = 1.8138$.

After 100 random steps, our final position has a variance of $100\sigma^2$ or standard deviation of $10\sigma = 18.138$. A Bayes factor of $e^{18.138}$ gives a probability indistinguishable from certainty, so we will most likely find strong evidence either in favor of lizardness or humanness.

[21] even though you only chose to start this investigation because you already had made some specific observations that are odd!

[22] I am eliding some details and assumptions here; basically, I have assumed that any evidence against humanity is for space lizards, as this maximizes the statistical power of the test. I also have a nagging suspicion that I have omitted a factor of 2 in converting to Bayes factors, but the end result appears not to be missing any factors of 2.

(Digression: for convenience to perform this calculation I have assumed all our observations are numerical with known distribution. If so, the appropriate statistical test would be a z-test (similar to a t-test) on the mean of the z scores of the observations, and not to do any kind of Bayesian calculation at all. More likely, some of our observations will not be numerical or come from unknown distributions, so that a z-test cannot be performed.)

However, what happens if we have a slight bias? Indeed even if all of our data points are individually fair, we might be biased about *which* data points we collect.

Suppose we conveniently overlook half the data in favor of humanity; so we have two thirds for lizards, and one thirds for humans. This may still look like we are being extremely fair – after all, we've included 33 data points *against* our desired conclusion! However the difference is massive, because random walks add noise that scales with the square root of the number of steps, but a *biased* random walk drifts at a rate that is linear with the number of steps. The total bias is $\frac{100}{3} \log 4 = 46.21$, completely overwhelming any random noise (which will also be slightly less than before, I didn't calculate how much) and leading us consistently to the conclusion in favor of lizardness with a confidence of about e^{46} to 1.

Of course, such bias in the data *could* also exist because lizardness is the right conclusion, and the data is accurately reflecting that! But the analysis of any one data point, or potentially all 100 data points, is insufficient to distinguish between a completely fair Bayesian analysis revealing an anomaly or a slightly biased Bayesian analysis with no true signal.

(What's more, we can get qualitatively the same result using far fewer than 100 data points; all we need is 100 *potential* data points. If there are 100 different ways we could make an observation, and we ignore the 90 least interesting ones (with Bayes factors closest to 1), they are the ones that make the smallest contribution to the final result, and most will cancel with each other. Thus by making only 10 interesting observations we get more-or-less the same result as if we had made 100, though with a bit more randomness.)

If instead of a haphazard collection of observations one makes a collection of evidence that is in some way *canonical* it is much harder to introduce bias; say, one considers all fingers, but only fingers.

Alternatively, while we are musing about space lizards, consider the Drake equation, which attempts to put numbers to the Fermi paradox.[23] While people may (greatly) disagree about the specific numbers to use for the terms of Drake equation, and vary slightly on exactly how the Drake equation is broken down into factors, the general structure of the equation is uncontroversial, as each factor represents an obligate step in the path from lifeless rock to observable intelligent life.

Similarly, if a hypothesis can be broken down into subparts, each necessary for the hypothesis as a whole to hold, then there is less potential for bias in the selection of evidence, though there may remain disagreement about the numbers assigned to the evidence or what breakdown into subparts is cleanest.

[23] If each star is a potential host for intelligent alien life, and we can observe billions of stars, why do we not observe any alien life?

3.1 "Animal" vs "Other"

As an example of a Bayesian calculation that fails to persuade me, let us look at the calculation presented by Rootclaim in session 1 to contrast two hypotheses labeled "Animal" and "Other". ([1] at 3:05:45)

The goal of their argument is to show that, even if one conditioned on the zoonotic spillover being true, it is significantly more likely that the spillover occurred outside of HSM than inside HSM. If so, then the proximity of the HSM SSE to animal sources of virus in the market is probably a coincidence, and provides very weak evidence in favor of zoonosis.

The argument is not cleanly presented; all of the numbers shown are being used as Bayes factors, but most are written with percentages, and verbally described as probabilities instead. The events as described are inconsistent: I will try to find the best interpretation of the information given.

Formally, let An be the event that the pandemic started in Wuhan from zoonotic spillover and the first (detected) SSE is *at the location of zoonotic spillover*,[24] and let Ot be the event that the pandemic started in Wuhan from zoonotic spillover and the first (detected) SSE is not at the location of zoonotic spillover.[25]

We wish to compare $P(An)$ to $P(Ot)$. To do this, they make a series of observations O_i and compute $P(O_i|An)$ and $P(O_i|Ot)$; these ratios are the Bayes factors. Their numbers, with abbreviated descriptions of the observations, are

[24] We are very specifically *not* conditioning on that place being HSM.

[25] or we could remove the condition on zoonosis for Ot, it does not play much role in this part of the calculation, and let Ot be any alternative to An

evidence	$P(O_i\|An)$	$P(O_i\|Ot)$	Bayes	(consv)
first SSE at HSM	0.15 to 0.30	1	6.67	3.33
no animal vendors infected	0.15 to 0.25	1	6.67	4
test results unrelated to wildlife	0.4 to 0.6	1	2.5	1.67
B dominant at HSM	0.2 to 0.3	1	5	3.33
WHO cases centered at HSM, other signals not	0.5 to 1	1	2	1
early cases associated with HSM, but not strongly	0.3 to 0.5	1	3.33	2
mahjong	0.3 to 0.4	1	3.33	2.5
infected person goes to HSM	1	0.001 to 0.0005	0.001	0.0005
HSM ideal superspreader location	-	-	10	4
more likely to detect SSE at markets	-	-	4	2
total	-	-	493.8	1.48

Each probability was given as a range, indicating a value they think is reasonable and a more conservative estimate (ie, more favorable to An). The column "Bayes" gives the Bayes factors based on the former numbers; the "(consv)" column gives the more conservative Bayes factors.

Let us walk through each of their observations – most will be discussed in greater detail later.

- "first SSE at HSM": They estimate that 15% of contact with potential animal-sources of pandemics is at HSM. As for the converse probability conditioned on Ot, they are not truly claiming it is 1, but rather separated out evidence awkwardly into separate factors, which are considered below.

  During the presentation, they framed this discussion to consider the probability that the SSE at HSM is the first *observed* SSE, not necessarily the first and only that existed at that time, so in part of this calculation they are taking into account observational bias; however, later Saar would concede the claim by Peter that the HSM SSE is the first one. If we accept this concession then several adjustments must be made.

- "no animal vendors infected": None of the *known* infections at HSM are among animal vendors. To my knowledge, no testing of animal vendors was conducted; we only have confirmed illnesses among people who were hospitalized. If sars-cov-2 spilled over only a single time (or perhaps twice; see section 5.1), it is very likely no infections among

animal vendors would be detected anyhow. I have no information about how the testing of people at HSM compares with the testing processes during SARS1.

- "test results unrelated to wildlife": This is an important area of discussion that will be considered later.
- "B dominant at HSM": Most sequenced illnesses linked to HSM were lineage B, so under *An* it remains to be explained where lineage A came from.
- "WHO cases centered at HSM, other signals not": That is to say, the claim is that the actual distribution of early cases is not centered on HSM, and that the appearance that they are centered on HSM is due to observational bias. This should be ignored following the presumption that HSM is the first SSE.
- "early cases associated with HSM, but not strongly": This is asking how to explain the existence of early cases not associated with HSM (I think the slide has a typo?). I think most likely such cases do not exist.
- "mahjong": It is a little unclear here what is in need of explanation when conditioned on *An*.
- "infected person goes to HSM": If the pandemic started far from HSM, then someone infected must travel to HSM for an SSE to occur there.
- "HSM ideal superspreader location": While an infected person might go to many different possible locations, not all will become SSEs; they estimate that HSM is 10 times more likely to become an SSE than the average location. This will prove an important point of discussion going forwards.
- "more likely to detect SSE at markets": Again, this should be ignored following the presumption that HSM is the first SSE.

The goal of Rootclaim is to get a Bayes factor comfortably greater than 1: if so, then even conditioning on zoonotic spillover, the hypothesis *Ot* better explains the observations of early case distribution than the hypothesis *An*. From then on, Rootclaim can argue against *Ot* rather than *An*; or equivalently, if the evidence shown here collectively has a Bayes factor greater than 1, then throwing this evidence out only favors zoonotic in future analysis.

I do not think Rootclaim succeeded at that subgoal: just removing the two factors that were conceded, the "conservative" Bayes factor already passes below 1. If you also remove the two factors I found most dubious ("early cases associated with HSM" and "mahjong"), even their central estimate for the Bayes factor falls to 5.555, not comfortably above 1.

Indeed I think most of the probabilities in this table are too *high*, and not by the same amount when conditioning on *An* versus *Ot*; I feel that in this analysis Rootclaim is consistently overestimating the probability of rare events.[26] I could be persuaded with difficulty that the

[26]In questioning ([2] at 1:15:00), it was clarified that this is to some extent deliberate: it is not the probability of this particular observation, but rather any observation akin to this. It also was implied that these should be regarded as Bayes factors rather than the probabilities, which explains why both sides are too high, as for a Bayes factor only the ratio matters. However this was not presented clearly, and I am left

probability under *An* that no known animal vendors became sick was below the 15% that Rootclaim presented; but certainly that probability under *Ot* is well below 100%. Overall, I think most of the Bayes factors given by Rootclaim are more favorable to *Ot* than is justified by the observations.

Furthermore I do not see how to make the numbers Rootclaim gives here consistent with their estimate that 10% of coronavirus pandemics that start in Wuhan would have their first SSE in HSM (conditioned on the location of the SSE being unrelated to the location of the primary case) [7]; this scenario corresponds to the *Ot* hypothesis, which here appears to be at most $0.1\% \cdot 10 = 1\%$, a factor of 10 lower.

However, even if all the numbers shown here are completely correct, I am not persuaded by the conclusion. The listed observations do *not* represent a canonical series of necessary events that must occur under the hypothesis *An* to get an SSE at HSM. If they did, one could argue that the choice of observations was natural or unmotivated. Rather, the listed observations appear to be a miscellany of loosely related things that happened in the context of the HSM outbreak and seemed vaguely odd. I expect that in any rare event there will be a variety of strange things that happen, simply because when many things happen inevitably some of them will be strange by chance.

This calculation feels like motivated reasoning: the answer was decided, and evidence was sought until it aligned with that conclusion. Of course I do not think Rootclaim deliberately did such a backwards method – I believe their process was completely honest and designed to be fair to both hypotheses – but (as they are well aware) human biases are hard to stamp out, and Bayesian calculations are not immune to being influenced by human biases. That half of the evidence found was mooted or (in my estimation) unsupportive reflects poorly on the argument; really, from the example above of the space lizard, one should expect hardly any bias needed to get far more decisive a result without using flawed evidence at all.

In any case, the result is an argument that failed to convince me; I do not find most of the numbers credible, and I do not find the process to combine them into a single Bayes factor credible.

3.2 Rootclaim's Bayesian calculation

Let us take a look at Rootclaim's other Bayesian calculation, which was their main analysis. I have reconstructed the numbers below; for each term they also gave low and high estimates for the probabilities, which I have omitted for space. Here the prior was taken to be 0.2321, which has a Bayes factor of 0.2321/0.7679.

in doubt of how carefully these estimates and calculations were performed. In any case I tried to take the most reasonable interpretation of what the intended meaning of the presented numbers are.

evidence	Bayes	log-odds
prior	0.302 25	−1.2
location of first SSE	13.48	2.6
adjust for BSL-2	4	1.39
FCS	2	0.69
Human ACE-2	5	1.61
N-glycan	1.5	0.41
12 nucleotide insert	50	3.91
`CGGCGG`	10.683 76	2.37
`PRRAR`, out of frame	0.4	−0.92
no contact tracing	2	0.69
no genome sharing	2	0.69
RATG13, ignore FCS	2	0.69
no whistleblower	0.5	−0.69
no intelligence leaks	0.7	−0.36
no published backbone	0.4	−0.92
no known WIV infections	1	0
no intermediate host	4	1.39
total	234 015	12.363 14

The positive log-odds terms (ie, those pieces of evidence which favor the lab leak theory) have a total weight of 16.44. Half of that weight comes from three pieces of evidence: the 12nct insert, `CGGCGG` , and the location of the first SSE (sum 8.882). These three pieces of evidence are claimed to be strong and deserve commensurate analysis. The *other* half comes from eight different pieces of evidence that are individually weak and appear ad hoc. Many of these individual terms will be discussed in detail later.

As per the argument in the previous subsection, that evidence is excluded; but if you believe the previous subsection should have a Bayes factor below 1, then it should be multiplied into the total for the main Bayes calculation.

This argument is considerably less bad than that given in the previous subsection. Nonetheless it bears some resemblance to the space lizard argument, and observations seem to be selected primarily on the basis of how much they stand out rather than any canonical or complete collection of evidence. This is fine if the observations really stand out – a factor of 50 is not bad, and you will be hard pressed to turn up many factors of 50 by chance. But Rootclaim includes five factors of 2 or less in their favor, and these should be plentiful to find for any hypothesis, true or false. (Though Rootclaim did intend some of these factors as being quite conservative.)

The numbers selected by Rootclaim also lead to some odd conclusions. For example, the lack of known intermediate host is given more weight than the prior; does Rootclaim believe more than half of pandemics where we have failed to identify an intermediate host are lab leaks? Similarly, Rootclaim evaluates 80% of pandemics that first appear in Wuhan to be lab leaks. It is a little difficult to understand what the "adjust for BSL-2" factor means in

isolation, but it seems to have the effect that half of all pandemics with highly contagious viruses for which BSL-2 is inadequate would also be lab leaks.

Overall, I found Rootclaim's Bayesian calculation sufficient to be intriguing but not sufficient to compel a conclusion.

3.3 Is Bayesian reasoning suited for this problem?

Outside of simple toy problems that do not reflect the complexities of the real world, Bayesian reasoning is best suited for dealing with a small number of strong but conflicting pieces of evidence. (Of course, if they were not conflicting, one wouldn't need to do any calculations at all.) Each additional multiplicand introduces a great deal of error in the final result, both from the uncertainty of its value, but also from the potential sources of bias. Including another factor of 2 has the potential to do more harm than good if you cannot be highly confident of controlling the error.

Ideally, one would assemble a list of evidence that is either canonical in some way, pre-registered (ie you decide which evidence to include before you know which hypothesis it favors), and/or extremely persuasive (has a large Bayes factor that is unlikely to arise from chance). Most likely, any naively-computed Bayes factors would have to be renormalized towards 1 to account for the steel-manning of both hypotheses; as Saar correctly observed, if you have any Bayes factors of 1000 and you are not doing a physics experiment then your calculation is probably flawed.[27]

Rootclaim's mission is to improve upon brittle intuitive reasoning with the use of careful Bayesian calculations; certainly a worthy goal and worthwhile method. However I take issue with their labeling more traditional methods – what I would call *deductive reasoning* – as "human inference", as the human component of the reasoning method has not been removed in their approach.

During the debates I asked Saar ([2] at 1:17:55) if he could give some justification for why one expects Bayesian reasoning to give the right answer, either in general or for this specific application. Part of justifying a method would be to identify the potential flaws of the method, so that one can know how to avoid them. Certainly Rootclaim are no fools, and are aware of the importance of careful analysis; they make a point of how their conclusions are robust against the possibility that one or more of their observations have errors. However their conclusions remain brittle against the possibility of small, consistent bias; a factor of 2 too generous applied to 20 observations yields million to one odds.

The goal of my calculations at the beginning of this section is to serve as a partial answer to my question. By understanding the potential ways that the Bayesian calculation can reach a flawed answer from valid data, we can accomplish two goals: we can reject Bayesian

[27] Say you make some real-world observation that appears to be a 1-in-1000 coincidence; what is the best explanation for that observation? The odds of a measurement error, or that you failed to understand the situation correctly, could easily be greater than 1-in-1000. If any particular observation carries a 95% chance that you fully understand it and have made no error, then the greatest Bayes factor you can get out of any single piece of evidence is 20.

arguments that are vulnerable to these flaws, and gain confidence in Bayesian arguments that do not.

To reduce errors in Bayesian reasoning, I suggest three (four) things:

1. Reduce one's bias. The best way to win more basketball games is to score more points – facile and useless advice, but true.
2. Focus on only the most extreme observations: it is harder to introduce a bias with 3 factors of 100 each than 20 factors of 2 each. Of course, this requires having extreme observations available.
3. Alternatively, use some canonical or clearly unmotivated selection of evidence. Rootclaim attempts to do this by using *all* the evidence, which certainly would be a canonical selection if it were possible; in practice I believe Rootclaim is aiming to include all the *relevant* evidence, which is much harder to do without human bias interfering.
4. Conservatively add a bias towards the zero point. If the bias you deliberately add against your side is larger than whatever bias you inadvertently add towards your side, then your result will be conservative. Rootclaim explicitly adds such a bias already.

Rootclaim is certainly aware of the potential for errors in Bayesian reasoning, and already tries to follow most of my suggestions above. And perhaps I am not the right person to be giving advice here; I do not trust myself to use a Bayesian calculation akin to Rootclaim's for this application.

4 Location, location, location

You've just got home from a long day of hugging animals and taste-testing lab samples when you realize you're feeling a bit under the weather. Congratulations: you are the primary case[28] in a new global pandemic! What happens next?

Most people in your position do not pass on their illness to others,[29] but to result in a global pandemic you would have to pass it on to one or more of your close contacts. Depending on your activities, you could easily have dozens or hundreds of potential transmissible contacts over the course of your illness; a few will be household members or coworkers, but the great majority often being strangers encountered in public. Of course, such contacts are not equally likely to acquire your illness.

To simplify matters, let us say you, and everyone else, have 100 potential contacts that do not vary with time and are equally likely to become ill.[30] We can build a massive graph whose vertices are all the people in the world and edges represent close contacts. On this

[28] I originally erroneously used the ambiguous term "patient zero" here. Fun fact: the term "patient zero" arose as a misreading of "patient O[ut of California]", referring to an early known case of HIV not in California.

[29] or, at least the chain of transmission usually stops relatively soon

[30] If you wish to account for the fact that contacts are not equally likely to become ill, everywhere below replace each mention of "an individual has 100 potential contacts" with "the probability distribution of which contact(s) will become ill has entropy log(100)".

graph we can track the path of a chain of disease transmission, though of course for the disease to grow to become a pandemic there will be more than one transmission chain going on at the same time.

Each successive transmission, as you pass it on to one of your close contacts, who passes it to one of theirs, acts as a random walk over the graph of close contacts. After some interval of time, the disease is discovered in the *index case* (ie, the first observed case). What we want to know is, for a particular hypothetical primary case, what is the probability distribution of who will be the index case?

4.1 A model for evaluating location of the index case

To estimate this, we will describe a crude model; while we do not know the quantitative details of the close-contact graph, fortunately our model is greatly insensitive to these details. Suppose the shortest path from the primary case to the index case is n steps; call this the *epidemiological distance* between them. Consider drawing a sphere around the primary case that contains all people at most n steps away; certainly any random walk of length n that starts from the primary case will end up at one of those people in the sphere, with some nonzero probability. Our simple model is to say that the index case follows a uniform distribution within that sphere; that is, that the index case is equally likely to be any of the people at most n steps from the primary case.

How good of a model is this? Well, let us consider two examples of simple graphs that might be compared with the real-world close-contact graph.

A *Bethe lattice* of degree d is a tree[31] where each node has d neighbors; Bethe lattices are infinite dimensional.[32] If your social network were a Bethe lattice, then you would have 100 close contacts, and about 100^2 people within 2 steps, and 100^3 within 3 steps, and so on. Random walks on a Bethe lattice very rapidly go away from the starting point; after n steps you will be a maximum of n steps away from where you started, and on average be slightly more than $n(1 - 2/d)$ steps away. Our model that the random walk ends up uniformly distributed within a sphere of radius n is very good on a Bethe lattice.

While the real-world close-contact graph does have loops (eg, especially within a household or workplace) respiratory diseases easily jump to strangers who will have few shared contacts. On a very small scale, a Bethe lattice is a fair representation of the connectivity of close contacts (this results in the "small-world phenomenon"). One could also imagine that each node on the Bethe lattice represents not a person but a small collection of closely connected people, such as a household or workplace; if so, when counting distances we would ignore transmissions within the same group.

On a larger scale, around the size of a city, this Bethe lattice approximation breaks down, because disease transmission is a physical process that requires proximity. As most individuals generally stay in roughly one area on the surface of the Earth during their illness,

[31] graph with no loops

[32] A graph has dimension D if the volume of a sphere of radius r scales like r^D. On a Bethe lattice, the volume of a sphere grows exponentially with radius.

the contact graph becomes approximately two-dimensional. If we use a 2D integer lattice to represent the contact graph at this scale, we see that diffusion is much slower; for any finite dimension, a random walk of n steps will on average end up a distance of $\sim \sqrt{n}$ from its starting location. Our simple model can be somewhat salvaged by using a sphere of radius $\sqrt{n}$[33], though the model is less good than with the Bethe lattice. Of course long-distance plane travel interferes with representing the close-contact graph as a Euclidean lattice. Regardless, we will not be interested in intercity transmission, so we can focus on the small scale approximation.

Now if we have some candidate for the primary case, we want to know the probability of observing a specific person being the index case. Had we knowledge of how many steps n of cryptic transmission to expect, we would draw a circle[34] of radius n around the primary case and see how many people are included; unfortunately, we do not have good information about what to expect for n, and we do not have exact information about the contact graph.

Suppose we take a flat prior distribution for n, as we don't know what to expect. Then draw larger and larger circles around the candidate the primary case until we first include the index case; the radius of this smallest circle that includes the index case is a posterior best guess for n, and the probability we expect for that index case is the inverse of the number of people in that circle.

4.2 Applying model to zoonotic and lab-leak hypotheses

Let Z be the event that zoonotic spillover of covid took place at an east Asian wildlife market, and LL be the event that covid resulted as a lab-leak from a lab conducting gain-of-function research. The observed index case is a shrimp vendor in HSM; call this observation H. We want to estimate $P(H|Z)$ and $P(H|LL)$ using the model from the previous subsection. So, for each hypothesis, how large of a circle do we have to draw around the primary case to include the index case?

For Z, the primary case would be any wildlife vendor at an east Asian market. We do not have to draw a large circle to get from such a vendor to the index case; the smallest circle we could draw would include all residents at a market which has wildlife.[35] At HSM, there were 1162 vendors ([32] annex page 182), which we round to 1200 as that makes the numbers easier. (When including all employees and daily visitors I imagine it would go somewhat higher than that.)

But HSM is not the only market, and there are potential primary cases at other markets, so our circle must include all residents at those markets as well. In their written responses,

[33] I found that in a d-dimensional Euclidean space, the Kullback-Leibler divergence of the true probability distribution relative to our simple model is a rather small $\frac{1}{2}\log(\pi(d+2)) - 1 + o(1/d)$. This fact is not useful here, but after going to the trouble of calculating it I felt compelled to mention it somewhere!

[34] This is a circle in epidemiology space, not physically on a map.

[35] By "resident" I mean not just vendors but any employee or daily visitor to the market; to a first approximation this is synonymous with "vendor". There is of course much room for variation in the choices I am making. For example, early covid cases were concentrated on the west half of HSM, so perhaps we should draw a smaller circle that includes only *half* the market. Conversely, perhaps we should be including household members of market vendors in our circle.

Saar gives the probability of the market in question being HSM as 0.2% [7], and Peter gives 2% [8], so by those numbers our circle should include 60 to 600 thousand people.[36]

For LL, the primary case would be any laboratory worker at a lab doing gain-of-function research on coronaviruses. In Wuhan, to draw a circle large enough to include the index case requires including the whole city, as there is no known potential chain of transmission from a lab worker to a resident at HSM that is significantly shorter than what you might expect to any other person in the city. Saar guesses that 20% of such GoF research is at WIV ([1] at 2:29:30), so multiplying the population of Wuhan by 5 suggests a circle containing 60 million people.

Our naive analysis then yields

$$P(H|Z) = 1/(6 \cdot 10^4) \text{ to } 1/(6 \cdot 10^5)$$
$$P(H|LL) = 1/(6 \cdot 10^7)$$

for a Bayes factor of 100 to 1000 in favor of zoonotic spillover.

Many of the inaccuracies in our model will apply equally to both Z and LL, and thus cancel out for the purposes of calculating a Bayes factor, but not all. We have supposed that the contact graph is homogeneous, so that everyone has the same number of contacts; but asymmetries in the graph could potentially cause a random walk to be more likely to end up in some places than another.[37] Saar claims that this effect causes outbreaks to be more likely to show up at HSM than one might expect otherwise; if so, this raises the probability of $P(H|LL)$, but does not change $P(H|Z)$ because under the zoonotic hypothesis our circle is already been restricted to HSM.

We should also ask whether our analysis should be so sensitive to who the index case is: the earliest cases were all close in time, and under a slight change in events the earliest among them could easily have been someone else, plausibly even someone not at HSM. We can address this by not considering a single index case but a sort of "average" of the earliest cases; enough of them are HSM residents, or closely linked to HSM, that the result is essentially the same as having a singular unambiguous index case who is a resident at HSM. Had the earliest case data a more ambiguous nexus, this would have required more careful analysis.

There are much more serious concerns with this analysis. We have chosen Z to represent zoonotic spillover at an east Asian wildlife market; if this feels arbitrary, that's because it is. Properly the debate concerns the possibility of zoonotic spillover at any location, not necessarily a wildlife market and not necessarily in east Asia. Notably, Saar observes [7] the possibility of zoonotic spillover at a restaurant that processes live animals. If we were to expand Z to include restaurants and locations outside of east Asia, our circle would grow enormously, and $P(H|Z)$ be reduced in proportion. Conversely, since we are interested in an outbreak that we know took place in Wuhan, we could reduce Z to zoonotic spillover at

[36] If this feels a little high, keep in mind most wildlife markets are smaller than HSM, so this should probably be lowered a bit.

[37] This is related to the well-known "friendship paradox" that most people's friends have more friends on average than they do. [15]

HSM, or perhaps even down to a specific shop. Each of these choices for Z is fully consistent with the parameters of the debate; which choice of Z is "correct"?

Similarly, one could ask if LL was chosen appropriately; would we be having the same debate if covid appeared in a different city that lacked WIV but had other research laboratories engaged in potentially dangerous coronavirus research?

For example, restricting Z to HSM and LL to WIV we get

$$P(H|Z) = 1/1200$$
$$P(H|LL) = 1/(1.2 \cdot 10^7)$$

for a Bayes factor of 10000.

It is problematic for our estimates of $P(H|Z)$ and $P(H|LL)$ to vary by orders of magnitude simply by taking a different perspective on the evidence: the truth of the origin of covid should not depend on the exact wording used by each side of the debate. The resolution to this problem is that as $P(H|Z)$ varies wildly as we adjust our interpretation of the event Z, the prior probability $P(Z)$ varies wildly in the opposite direction. If we broaden Z to include all sorts of potential sources around the world, $P(Z)$ increases; if we narrow Z to a single specific resident at HSM, $P(Z)$ decreases in accord. The joint probability $P(H \cap Z)$ is in each case unchanged.[38]

Therefore any discussion of $P(H|Z)$ and $P(H|LL)$ is intrinsically linked to the more difficult discussion of the prior probabilities $P(Z)$ and $P(LL)$. We will delay further analysis on this to that time, in section 7.

4.3 Where is a person?

The distinction between *residents* at HSM (by which we mean any person who returns to HSM on a daily basis, which includes vendors, other staff, and daily shoppers) and *visitors* is at times subtle yet impactful. With estimated 1200 residents and perhaps 50000 weekly[39] unique visitors [7], these subpopulations disagree by a factor of 40; most of our analysis scales linearly with population size, so using the wrong subpopulation introduces an error of 40 times.

We can imagine making a list of all people in Wuhan, and all locations in Wuhan where that person might be. The typical person will be strongly associated with two places, their home and their work,[40] with weaker associations with many other places (eg markets they shop at). Instead of talking about the index case as being a specific person, we could refer to an "index location" where the first case *cluster* arises. Looking at transmission from location to location is much harder than person to person because locations vary wildly in how prominent they are: locations associated with more people will of course have a

[38] if we assume that zoonotic spillover didn't occur at a *different* location and then come to HSM

[39] I use weekly throughout as that is roughly the time scale over which a person is infectious with covid.

[40] This gives rise to the term "third place" to mean a location where one regularly goes to freely socialize; a characteristic feature of such a place being that one can spontaneously interact with people there without requiring advance planning.

much higher chance of having outbreaks. To a first order, we can scale the probability of an outbreak occurring at a location by the size of that location's subpopulation. Doing so may require introducing a factor of roughly 2 to account for the fact that most people are strongly associated with 2 locations. (We will frequently just ignore this factor of 2.) If we include all the weekly visitors at HSM as part of its population, then we must use a much larger adjustment factor, as the typical person is a visitor of many more than 2 locations.

While the initial outbreak at HSM includes both residents and visitors (and people who were neither, but instead indirectly linked to HSM), for most purposes what is relevant is the resident population, because the epidemiological "center" of the outbreak is amongst the residents. Throughout this analysis I will take care with which subpopulation I use, but not go into detail of my reasoning each time.

In section 4.5 we will see Rootclaim's argument that HSM was a likely place for an earlier outbreak to arise; a key part of that argument is the observation that HSM has "permanent residents". In contrast, consider for example a train: an outbreak on a train will likely be between visitors (ie passengers) who will mostly not return to the same train for further transmission. To a first order, this is appropriately accounted for by the fact that we scale the relevance of a location with the number of residents it has; for a train, the "residents" would be the train staff, which is quite a small number compared to the number of residents at HSM, so we'd expect an outbreak *among train staff* to be proportionately less likely than one among HSM residents. This expectation does a poor job of representing a super-spreading event that could happen among train passengers without involving the staff. However since the real-world HSM outbreak is unambiguously centered on HSM residents, we don't have to think too hard about how best to model such scenarios.

4.4 Early ascertainment bias

The *ascertainment rate* of covid refers to the fraction of covid cases that become known; before covid became known, the main way for people to be diagnosed is retrospectively in samples that were collected while they were in a hospital. People who became sick with covid before 2020 and recovered without going to a hospital will mostly be unknown. There is some nuance here, in that people who visit local hospital clinics are less likely to have samples amenable to retrospective analysis than those who went to a major tertiary hospital; and the possible distinction between "going to a hospital" versus "being hospitalized". We can roughly estimate the early ascertainment rate as somewhere around 5% to 10% based on the hospitalization rate of covid.

Ascertainment rate bias refers to any biases in how likely some cases are to be ascertained than others; most notably that in the first half of 2020 January having an epidemiological link to HSM was required for diagnosis. Other, smaller biases likely also existed. However I think the ascertainment rate for hospitalized cases to be quite good, and effectively zero for non-hospitalized cases; if so, that places a sharp upper bound on how much bias could exist.[41]

[41] other than any biases that influence whether patients become hospitalized at all

Rootclaim initially argued that ascertainment rate bias made covid more likely to be *detected* early at HSM even when there could be clusters in multiple locations. However they eventually mooted this point (correctly, in my opinion), which makes any ascertainment bias irrelevant to the discussion of where the first SSE took place.

4.5 Is covid attracted to HSM?

The simplistic model used in the previous subsection suggests that the early outbreak being observed at HSM gives a Bayes factor of 100 to 10000 in favor of Z over LL, but in Rootclaim's Bayesian calculation the location of the initial outbreak is counted as favoring LL by a Bayes factor of 13.48.[42] This substantial discrepancy is largely due to a single factor which we have not included so far: whether outbreaks are disproportionately likely to appear in HSM compared to other locations. Rootclaim estimated that, given that the first cluster appeared in Wuhan, the chance of it being at HSM under LL is about 10% [7]. Since residents at HSM are 0.01% of the population of Wuhan, this is equivalent to saying that residents at HSM have 1000 times the chance of being the index case than the average Wuhan resident.[43]

I do not think there is a 1000 times tendency towards HSM, but a smaller number might be justifiable, so let us more carefully consider Rootclaim's argument.

The core of their argument is the claim that HSM gives uniquely good conditions for the spread of covid between vendors. They identify several properties:

- "High traffic" ([1] at 3:12:00); ie, large number of unique weekly visitors
- "Permanent residents" ([1] at 3:12:00); ie, people repeatedly return to the location
- "Enclosed, poor ventilation" ([1] at 3:12:00)
- "Density" ([5] at 2:30:40); unclear what this refers to
- "Organic materials" ([5] at 2:14:36)
- "Temperature" ([5] at 2:14:00)
- "No UV radiation" ([5] at 2:14:00)

The idea is that lots of people visit HSM, so even when there are few sick people in Wuhan there is a decent chance of one visiting the market; then if a resident becomes sick, the resident will return to HSM to infect others; and the growing cluster of cases at HSM will outpace infections elsewhere.

At an intuitive level these factors do not seem to especially highlight HSM: restaurants, schools and universities, dorms and apartment complexes, and community centers all seem comparable to HSM in these factors. Hospitals, while *usually* well-ventilated, are often

[42] This factor comes entirely from the outbreak occurring in Wuhan, and thus proximate to the WIV, with the specific location in Wuhan being neutral between Z and LL.

[43] This is difficult to reconcile with their estimate elsewhere that HSM is 10 to 4 times more likely than a typical place to be a superspreader location ([1] at 3:14:00).

nexuses of outbreaks early in a pandemic. Households are individually too small to have a significant chance of being the first cluster, but are collectively likely due to the large number of them and the quite high risk of transmission within a household. Indeed the very first covid cluster detected was a household cluster consisting of an elderly couple and their asymptomatic son found by Dr Zhang; their connection to HSM was only identified later.

(I am also modestly skeptical of how true or relevant some of these factors are for HSM. While I don't see any obvious flaw with Rootclaim's source [10] that UV radiation is correlated with lower growth rate of covid, I am very skeptical that there is any causative relationship; in any case, it is equally relevant for any indoors location. While more plausible than UV, the *causative* role of temperature in flu transmission remains uncertain, and covid likewise. The presence of organic materials is mostly irrelevant to human-to-human transmission which is dominated by respiratory transmission, not fomite transmission.)

(As for the ventilation of HSM, based on photographs it appears to be not great but not uniquely bad; the end of each alley is open to the air, and ceilings are generally spacious. Being a semi-open-air structure I would think it is better ventilated than an enclosed building. However multiple independent first-hand accounts describe it as poorly-ventilated; I'm not sure if this is in comparison to other open-air structures or in general. A few parts of the structure, such as in the south-west, appear quite cramped and enclosed. The building had previously had a ventilation system but it was shut down. Extensive pictures and videos can be found at `http://babarlelephant.free-hoster.net/visiting-the-wuhan-seafood-market/`.)

(In section 4.3 I briefly discussed the relevance of having "permanent residents"; this warrants further consideration, but I lack the time.)

4.5.1 Rootclaim's market model

Rootclaim created a quantitative model ([5] at 2:29:40) to illustrate how an initial patient unassociated with HSM could yield the observed outbreak; the goal is to show in principle how one might achieve the WHO numbers, which were 55 HSM-linked cases and 119 non-HSM-linked cases in 2019. It is not meant as a definitive explanation, just an example. Let us walk through it:

1. The primary case occurs at WIV.
2. After 7 doublings, there are 128 cases in Wuhan out of a population of 12 million. With 10000 daily visitors at HSM, it takes on average 10 days for a sick person to visit HSM.
3. The visitor infects a resident at HSM.
4. The doubling-time of HSM-linked cases is half that of other cases. After 5 more doublings, there are 4096 cases outside HSM and 1024 cases associated with HSM.
5. 1/32 non-HSM cases are hospitalized, and 1/16 HSM cases are hospitalized, yielding 128 known non-HSM cases and 64 known HSM cases.

I find this model unrealistic in three places.

First, it requires covid to be strangely passive before reaching HSM. At 5% ascertainment, there is only a 0.14% chance none of the 128 pre-HSM cases are found; at 10% ascertainment, this falls to 0.00014% chance.[44] [45] (Indeed non-HSM cases should be observed exceeding HSM cases until the final doubling period.) Also, none of these 128 cases directly lead to super-spreading events unconnected to HSM.

Second, Saar entirely omitted step 3, which is a low probability event. A person going shopping might have face-to-face contact with a vendor for several minutes, facilitating transmission; but I expect in a typical shopping trip they will have broadly comparable levels of contact with dozens or hundreds of other people, such as other shoppers and on public transit. (For comparison, the US CDC used a minimum of 15 minutes of exposure to even be considered a "close contact"; though of course it would be foolish to trust your health to such a rule.) Either it was unlikely for the vendor to be one of the few people infected by the visitor, or the visitor was infecting dozens of people.

Third, the great disparity in growth rates between HSM-linked cases and non-HSM-linked cases after covid reaches the market is implausible, as Rootclaim ignores the key factor that HSM-linked cases become non-HSM-linked cases if they lose their epidemiological connection to HSM. There are several ways one could model this process, and which is most appropriate depends on what is known about the early case data. As just one example, let us suppose that in the ordinary doubling time of covid, a resident with covid on average infects two vendors and two visitors (this retains Rootclaim's assumption of HSM being twice as infective, which I think is too high), and that "HSM-linked" includes all infected resident and visitors but not any indirect infections.[46] Then the number of cases after each doubling period tracks as follows:

doublings	resident cases	visitor cases	connection lost	non-HSM cases
0	1	0	0	2^7
1	2	2	0	2^8
2	2^2	2^2	2^2	2^9
3	2^3	2^3	2^4	2^{10}
4	2^4	2^4	$2^5 + 2^4$	2^{10}
5	2^5	2^5	2^7	2^{11}
6	2^6	2^6	$2^8 + 2^6$	2^{12}
7	2^7	2^7	$2^9 + 2^8$	2^{13}
8	2^8	2^8	$2^{11} - 2^8$	2^{13}
9	2^9	2^9	2^{12}	2^{14}

[44] Using the 1/32 ascertainment rate Rootclaim had, it is a 1.7% chance.

[45] For any reasonable ascertainment rate, Rootclaim's model can be improved by greatly reducing the number of people to get ill before one goes to HSM: at 32 people, for example, the chance that one of them goes to HSM is reduced by a factor of 4, but this is more than made up for by lowering the chance of one of the pre-HSM cases being ascertained.

[46] Rootclaim is aiming to replicate the WHO numbers, which have only 2 indirect cases of the 55 HSM-linked cases; the other 53 were infected at HSM. See appendix B.1.

"Connection lost" refers to cases descended from the initial resident case, but whose epidemiological connection to HSM is forgotten. We are being a bit loose about distinguishing cumulative and current infections, as I am trying to imitate Rootclaim's numbers and they were also loose; for exponential growth the distinction is not important.

Note that under these assumptions (or any reasonable assumptions), the growth rate of HSM cases after the first doubling is entirely driven by the resident-to-resident infection rate, so no matter how long we continue the above table, HSM cases will never catch up to non-HSM cases. To get 2^{10} HSM cases within 5 doublings we need each resident to infect (almost) 4 other residents. Also to maintain the observed half-half split of HSM cases between residents and visitors,[47] we need each resident to infect 4 visitors – so the total infectivity rate at HSM is 4 times that elsewhere in Wuhan, when I already find the 2 times rate Rootclaim proposed unreasonable.

Using 4 and 4 we get:

doublings	resident cases	visitor cases	connection lost	non-HSM cases
0	1	0	0	2^7
1	2^2	2^2	0	2^8
2	2^4	2^4	2^3	2^9
3	2^6	2^6	$2^6 - 2^4$	2^{10}
4	2^8	2^8	$2^8 - 2^5$	2^{11}
5	2^{10}	2^{10}	$2^{10} - 2^6$	2^{12}

(Observe that even with HSM having 4 times the infectivity of non-HSM, HSM-linked cases still can never exceed 2/3 of all cases.)

There are other ways to represent the loss of epidemiological link to HSM (my first version was to assume that each transmission has a fixed probability of lost connection, but I think that is less realistic), but regardless of choices I don't think there is any believable way to get the observed case distribution from Rootclaim's model.

4.5.2 Rootclaim's market model, with timeline

Let us make a timeline of pertinent events assuming Rootclaim's model is correct:

event	observed date	total non-HSM cases	total HSM cases
primary non-HSM case	?	1	0
index non-HSM case	Dec 16	2^5	0
primary HSM case	?	2^7	1
index HSM case	Dec 10	2^9	2^4
market closure	Dec 31	2^{12}	2^{10}
lockdown	Jan 23	$2^{16.76}$	-

Here, "total cases" means cumulative. On January 23 there are around 2000 new ascertained

[47] 30 vendors, 25 non-vendors; see appendix B.1.

daily cases, and the doubling time is about 3.5 days [23], so assuming an ascertainment of 10% this gives $2000 \cdot 10/(1 - 2^{-1/3.5}) \approx 111000 \approx 2^{16.76}$ cumulative cases.[48] [49]

The timing of the index cases is based on the ascertainment rates used by Rootclaim; of course the exact timing will be highly uncertain. Nonetheless we see right away the difficulty in explaining why cases were not ascertained outside of HSM for so long; their model was calibrated for producing the right number of cases on December 31.

We get a reasonable chronology for January: 4.76 doublings in 23 days is 4.8 days doubling time, which is definitely higher than the true value, but not bad for a simplistic model that wasn't intended to be extrapolated to January. In December we have 3 doublings in 21 days for a doubling time of 7 days, which is pretty far from a reasonable value.[50] Comparing the index non-HSM case to market closure we have a doubling time of 2.14 days, which is far too short; and of course to get 4 doublings in -6 days is definitely wrong.

Rootclaim's model looks good after the HSM closure but does not line up with observations early in the market outbreak; it overestimates cases outside HSM and underestimates cases within. Assuming some randomness in the timing of the index cases and the ascertainment rates and the doubling time, we can finangle the model numbers to be kind of consistent with observations, but it clearly gets worse and worse the earlier in the outbreak one looks.

4.5.3 Empirical evidence of market outbreaks

It is a great challenge to argue from first principles that a certain environment is especially conducive to covid transmission. Recall how contentious even (comparatively) easily-measured aspects like fomite transmission or size of respiratory particles were in 2020; and now we hope to quantify the infectivity rate of being in one particular shopping center?

Much better would be to observe empirically how covid behaves; and Rootclaim attempted to do just that by introducing a list of known covid outbreaks in China centered at or involving markets ([5] at 2:21:50). Here is an edited version of their list:

[48] We are not using Rootclaim's ascertainment rate of 1/32 because the ascertainment rate has certainly risen by late January compared to December.

[49] If you feel queasy about using the geometric formula to extrapolate from daily cases to cumulative cases, that is a sensible reaction, but we can get the same results by instead converting the model's cumulative cases into daily cases before comparing with observations; since the model is exactly exponential, this conversion is exactly correct. I did not do the conversion that way because the conversion is sensitive to the doubling time, and I wanted to delay discussion of the doubling time until after the table of numbers has been made. We would also have to deal with problems like what if the doubling time varies with time.

[50] Note, however, from [21]: "In its early stages, the epidemic doubled in size every 7.4 days."

location	date (2020)	case count	market?	source of index case
Xinfadi (Beijing)	June	335	yes	frozen food
Dalian	July	144	yes	frozen food
Qingdao	October	14	NA	frozen food
Kashgar	October	428 to 430	yes	"aviation container"
Tianjin	November	10	NA	frozen food
Shanghai	November	3	NA	"aviation container"
Manzhouli	November	25	NA	human to human
Qingdao	December	2	NA	cold food
Chengdu	December	14	NA	trash
Beijing	December	3 to 41	NA	human to human
Dalian	December	69	NA	cold items
Shenyang	December	38	NA	human to human
Shijiangzhuang	Dec to 2021 Jan	1129	no	human to human
Northeast China	Dec to 2022 Jan [sic]	1650	no	human to human

(Rootclaim considered only the 5 biggest outbreaks, and did not investigate the possibility that markets could be involved in the others. Discarding the 2 and 3 person outbreaks as irrelevant seems reasonable, but the others I am not sure about.)

Consider the three market-linked outbreaks listed above. There is an immediate problem with Rootclaim's argument: *none of them started from human to human transmission.* Rootclaim needs to show that there is a high probability of an infected person going to HSM and creating an outbreak there. Frozen food causing an outbreak does not support Rootclaim's case, except in so far as it demonstrates that markets can sustain an outbreak once one is started there. Indeed it is almost inevitable to get a list of non-human-to-human outbreaks as Rootclaim's main source for building this list is subtitled "An Overview of Environment-To-Human Transmission Events" [16].

This is not as bad for Rootclaim as it appears. While the outbreak in the Xinfadi market in Beijing did appear to start from frozen salmon brought to the market, the Dalian and Kashgar outbreaks only secondarily involved a market, so the *market* index case was human to human transmission. Let us first look at Kashgar, in Figure 7.

In the Kashgar outbreak, 75 different infected individuals attended a large farm market event that had approximately 20 thousand participants (out of a population of 700 thousand in Kashgar); this lead to 20 others at the market becoming infected. Of those 20, 5 attended a wedding, leading to 36 infections at that wedding; this makes weddings appear to be a far greater risk of spreading covid than markets. Arguably the R_0 at the market is less than 1. This would be a relevant analog to the HSM outbreak if half a million people came to HSM in one day, and more people brought the illness to it than from it.

The Dalian outbreak is more favorable to Rootclaim. The only source I have is an article in China Daily [38]:

> Farm produce markets and buses have become "amplifiers" in the spread of COVID-19 during the latest outbreak in Dalian, Liaoning

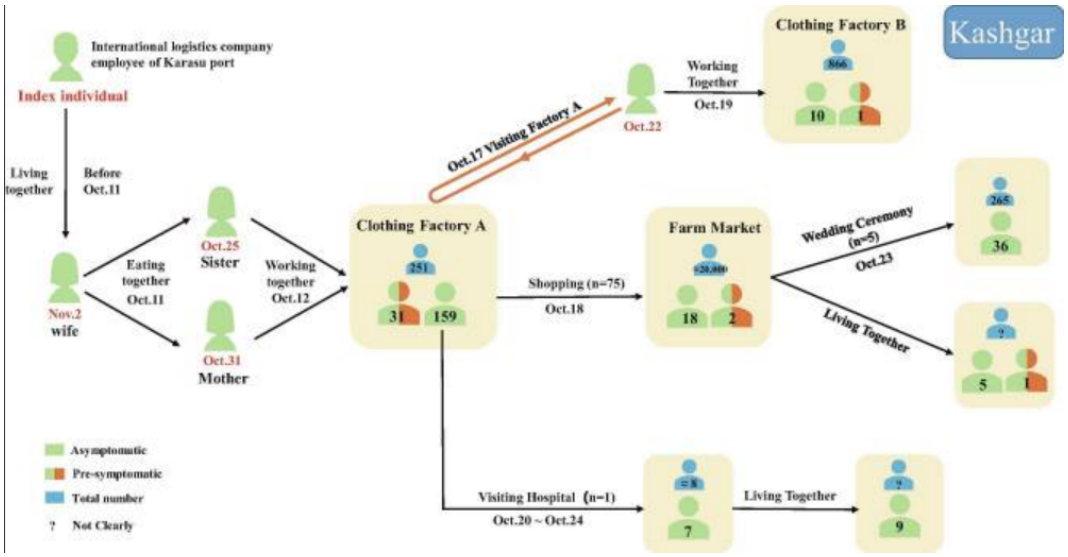

Figure 7: Outbreak in Kashgar. This is Figure 1 of [36].

province, a local disease control expert said.

On Thursday, the city reported three confirmed COVID-19 cases, including one previously reported as an asymptomatic infection, along with 112 asymptomatic carriers, the city's COVID-19 prevention and control headquarters said on Friday.

"It is quite different from previous outbreaks in our city, which reported sporadic cases. The latest one mainly influenced the downtown regions and was amplified through places like farm produce markets," Meng Jun, deputy director of the Dalian center for disease control and prevention, told Dalian Daily on Thursday.

Genetic sequencing found that the cases reported since Aug 26 were highly homologous to those found in Dalian since Aug 20. Preliminary results showed the source was accidental exposure of non-closed-loop management personnel in a centralized isolation hotel for international travelers.

Infected people, reported since Aug 20, had visited the Taoyuan market, a well-known market with fresh vegetables, meat, seafood and groceries.

"Some cases since Aug 26 are owners or customers of the market. Therefore, they belong to the same outbreak, introduced by infected people into the Taoyuan market, which then became an amplifier and spread out," Meng said.

[...] Meng said genetic sequencing had identified the virus responsible for the outbreak as a mutation of the Omicron subvariant BA.5.2.1,

which is more contagious and harder to discover.

However it is impossible to draw any conclusions from the text above, as it is far too vague.

Moreover, omicron did not exist in 2020 July; the article appears to refer to an unrelated outbreak 2 years later. There is no reason to think the 2020 July outbreak involves a market.

In sum, the empirical evidence of market outbreaks show one outbreak in 2022 August with a vague description of being amplified by "buses" and "places like farm produce markets", and the 2020 June outbreak in Xinfadi which started due to frozen food in a market. Neither is a good analogy for Rootclaim's claimed account of events in HSM. In the meantime, China had at least a dozen non-market outbreaks in 2020, and many more since then.

4.5.4 Super-spreading events are highly random

To argue that a super-spreading event is expected to occur first at HSM *and not concurrently elsewhere* it does not suffice just to show that covid is highly infectious at HSM, but also it is not especially infectious elsewhere ([4] at 1:02:10).

Covid did not go away after HSM was closed, but rather continued to create case clusters throughout Wuhan; this gives a lower limit on how infectious covid is away from HSM, and so bounds the potential disparity between HSM and non-HSM case growth rates.

Most people with covid do not pass it on to anyone, while a small minority of cases will give it to dozens or a hundred other people in single incidents; this same pattern was observed with SARS-1. Rootclaim's smooth exponential model gives an average growth rate curve, which is more accurate the more people are involved to average over. If one location has twice the infectivity of another, we expect its case count to grow twice as exponentially-fast; but when there are only thousands of cases or fewer, a single super-spreading event can completely overwhelm such a difference and drive the less infectious location to have more cases sooner.

As the location of early clusters will be driven largely by the happenstance of where such events happen to take place, the location of the first cluster will have a flatter distribution than would be expected by an analysis that considers only average growth rates at each location.

For example, if we say that instead of each person giving covid to two people, half of people give it to none and half give it to four,[51] then there is a 50% chance the HSM outbreak stops immediately.

Let us try to be a bit more precise and suppose that outside of HSM the number of people a given case infects follows a certain distribution with mean μ and variance σ^2, and that within HSM it follows a similar distribution with mean 2μ and variance $4\sigma^2$.

If we consider a protocluster of n cases at one location (outside of HSM), the number of

[51] Some people might develop symptoms quickly and stay home right away; these will not pass it to others outside the household no matter how infectious their workplace is. Also, there is enormous variation in viral load from person to person.

people infected in total by those n people has mean $n\mu$ and variance $n\sigma^2$. What is the probability that protocluster is growing faster than a same-sized protocluster at HSM?[52] Well, the difference between the two clusters follows a mean of $n\mu$ and variance $5n\sigma^2$,[53] so for that to be negative would be a z-score exceeding

$$z = \frac{\text{difference}}{\text{stddev}} = \frac{n\mu}{\sqrt{5n\sigma^2}} = \frac{\sqrt{n}\mu}{\sqrt{5}\sigma},$$

which increases slowly with n.

For a normal distribution (which this follows if n is large enough, by the Law of Large Numbers), the probability of exceeding a given z-score is the Q-function or error function. The Q-function is approximately

$$Q(z) \approx \frac{\phi(z)}{z} = \frac{e^{-z^2/2}}{\sqrt{\tau}z}$$

where $\tau = 2\pi$. Substituting for z we get a probability of

$$p \approx \frac{\sqrt{5}\sigma}{\sqrt{\tau n}\mu} \exp(-n\mu^2/10\sigma^2)$$

The ratio μ/σ could be any positive number (as $\mu > 0$ because the number of people with secondary infections cannot be negative), but if we impose the constraint that at least half of people with covid pass on no secondary infections, then it follows that $\mu \leq \sigma$ and so μ/σ has to lie in the interval from 0 to 1. Taking the extreme $\mu = \sigma$,[54] which is most favorable to Rootclaim's model, we get

$$p \approx \frac{\sqrt{5}}{\sqrt{\tau n}} \exp(-n/10).$$

Trying a few values of n, we get

n	p
5	24.2 %
10	10.4 %
20	2.7%
30	0.81%
40	0.26%
50	0.085%

This is a lower-bound on the probability that a cluster outside of HSM outcompetes a cluster within HSM of the same size for a single generation. Even if the *average* growth rate within HSM were twice as high as without HSM, large variation at small n makes

[52] which has mean $2n\mu$ and variance $4n\sigma^2$ of number of people infected next generation

[53] $E(X-Y) = E(X) - E(Y)$ and $Var(X-Y) = Var(X) + Var(Y)$ if X and Y are independent.

[54] This extreme is only attained by requiring exactly half of people have no secondary infections and the other half have exactly the same number of secondary infections.

attaining that average uncertain. Of course, HSM is not just competing against a single cluster elsewhere, but perhaps 127 such clusters under Rootclaim's model; and to produce the numbers claimed, the HSM cluster must outcompete all of them. The probability that HSM becomes the biggest of the 128 clusters is more than 1 / 128, but not much more.

4.5.5 Are other diseases drawn to HSM?

Rootclaim estimates that there was a 10% chance of the first covid outbreak appearing in HSM; their reasoning applies equally well to other respiratory diseases. We should expect, then, that in a typical winter HSM has one of the first 10 flu outbreaks in Wuhan. Comparable markets elsewhere in Wuhan or in other cities should similarly be early nexuses of flus, colds, and other diseases.

Is this true? I have no direct evidence either for or against this, but I strongly suspect that if it were true someone would have noticed by now. China, like the US, invests in flu surveillance and has a significant motive to identify potential infection sources or focuses.

Saar contests that the factors favoring HSM are unique to when there are very few cases present in a city, as the high rate of visitor traffic becomes less relevant when the disease is abundant everywhere ([4] at 20:40). However disease transmission should remain linear at low case counts, and nonlinearities only begin when some subpopulation has so many cases as to create significant immunity. If the number of infected people in a city is doubled, the rate of outbreaks at every location is doubled, and the chance that a certain location is first remains unchanged.

(However people have cross-year immunity to the flu and other endemic diseases, which will slightly dampen inter-location variation in infectivity.)

4.5.6 Assessment of Rootclaim's market model

Rootclaim explains the outbreak in HSM preceding outbreaks elsewhere as a consequence of the infection growth rate in HSM being so much higher as to overcome the head start of cases elsewhere in Wuhan. For this to be correct requires three unlikely things to happen:

1. One of the first covid patients goes to HSM
2. The first (or one of the first) covid patients to visit HSM infects an HSM resident, and the infection takes hold as a local cluster
3. The HSM cluster grows so rapidly as to outpace non-HSM cases

It is possible to trade off one of these factors against the others by tweaking the numbers, but there is no avoiding that at least one of them will be improbable:

1. Assuming 50000 weekly visitors at HSM, and that each sick person is contagious (and active!) for one week, then for every sick person who visits HSM there are on average 245 sick people in Wuhan who do not; with a more realistic infection period that would be closer to 500 sick people in Wuhan. We can make the first step reasonably likely

by assuming only the 500th sick person went to Wuhan, but now the non-HSM cases have a much greater headstart to overcome.

2. Maybe the first sick person to visit HSM was super contagious and got lots of people sick, but now we have even more non-HSM cases to overcome, as well as needing to explain the coincidence that the person who happened to visit HSM was also super contagious. Or maybe all of the pre-HSM cases were super contagious, but then how did HSM outpace non-HSM? Or maybe it took multiple sick people visiting HSM for the outbreak to take hold there, but that makes it even more improbable that more than one of the earliest covid patients happened to go there.

3. This is the mirror image of the previous bullet points. Perhaps non-HSM cases did not have a big head start, but then it is quite unlikely one of the first non-HSM cases happened to visit HSM.

No matter how the model is adjusted, for this model to explain the observed cumulative case numbers at 2019 Dec 31 requires either multiple unlikely events or one especially unlikely event. This is not due to some simplification or model error, but inherent to the difficulty of explaining how the first cluster of cases is observed far from the purported origin. When the 2019 cases are broken down by time, the model does even worse at explaining events earlier in the outbreak; it predicts HSM cases will be heavily backloaded to the end of December; a week before the end of December, (ascertained) non-HSM cases should outnumber HSM cases by 8 to 1, and two weeks before by 32 to 1, whereas we instead observe that HSM cases are slightly frontloaded to earlier in the outbreak. The first HSM case should be ascertained four doubling times (two weeks) after the first non-HSM case, instead of 6 days *before*.

I think the maximal likelihood way to make this model consistent with observations is to suppose that one of the first 5 or 10 covid patients went to HSM and infected a vendor there, and the HSM outbreak steadily outpaced non-HSM cases. While this requires two quite unlikely events and one somewhat unlikely event,[55] it allows for a distribution of cases over time that is largely indistinguishable from one where covid started at HSM.

This can be substantially ameliorated if you believe Connor Reed, or other people, had covid before the HSM outbreak, but I found no evidence in support of that (see section A.2). Even a small change like Chen being sick on Dec 8 instead of Dec 16 makes Rootclaim's model something like 4 times more likely, as it alleviates part of the burden of explaining the absence of ascertained pre-HSM cases.

We are, however, ignoring one of the biggest difficulties of Rootclaim's model, which is explaining not only that HSM was the *first* covid SSE, but by quite a large margin in time. Even if we accept that one of the first people with covid went to HSM and started an outbreak there, this does not imply that the other people with covid now disappeared because they no longer relevant to the story. We know that covid is contagious outside of HSM because it continued to grow even after HSM was closed. If there are 128 or 245 or 500 sick people going around Wuhan at the time of the HSM primary case, shouldn't we expect other large

[55] Somewhere in the ballpark of $1/100 \cdot 1/20 \cdot 1/5$, not coincidentally very close to the Bayes factor I will eventually assign in section 7.6 to the outbreak starting in HSM.

SSEs to start popping up in other locations? Why was it another month before the first trickle of covid cases outside of China (or Wuhan) emerged?

4.6 Locating sars-cov-2 within HSM

4.6.1 Covid testing of environment and animals

Extensive environmental testing of HSM began on 2020 January 1, after the market was closed and disinfected that morning.[56] A total of 515 environmental samples were collected in HSM that day, and an additional 364 environmental samples were collected in HSM from January 12 to March 2. Additionally, 457 samples from animals were collected from January 18 to March 30. [22, 12]

None of the 457 animal samples tested positive for sars-cov-2. This is to be expected, due to the nature of the sample selection. Most samples were taken from dead animals in cold storage, but about 100 of them were from stray live animals or their feces around the market. The most commonly sampled animals were rabbit, stray cat, snake, and hedgehog. Of the 457 samples, only 16 came from animals (six different bamboo rats[57]) that were plausible carriers of sars-cov-2; the other 441 samples are irrelevant to identifying the origin of sars-cov-2. Furthermore, most samples were collected 2 months after HSM was closed, giving ample time for decay of virus genome. [22]

The environmental testing is much more informative, though not definitive. First, DNA testing of the samples gives insight into the types of animals being sold or processed at the market, including their origin, confirming that animals were imported from south China where wild coronaviruses are prevalent. Testing for other viruses also directly supports the possibility of sars-cov-2 being imported to the market in an animal: in particular, a bamboo rat betacoronavirus was found that likely originated in a bamboo rat farm in Gaungxi, which borders Vietnam and Yunnan. [12]

Second, 72 of the environmental samples tested positive for the presence of sars-cov-2 RNA by PCR, and from 3 of them live virus was able to be extracted and cultured, which were the same 3 that tested most strongly for sars-cov-2; 2 of the 3 came from stores with known covid cases. Of the 72 samples, only 27 were attempted to be cultured, all collected on January 1. In addition to the 3 samples from which live virus was extracted, one other sample was able to be sequenced; between them 3 of the samples were lineage B and one was lineage A. [22]

The positive environmental samples were not distributed uniformly around the market, and their distribution narrowly identifies a very specific location for a zoonotic spillover to have taken place. Sampling was targetted at shops that had live animals or had employees known to have had covid, so any statistical analysis of positive samples (such as figure 4B of [34]) must account for the non-uniform prior. However, the signal is strong enough that no statistical analysis is necessary.

[56] It would be disinfected regularly thereafter. Personally I do not find it encouraging that researchers none-the-less continued to recover positive sars-cov-2 samples for some time.

[57] Though Christoph et al. variously say bamboo rats are or are not susceptible to covid [12]; perhaps there is an important distinction between the hoary bamboo rat and the Chinese bamboo rat that I am missing.

22 stores in HSM had positive environmental samples; 18 had only one positive, 1 had two positives, 1 had three positives, and 2 had five positives (see [22] figure S1A for details). Of course, different stores had a different number of tests (many were not sampled at all), so we should consider the fraction of samples at a store that was positive. This becomes slightly difficult as many stores had 2 or fewer samples, so that the ratio has very low precision. Of the 18 stores with a single sample, one (east 9:22, and thus far from the origin of the HSM outbreak) was positive; of the several dozen with two samples, five had one positive and the rest none.

We are left with two stores that stand out: 4:26,28 which had 5 positives of 9 samples (and a known positive covid case), and the nearby 6:29 which had 5 positives of 10 samples. Both stores are located in the southwest region of the west half of HSM, generally the region of stores with wild animals and where most early covid cases were infected. (Of course, keep in mind most stores with 3 or more samples were in or near that region of HSM.) The latter store in particular had live raccoon dogs (one of the species suspected to be intermediary for sars-cov-2) and was specifically identified in 2014 as a potential location of a novel disease spillover [26, 39].

Besides the 33 positive samples in stores, other locations in and around HSM also tested positive for sars-cov-2. Visible under grates running along storefronts in HSM (though obscured in the photographs of HSM I included above) is a shallow drainage system which would have received biological waste from shops, including waste from any live animals. Of 60 drains sampled January 27 through 29, there were four positives, including the drain in front of shop 6:29. The drains were tested again two weeks later: of 17 drains sampled February 9 through 15, only three were positive – again the one in front of shop 6:29, and two downstream of it (which would have been exposed to waste from most of the west half of HSM). The drainage path of shop 6:29 is illustrated in figure 8.

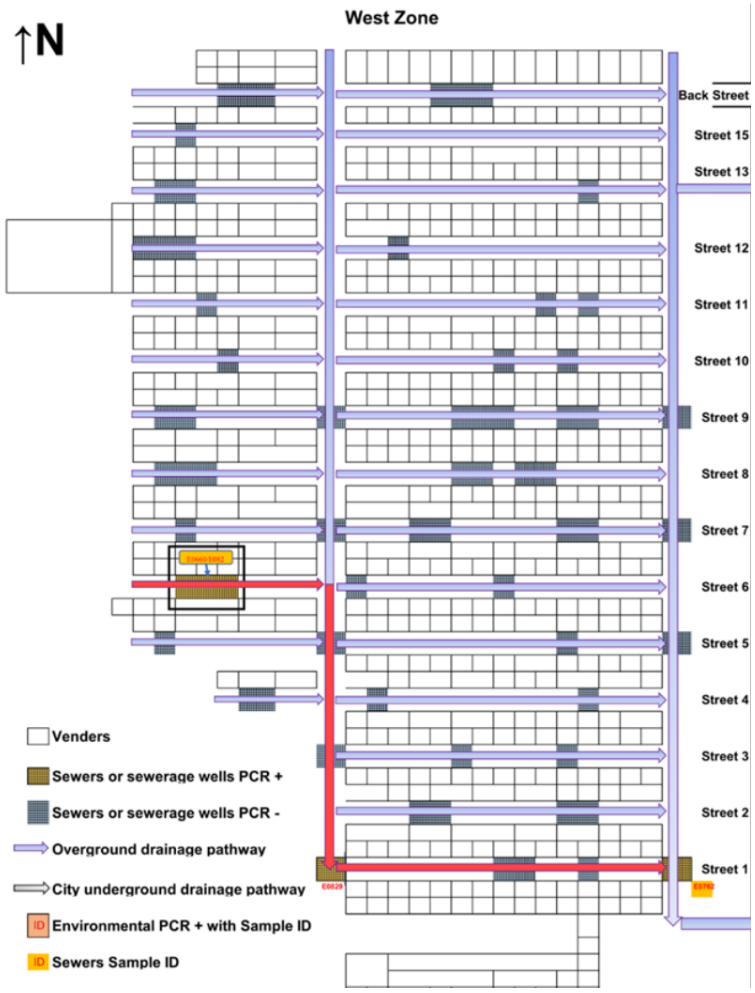

Figure 8: The topology of the ground-level drainage system for west HSM and the location of the three positive samples collected in February, with the drainage path from shop 6:29 shown. The locations marked in the legend as having negative tests include all samples collected in January or February and should be ignored; only 17 samples were taken in February. Figure is Extended Data Figure 1 of [22], with alterations and annotations by Peter Miller ([1] at 1:29:45).

4.6.2 Correlations between animal DNA and sars-cov-2 positivity

Rootclaim observed ([1] at 2:56:30) that statistical analysis of environmental samples show no or negative correlation between samples that contain DNA from species such as raccoon dog and samples that test positive. It would not have occurred to me to perform this comparison, and I do not think it is useful or relevant.

The animals at HSM were caged and did not have freedom to move about the market. While there would have been some disease communication between them, particularly within a single shop, disease communication between animals would presumably have been lower than between humans at HSM who could and did move between multiple shops and interact with other humans. I expect that at the time of spillover, very few animals at HSM would have been carrying sars-cov-2 infections, compared to the number who might be susceptible to it. (It is quite possible that some would later be infected by humans.)

Any positive correlation between samples that contain raccoon dog DNA and samples that test for sars-cov-2 would only occur if the samples happened to contain material from the specific raccoon dogs that were carrying sars-cov-2. There are probably few or no such samples, so the expected correlation under the zoonotic hypothesis would be weak or zero. Moreover, there are several species in HSM suspected as being potential sars-cov-2 intermediaries (such as raccoon dog, civet, and bamboo rat); presumably only one of them could have any positive correlation, with the others being zero as they are irrelevant.

Even if there were a positive correlation, it could be overwhelmed by confounding factors. After the first round of testing appeared to identify the region with live animals as a potential source of the pandemic, a second round of testing included 10 samples at each of 7 different shops that sold wildlife [22]. This sampling, two weeks later on January 12, would have had a lower positivity rate due to the additional time for the sars-cov-2 RNA to decay. Consequentially this would induce a negative correlation between samples with wildlife and samples with sars-cov-2.

4.6.3 Covid testing of animal vendors

Rootclaim claims that no animal vendors at HSM tested positive for covid, in contrast to a significant association between people who worked with wild animals and seropositivity for SARS-1 ([1] at 3:07:50, [5] at 1:46:00, [13]).

There are two reasons I do not find significance in this. First, as with the animals themselves, I would not particularly expect to find a trend connecting animal vendors with testing positive for covid even if sars-cov-2 did spill over at HSM. Only one animal vendor would have been directly infected by an animal (or two if there were two spillovers, as discussed in section 5.1), with further infections being evenly distributed between all sorts of vendors, with a slight bias towards animal vendors as they are loosely clustered in the west half of HSM.

As early cases have a lower ascertainment rate, we expect a low chance that the primary case would be ascertained directly: roughly 5%, the hospitalization rate. In the 95% chance that

the primary case was not hospitalized, we have no strong reason to be looking specifically to animal vendors to find early infections.

Second, it is not clear to me that the claim is true. Rootclaim gives no source for the assertion. Presumably if this information is available at all, the source would have to be the WHO report, which has a near monopoly on data on early cases, along with other publications co-authored by members of the WHO investigative team. The HSM map published by the WHO labels 10 stalls known to sell live animals, and does not show any cases at those stalls; this makes me think it is probably true ([32] page 46; or see section B.2). However I could not find any explicit mention in the text that no animal vendors were positive,[58] despite very lengthy discussion of testing of the animals themselves, and the WHO report only specifies the types of goods vendors sell in a very few cases. It does say that 20 of 168 December cases self-report exposure to live animals ([32] page 45), but does not clarify how many of those were at HSM, nor the nature of that exposure (could it include pets? but would only 11.8% of people have pets?). The breakdown of all vendors at HSM by type of goods sold ([32] annex page 182) does not mention live animals at all; perhaps this is included in "others", or perhaps they are included in meat sellers if the animals are being butchered on site, or perhaps the live animal sellers also sold other products.

5 Genetic evidence

5.1 Were there two spillovers?

As we saw in section 1.2, the early sars-cov-2 genomes can be neatly organized into two distinct taxa,[59] called lineages A and B. Each of these taxa has an unambiguous root, whose genomes differ by exactly two nucleotides: lineage A having T8782, C28144 (which we will call "T/C") and lineage B having C8782, T28144 ("C/T"). This is an unusual phylogenetic structure that cannot be easily explained by either the zoonotic nor lab leak hypotheses.

Simple analyses like minimum spanning tree can construct a reasonable unrooted phylogenetic tree, but do not help identify the root. As few or no early intermediate samples (T/T and C/C) are known, a simple geometric argument suggests that it is very unlikely the root is an intermediate genome. While coronaviruses differ in what nucleotides they have at these two positions, the closest relatives to sars-cov-2 are in agreement with lineage A, so outgroup analysis suggests that lineage A is the root of sars-cov-2. Conversely, when considering the collection dates of early sars-cov-2 samples, it appears that lineage B may predate lineage A by a few days or weeks and would therefore be the root.

Pekar et al [27] suggests a different possibility: what if lineages A and B represent separate introductions of sars-cov-2 into humans? They find in simulations that it is unlikely for a single introduction to yield a phylogenetic tree with two clearly distinct and prominent taxa whose roots are exactly two mutations apart; when this does happen, it is usually because

[58] Despite reading it multiple times I could easily have missed it – it's 300 pages long!

[59] Well, I think "taxon" is not usually used below the species level; I would say "clades", but one of them is not monophyletic.

an early infection underwent two successive mutations rapidly. On this basis they argue that two separate introductions better explains the observed phylogenetic evidence.

If true, and sars-cov-2 underwent multiple introductions into humans, then we could be all but certain that the zoonotic hypothesis is correct. However, there are several weaknesses in the argument of Pekar et al. One significant issue is that their simulations [24] appear (if I understand correctly) to model disease transmission as following a poisson distribution, whereas real-world transmission of covid may substantially involve sporadic super-spreading events, leading to a much more clumpy structure in the resulting phylogenetic tree.

A much more serious objection is that the two-introductions hypothesis does nothing by itself to explain the strange shape of the phylogenetic tree. Regardless of the number of introductions, both taxa definitely belong to a single tree; labeling certain nodes in the tree as "animal host" or "human host" does not change the shape of the tree. The only way that the number of introductions is relevant is if disease transmission has different dynamics in animals and humans. While this is likely to be true, Pekar et al does not explore this distinction, nor make any attempt to model covid transmission amongst animals. For the two-introductions hypothesis to make any sense, we must suppose that animal-to-human transmission is comparatively rare, and that covid was circulating in a small reservoir of animals; this is plausible, but significant work remains to quantify how well this hypothesis explains the data relative to other hypotheses.

I find myself intrigued but unconvinced by the possibility that there were multiple introductions of sars-cov-2 into humans.

If I had to guess whether lineage A or lineage B were ancestral, I would be inclined to believe Pekar et al that lineage B is more likely to be ancestral; while other researchers identified lineage A as ancestral, their work was more broadly looking at the whole phylogenetic tree for sars-cov-2 rather than narrowly focused on this single question. Fortunately I don't have to make such a guess. The most likely explanation for the lab leak hypothesis involves an early case going to HSM and starting a cluster of covid cases; if so, the shape of the phylogenetic tree will be essentially indistinguishable from that of the single-introduction zoonotic hypothesis. Neither hypothesis well explains the observation of early covid cases cleanly being grouped into two distinct taxa. While the zoonotic hypothesis does not explain why lineage A was mostly absent from HSM, the lab leak hypothesis does not explain why early lineage A cases were initially found *near* HSM, so again neither hypothesis does particularly well here.

5.2 Spatial distribution of lineages A and B

Among people who had onset of covid in 2019 December, 13 people had sars-cov-2 samples sequenced; 11 of these people had a known epidemiological association with HSM ([32] page 76). 12 of the 13 samples were found to be lineage B. This raises the obvious question: where was the lineage A?

More specifically, all 11 associated with HSM (onset dates: Dec 13, 17, 19, 20, 20, 20, 20,

22, 23, 23, 23) were lineage B, whereas of the other two cases, one (onset date Dec 16[60]) was lineage B and one (onset date Dec 26) was not. This gives the rather more pointed question – why was there no lineage A in HSM?

The lab leak theory has an elegant explanation for this: based on outgroup analysis, it appears that lineage A is ancestral to lineage B, so presumably it was lineage A that was leaked from WIV. One of the lineage A descendants mutated to become the root of lineage B, which would include the HSM primary case. Thus, the HSM outbreak, and all cases descended from the HSM primary case, would be entirely lineage B, whereas Wuhan as a whole would have a mix of the two lineages.

The zoonotic theory does not have a clean explanation for this observation. The most obvious, and simplest, is to suppose that lineage B was ancestral, so there were fewer lineage A cases, and by chance did not include any of the 11 associated with HSM. The sample size involved is so small that this could easily be simple coincidence. Supposing we knew that 1 of the 13 early cases were lineage A, and we were then to ask the chance of the observation that all 13 HSM cases were lineage B, we would get a p-value of $2/13 = 0.154$, which is not so unlikely.

An alternative way of regarding the situation is that if the primary covid case were in HSM, then the genetic diversity of sars-cov-2 within HSM should be equal to the genetic diversity within all of Wuhan. Observing that the one lineage A case was not in HSM is weak, but nonzero, evidence that the genetic diversity within HSM is lower than that of Wuhan as a whole, and therefore that HSM is not the origin of the pandemic.

However the lab leak explanation for this observation is not consistent with further observations. A glove in HSM (shop 7:15,17) was tested for covid, and was the one environmental sample to be sequenced with lineage A (out of four environmental samples in HSM that were sequenced). An employee of this shop became ill with covid with an onset before December 15, and therefore presumably was lineage A; but if lineage A were ancestral and the primary HSM case were lineage B, how did lineage A reach HSM so early?

Furthermore the first two (three) known lineage A cases were both found in close proximity to HSM. The first case sequenced with lineage A, with onset December 26, lived 2 km south-southeast of HSM [33]; he got it from his wife, onset December 15, who was presumably also lineage A. The second case sequenced, onset December 27, was infected while residing at an unspecified hotel "near" HSM [34].

It is hard to explain the first observed presence of lineage A in close proximity to HSM with the lab leak explanation above. Whether the origin of sars-cov-2 was zoonotic or laboratory, it appears that both lineages A and B were located in or near the market, and the best explanation for the observations is the same for both theories.

[60]Though the WHO reports Chen as having an onset of Dec 8, this appears to be incorrect.

5.3 CGGCGG

Much attention has been drawn to the sars-cov-2 FCS, shown below as found in the Wuhan-Hu-1 reference sample [25]:

```
CAG ACT CAG ACT AAT TCT CCT CGG CGG GCA CGT AGT GTA GCT AGT
 Q   T   Q   T   N   S   P   R   R   A   R   S   V   A   S
```

Recall the `RRAR` sequence is the FCS, with the putative insert relative to the closest known viruses being `T CCT CGG CGG GC` (or perhaps `CCT CGG CGG GCA`).

There are six different codons that code the arginine amino acid; in sars-cov-2 and similar viruses, `CGG` is the rarest of the six possible encodings of arginine, at only 3-5%. Rootclaim argues that this is evidence against a natural origin to the FCS, as it has two consecutive `CGG` codons.

However, I would not expect a priori the codon frequency in an insert to match that of the surrounding genome, but rather to reflect the frequency of the source of the fragment; note that `CGG` is the most common encoding of arginine in humans, for example. The main reason `CGG` is rare in sars-like viruses is that there is a mutational bias against the C and G nucleotides; C is the rarest at 18.4% of nucleotides in Wuhan-Hu-1, followed by G at 19.6%, whereas A has 29.9% and T has 32.1% [25]. After a fragment from an outside source has been inserted, it takes time for this mutational bias to impact the novel fragment.

Furthermore, if the fragment was inserted naturally, we do not know if the source had the same alignment, or even was a coding region at all. We would have to analyze the frequency of the non-aligned codons to properly assess the likelihood of the `CGGCGG` sequence from a natural source.

I do not see any particular significance to the `CGGCGG` sequence. It is slightly odd, but one expects to observe many slightly odd things through random chance.

5.4 Project DEFUSE

In 2018, the EcoHealth Alliance submitted a grant request to DARPA for $14 million for "Project DEFUSE", a 3.5 year project on bat coronaviruses in collaboration with 5 other groups (located at UNC, Duke-National University Singapore, WIV, USGS National Wildlife Health Center, and Palo Alto Research Center) [14, 20]. To my knowledge, the project was not funded. The project involved quite a large number of goals and steps and can be organized as follows:

1. Collect bat coronaviruses and perform computer and lab testing on them
2. Deploy vaccines and immune boosting agents to wild bats
3. Write an app for soldiers to figure out how likely bats are to have dangerous diseases (gotta appeal to DARPA somehow?)

5.4.1 Excerpts from proposal

I will be omitting quite a few details about irrelevant activities in the proposal. There are two relevant passages:

> We will screen samples for SARSr-CoV [sars-cov-1-related coronavirus] nucleic acid [...] Full-length genomes or S [spike] genes of all SARSr-CoVs will be high-throughput sequenced followed by genome walking. We will analyze the S gene for its ability to bind human ACE2 by Biocore or virus entry assay. *Synthesis of Chimeric Novel SARSr-CoV QS*: We will commercially synthesize SARSr-CoV S glycoprotein genes, designed for insertion into SHCO14 or WIV16 molecular clone backbones (88% and 97% S-protein identity to epidemic SARS-Urbani[61]). These are BSL-3, not select agents or subject to P3CO (they use bat SARSr-CoV backbones which are exempt) and are pathogenic to hACE2 [human ACE2] transgenic mice.
>
> [...] We will conduct in vitro [in test tube] pseudovirus binding assays, using established techniques, and live virus binding assays (at WIV to prevent delays and unnecessary dissemination of viral cultures) for isolated strains.

So, they will look for sars-like viruses, sequence their spike proteins, simulate how well those spike proteins bind to human ACE2 receptors, and put the genes for those spike proteins into known viruses which have been selected for their low pandemic potential. Those viruses will be given to mice with human ACE2 receptors. The proposal goes on to explain that after evaluating the chimeric viruses, some of them will be verified by testing the original wild viruses the spikes came from; the latter would be done at WIV where the wild viruses were originally sampled.

> *S2 Proteolytic Cleavage and Glycosylation Sites*: [...] We will analyze all SARSr-CoV S [spike] gene sequences for appropriately conserved proteolytic cleavage sites in S2 [spike subregion 2] and for the presence of potential furin cleavage sites. SARSr-CoV S with mismatches in proteolytic cleavage sites can be activated by exogenous trypsin or cathepsin L. Where clear mismatches occur, we will introduce appropriate human-specific cleavage sites and evaluate growth potential in Vero [monkey kidney] cells and HAE [human airway epithelial cells] cultures. In SARS-CoV [ie, sars-cov-1], we will ablate several of these sites based on pseudotyped particle studies and evaluate the impact of select SARSr-CoV S changes on virus replication and pathogenesis. We will also review deep sequence data for low abundant high risk SARSr-CoV that encode functional proteolytic cleavage sites, and if so, introduce these changes into the appropriate high abundant, low

[61] SARS-Urbani refers to a a reference genome of sars-cov-1, collected from Carlo Urbani. Urbani was a doctor and biologist who first realized SARS-1 was a new disease, and whose warning to the WHO likely saved many lives. He died of SARS-1 at age 46.

risk parental strain.

This passage is part of a long list of planned genetic experiments. They will look for cleavage sites, including furin cleavage sites, in existing cells, and genetically modify some of them to introduce "human-specific" cleavage sites. It is unclear to me whether furin is human-specific. I understand this passage to be referring to the chimeric viruses with known backbone (SHCO14 or WIV16). They also discuss adding cleavage sites to common, low risk sars-like viruses. After this passage, it goes on to discuss N-linked glycosation.

5.4.2 Proposed activities that would have involved WIV

The proposal also contains a detailed breakdown of the planned activities into 61 tasks, and which groups will be doing each. Here I have summarized each step that WIV would have been involved in:

1. Phase 1, tasks 3.1 and 3.2: Screen bat samples for sars-like viruses and sequence spike proteins
2. Task 6.1: make assays for identifying antibodies to known sars-like viruses
3. Tasks 6.2 and 6.3: verify assays on rabbits and blood samples from people who had SARS-1
4. Task 6.4: test assays on people in Yunnan to look for sars-like spillover events
5. Tasks 7.5 and 8.5: test viral[62] vector vaccine on captive bats
6. Phase 2, tasks 5.1 through 5.4: Identify cave features; surveil bats; deploy vaccines or immune boosters to caves; surveil bats

As far as I can see, no genetic engineering would be performed at WIV, nor would any chimeric viruses be brought there, except prototype vaccines which use a raccoonpox chimera. WIV was to be involved with samples collected from bats, tests on bats, tests on people who live near bats, and experiments in bat caves.

5.4.3 Previously documented research activities at WIV

In response to written questions, Rootclaim [7] identified two papers that described relevant laboratory work conducted at WIV. In both cases the work involved editing a backbone of WIV1, a specific sars-like coronavirus genome chosen for its lack of "pandemic potential", and thus presumably safer to work with than a pathogenic virus. In one project, they edited WIV1 by replacing its spike protein with spike proteins found in viruses collected from bats in Yunnan [19]. Human and monkey cells were then exposed to the recombinant viruses to determine its infectivity. In another project, they removed a gene in WIV1 that is absent in sars-cov-1 to see if the virus would still work in human and monkey cells [37].

[62] raccoonpox, not sars

5.4.4 Summary of significance of proposal

The proposal does not explicitly state it would add furin cleavage sites to viruses, but it does discuss adding cleavage sites in general shortly after referring to furin cleavage sites by name.

Most, but not all, of the genetic modification involved using backbones from known, pre-existing viruses.

The proposal does not, as far as I understand, discuss any *in vivo* genetic modification; all editing was to be done by directly altering sequences, not by passaging through humanized mice or other organisms. Viruses would be used *in vivo* only for testing pathogenecity.

The proposal does specifically say it would screen for wild viruses with high affinity towards human ACE2.

Overall, the Project DEFUSE proposal is a bit eerie in how it calls out multiple features that are quite distinct in sars-cov-2. However these features were previously known or suspected to be important to human coronaviruses, and the proposal is hardly a blueprint for building sars-cov-2, containing many irrelevancies and missing key components.

Project DEFUSE is most relevant in that it shows that the community of virologists was broadly interested in making chimeric viruses with certain features which have the potential for lab leak risk if handled improperly, and not that WIV in specific planned to do a particular research path that would lead to creating sars-cov-2.

5.5 ACE2 binding affinity

Piplani et al [28] found in computer simulations that sars-cov-2 binds best to human ACE2 receptors out of 14 species studied; note that other studies found conflicting results.

This is hard to explain under zoonotic theory, as the sars-cov-2 sample used was from 2019, before significant evolution in human hosts should have happened. However, the second highest affinity studied was pangolin ACE2, which is similar to human ACE2; if pangolins were an intermediate host, then sars-cov-2 could have adapted to pangolin ACE2 and thus partially to human ACE2.

Under the lab leak theory, this observation is explained as a result of lab workers screening many different wild viruses to identify those that bind well to human ACE2, and possible then passaging the virus through mouse hosts that have been genetically modified to have human ACE2 receptors. While this may be a better explanation then the zoonotic theory, it is still not a great explanation. This process requires screening a large pool of wild viruses to find one that happens to be well-adapted to humans already; and presumably researchers are screening to find which one best binds to humans, indifferent to whether it binds better to humans than other animals. Also, this screening process may use different simulations than Piplani et al used.

A control is also required; maybe human ACE2, either in reality or in the simulation of Piplani et al, is just particularly good at binding to spike proteins?

The high binding affinity found by Piplani et al appears to be weak circumstantial evidence in favor of the lab leak hypothesis.

5.6 Leading proline

The amino acids near the sars-cov-2 FCS are

```
QTQTNSPRRARSVAS
```

whereas the most closely related known virus banal-20-52 has

```
QTQTNSRSVAS
```

at this location. The difference between these is `PRRA`.

The `RRAR` subsequence constitutes an FCS, which explains why a laboratory worker might insert a sequence like `RRA`, but wherefore the proline `P`?

I see no satisfying explanation for what might motivate the decision to include a proline at this location, and it seems quite unlikely that the natural precursor virus happened to have such an inserted proline at the exact location that researchers add an additional insert. However we should not expect every decision someone makes to have an obvious post hoc explanation. There are many arbitrary choices during the course of a genetic engineering experiment, and it is hardly strange if a few of them are not the most canonical or optimal.

I do not think the proline requires a specific explanation under the lab leak theory (and likewise the "out-of-frame" nature of the insert).

5.7 12 nucleotide insert at FCS

As shown in section 1.2.2, sars-cov-2 has 12 extra nucleotides at the FCS when compared to the most closely known bat viruses, such as banal-20-52.

Certainly this looks suspicious, but there is a lack of definite information to say anything more than that. There is no particular reason that this mutation could not have evolved by chance. The only thing odd about this insert is how large of a change it is compared to the smaller changes elsewhere in the genome; however, we do expect the S1/S2 boundary region to be under high evolutionary pressure, especially as the virus crosses species into mammal hosts that could make use of the FCS.

While banal-20-52 is the closest known virus, it is not close enough to make the kind of comparison that Rootclaim seeks to do with sars-cov-2. The two viruses are separated by decades of evolutionary time and likely multiple recombination events with other viruses.

6 Can you keep a secret?

If sars-cov-2 were made in a lab, could it be kept a secret?

This is difficult to estimate in the best of circumstances, and here I am concerned with the practices of a lab about which I know nothing, all primary sources about which are in a language I cannot read, and in a culture I have no experience with. I do not know any of the people involved or how they usually operate, and I am restricted by the rules of the debate to the information brought by each side, which includes almost nothing that bears on this question.

So I must keep in mind my great uncertainty; and yet I find it hard to imagine how to draw a line around the bubble of people aware of the artificial origin of sars-cov-2 so that that information does not escape that line.

Possibly zero people know. Researchers could have inserted an FCS into a virus, created sars-cov-2, gotten sick and released it from the lab without ever realizing – even in retrospect – that the virus in the lab was the same as the one in the wild. But surely they would have figured it out eventually? Researchers at the WIV had already examined a preliminary sequence for sars-cov-2 by 2019 December 27,[63] wouldn't they have thought at some point to look at that sequence and see if it is like one of the viruses in their lab?

Possibly one person knows. Some overworked, underpaid grad student took initiative to charge ahead on a (major, multi-year) research project unsupervised and got themselves sick. Shocked and overwhelmed by what happened, they covered up their research, threw away the materials, and vowed to take the secret to the grave.

Possibly most of the lab knows. Sometimes research groups will write proposals based on work that is largely already completed; maybe the PI had started a major research project on the speculative hope of it getting funded down the road. The group worked hard and had success – until the pandemic, and they were afraid that if it became known how reckless they were, the whole lab would be shut down. The group agreed, perhaps tacitly, that no one was to know.

Possibly people outside of the lab know. Funding sources in China might have approved the project. Researchers might have discussed their work with professional colleagues, friends, and family members. Domestic and international collaborators might have contributed to the research. Chinese officials might have interviewed lab personnel and found out the truth, or found physical evidence of the virus in the lab. Lab workers might have accidentally or anonymously let slip the truth, or pieces of the truth, on social media. (Could the truth of covid's origins be openly circulating on Chinese social media and western investigators just never see it?)

No matter how I draw the line, it is challenging for me to write a plausible story in which the secret stays inside. I think the most reasonable of these is the third, that the secret does not leave the lab. In this version, maybe a half dozen or dozen people directly worked on the project to create sars-cov-2, and then told no-one, and several took active steps to conceal it. Or perhaps the true explanation of how the secret kept is too strange for me to ever think of it.

If we imagine that at least six people knew about the research work that created sars-cov-2

[63]The source I saw for this claim was a broken link, but the exact day isn't important here.

and each has at least 50% chance to reveal some information about the existence of their research, then there is only 2% odds of the secret staying secret. Conspiracies are hard to keep, especially when the participants are unwilling (ie, they didn't specifically choose to join the conspiracy) and secrecy is required weeks to years before there was any known reason for it. 2% sounds about right, but considering my great uncertainty I could not give this lack of evidence a Bayes factor greater than 10.

Rootclaim evaluates lack of whistleblower, intelligence leaks, and published backbone (any of which would constitute the secret escaping) as a combined probability of 14%, not far from the 10% I quite conservatively gave it.

7 Analysis

Having criticized others' arguments at great length, at last I come to the more difficult task of presenting my best guess as to the origin of the covid pandemic. I will present this in the structure of a Bayesian inference computation, though not wholly quantitative, with my choice of numbers being subjective and uncertain.

Let Z be the hypothesis that covid emerged from a zoonotic spillover in HSM, and LL that covid resulted from gain-of-function research in WIV; these are very slightly narrowings from the actual hypotheses at stake.

Note that if you accept Rootclaim's argument in section 3.1, then the definition I gave for Z is quite a significant narrowing, as they believe the observational evidence is equally consistent with zoonotic spillover anywhere in Wuhan as in HSM, and so Z would have a suitably greater prior (but much lower posterior).

The first step is the most challenging, to estimate the priors $P(Z)$ and $P(LL)$. These numbers will be given as annual risk, so can also be interpretted as the yearly rate at which pandemics emerge from those sources.

7.1 Zoonotic prior

Before the debate I had a fundamental misunderstanding of what drives new pandemics. Consider Figure 9, which is a subset of known sarbecoviruses most closely related to sars-cov-2; recall that sarbecoviruses are themselves a branch of betacoronaviruses, which are a branch of coronaviruses. Conversely, while sars-cov-2 is considered one virus "species", it can be subdivided endlessly into lineages or strains. This great diversity of viruses is driven by their high mutation rate (compared to eukaryotes) and genetic recombination.

I therefore had (wrongly) imagined the gamut of viruses as akin to a *continuum* of possibilities. Spillovers from animals to humans are rare, and when they happen they effectively sample a random virus from that continuum of viruses that exist in animals. Depending on the properties of the virus, the spillover could continue to create a pandemic or not.

This is wrong, as if it were true it would be essentially impossible for the "same" virus to spillover multiple times, like picking random real numbers and getting the same number

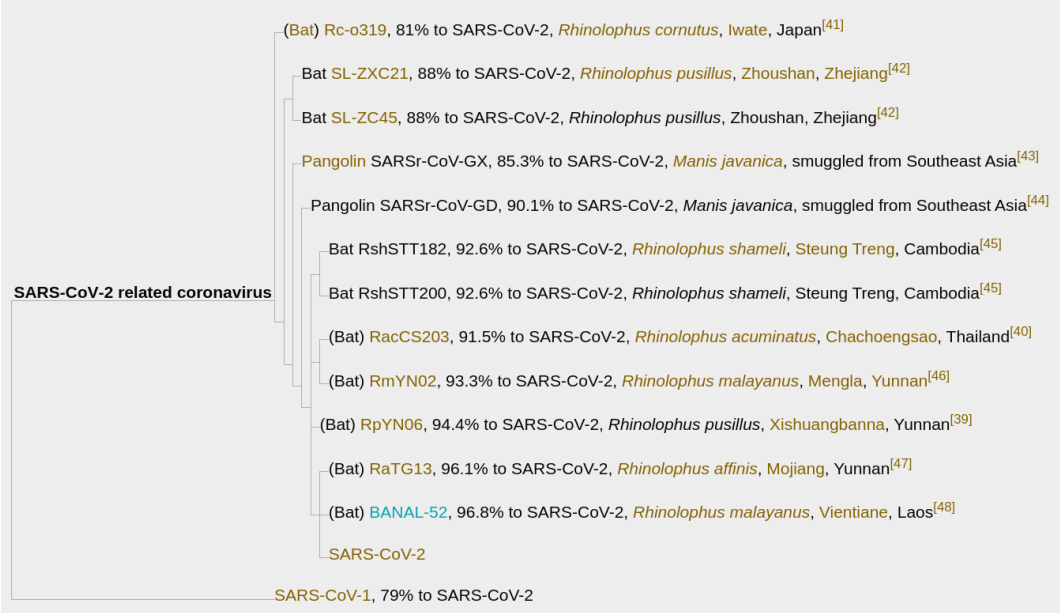

Figure 9: Some of the closest known related sarbecoviruses to sars-cov-2, from Wikipedia.

twice. Instead, it appears that (potential) spillover events are quite common,[64] and the primary limitation on pandemics is not whether the animal-to-human transmission occurs but rather whether a virus with the necessary properties evolves in an animal population humans are exposed to. Once such a virus evolves, it easily passes to human hosts, possibly many times; SARS-1 had multiple known independent spillover events,[65] and MERS has had dozens to hundreds since its discovery in 2012 [11].[66]

We will be using the emergence of previous human coronaviruses to choose a prior $P(Z)$. The multiple spillovers of SARS-1 and MERS are not independent events; since the limiting factor seems to be the evolution of a suitable virus, not its spillover, these should only count once each. There are 7 known human coronaviruses, two discovered in the 1960s, and the rest since 2002. Most of these only cause mild disease, and may have been circulating long before their discovery; the 1889 pandemic that killed 1 million people may have been the spillover of hcov-oc43, which today typically causes only colds. SARS-1 and MERS are the only "sars-like" viruses I consider a suitable precedent, and both emerged in humans since 2002; this would give a rate of 1.5 per 21 years or 1 per 14 years.[67]

[64] According to Peter Daszak of EcoHealth Alliance, more than 1 million people in southeast asia are exposed to bat coronaviruses each year ([32], annex page 71).

[65] The US CDC found in 2003 that 13% animal traders in Guangdong, where SARS-1 was discovered, had antibodies to SARS-1 despite none of them having had known illnesses; the subgroup with highest incidence was palm civet traders, at 73% (16 out of 22), compared to 1-3% in control groups. This suggests spillover events are commonplace. [13]

[66] I infer the paper as implying an estimated 200 spillover events, but it doesn't say so directly, so I might be misunderstanding.

[67] What we want to know is the expected *gap* between consecutive novel coronaviruses; ignoring sars-cov-2, from 2002 to 2023 there were sars-cov-1 in 2002 and MERS in 2012. We get 1.5 by assuming we are (on average) half way between MERS and the next novel coronavirus, so there have been 1.5 "gaps" since

We should also consider prior to 2002; but how much prior? I believe the risk of pandemics has grown greatly on the timescale of decades, due to changes of land use, agricultural intensity, and population; consider the increasing size and frequency of Ebola outbreaks in the last 10 to 30 years. If we limit our attention to the last 40 years, under the belief that outbreaks were less likely to occur or be noticed before then than today, we get an overall rate of 1 per 20 years.[68]

Not all sars-like viruses can cause a pandemic. Estimating from sars-cov-1 and MERS we get a $(0+1)/(2+2) = 1/4$ chance (I forget the name of this correction method) of it happening; and while sars-cov-1 is quite like sars-cov-2 in some ways, the sars-cov-2 virus does seem uniquely terrible, in part due to its initially very high presymptomatic transmission rate. Perhaps a less terrible virus would not have lead to a pandemic. On this basis I will add another factor of 1/2.

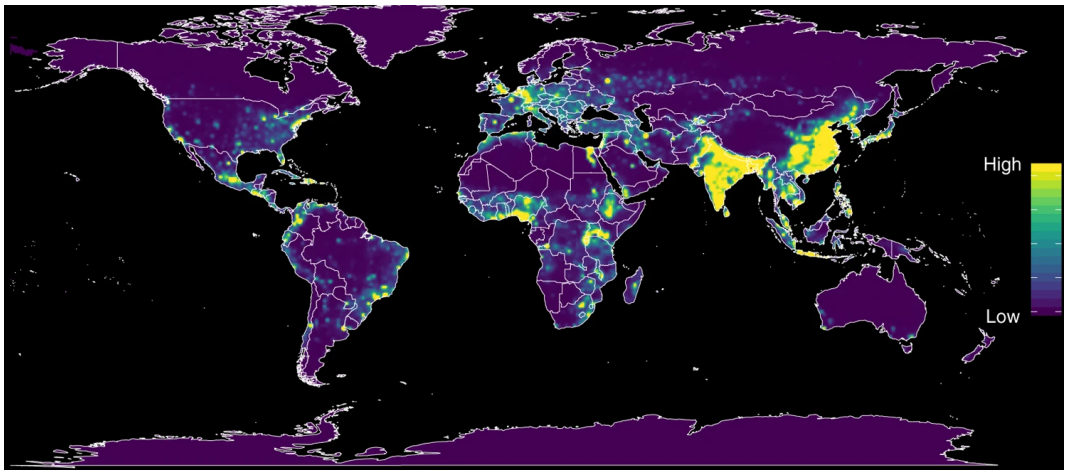

Figure 10: Estimated relative risk distribution of emergence of zoonotic infectious diseases. A large fraction of novel zoonotic diseases are expected to emerge in south or east asia; in addition to population density, other important factors are "forested tropical regions experiencing land-use changes and where wildlife biodiversity (mammal species richness) is high". Figure 3b from [9].

If a new sars-like virus emerges, one expects probably to see it somewhere in east Asia, not necessarily HSM in specific. Conditioned on the emergence being a wildlife market in east Asia, Peter gives the probability of HSM as 1/50 [8]; Saar gives it as 1/500 [7]. Both sides' calculation involved naively using the population of Wuhan compared to the population of China / Asia; but Saar observed that HSM is the largest market of its kind in central China, so using raw population figures probably yields an underestimate. My vague impression is that HSM is atypical in the extent of its live wildlife presence. Furthermore, biologist Eddie

sars-cov-1. We don't count a gap before 2002 because we chose 2002 to line up with sars-cov-1. We exclude sars-cov-2 because we are trying to estimate a prior for its emergence. If you were to include sars-cov-2, you'd use the interval 2002 to 2019 and so estimate 2 per 17 years; you can't go past 2019 because the timing of when this analysis is being conducted is influenced by the timing of sars-cov-2.

[68] Again, we count gaps: there was half a gap before sars-cov-1, half a gap after MERS, and a full gap between, for 2 gaps in 40 years.

Holmes and the Wuhan CDC specifically called out HSM as a possible location for a future pandemic spillover [26, 39].[69] Conversely, pandemics can emerge in locations other than wildlife markets. In toto I will assess this as 1/200.

(One could also get 1/200 through a very crude calculation like the following: suppose there is a 50% chance spillover is in China, conditional 50% chance in one of the 10 largest cities in China, and 20% of the relevant wild animal contact in Wuhan takes place at HSM. This last figure feels a little high to me. Note that Rootclaim gives a figure of 15% to 30% chance of the spillover being in HSM conditioned on it being in Wuhan.)

Overall this gives

$$P(Z) = \frac{1}{20} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{200} = \frac{1}{32000},$$

which is to say that HSM should produce a pandemic akin to sars-cov-2 every 32000 years.

Because the dynamics of flu evolution is fundamentally different than that of coronaviruses, I do not find flu pandemics/epidemics/outbreaks useful for assessing the prior $P(Z)$.

7.2 Lab leak prior

To estimate $P(LL)$, rather than extrapolating from historical lab leak events, I will work forwards from first principles. This is an error-prone approach, not just because it requires speculating about the activities and safety of a lab I have no first- or second-hand knowledge of, but also because any errors introduced are unlikely to cancel out with errors in $P(Z)$. However this is necessary as I know no analogous precedent for LL.

As my numbers are more-or-less conjecture from thin air, I will just give them in tabular form:

WIV conducts GoF research a la "DEFUSE"	1/50
(annualized)	1/1.7
WIV uses suitable starting virus	1/2
WIV research is successful	1/2
lab leak of that virus occurs at WIV	1/50

The Project DEFUSE research plan [14, 20] looks eerily similar, in parts, to sars-cov-2. However following Project DEFUSE exactly is not consistent with LL. Here are some of the most important discrepancies:

- Project DEFUSE was a collaboration of multiple research groups; but the more people involved, the easier for people to become aware of the research
- most of the work, including all work on virus chimeras, was to be done in UNC or other US labs

[69] The proper way to include this is to first estimate the chance of spillover at HSM excluding Holmes's prognostication, second estimate the chance the spillover site would have been successfully forecast (and applying a Bonferroni correction), then take some sort of average of these two estimates according to your relative confidence in them.

- the project was planned to last 3.5 years
- the project was interested in viruses related to sars-cov-1
- the project planned on using backbones from previously known viruses, selected for lack of pandemic potential in humans

For LL to occur, WIV had to undertake the contents of Project DEFUSE alone without collaborators (presumably their former collaborators would know of a falling out between them?), abandon the idea of using wild viruses closely related to sars-cov-1, use wild backbones instead of safe backbones, and rush the project in half the proposed time. It is also necessary for WIV to keep the elements of the proposal that are in accord with sars-cov-2 without alteration. The only known precedent for activities like this taking place at WIV is the construction of a chimera of the WIV1 backbone with spike gene collected from wild viruses; I know of no precedent for WIV inserting FCSs, filtering viruses based on infectivity, nor using wild-type backbones. I find the probability of all of this to be *at most* 1/50.

That gives the probability of WIV undertaking the GoF research as described, but we want an annual risk. Assuming a Poisson distribution of lab leaks with unknown parameter, after 1.7 years the annualized probability of the first lab leak occurring is at most 1 per 1.7 years.

Under LL, WIV must find and choose a sufficiently dangerous virus to start with, which at minimum requires such a virus to exist. Arguably the latter is the same factor of 1/8 as from Z, but I have weakened this to 1/2 as GoF research lowers how dangerous the starting point would need to be. This assumes that WIV is able to find the virus, though in reality only a small fraction of wild viruses are ever sampled.

Finally, we estimate the chance of the virus actually escaping. Rootclaim describes this as almost certainty once sars-cov-2 is manufactured, but I cannot take this seriously when Rootclaim does not quantify the risk of a outbreak of covid at the lab origin, and suggests that 128 sick people in Wuhan might lead only to a cluster at HSM. Even supposing WIV chose not to use a pseudovirus and performed aerosol-generating procedures on the virus under BSL-2 conditions, is this riskier than having a colleague with covid? However likely it is for a researcher to get covid from a vial in a fume hood, it seems far more likely to get it from a fellow researcher – they aren't bringing their BSL-2 equipment out to group dinners. The chance of a lab leak is surely lower than the chance that leak is followed by a outbreak in the lab, and Rootclaim finds the latter to be too low to be noteworthy.

Also, while lab leaks of infectious agents do happen, they still are very rare, *especially* if you restrict to lab leaks that had secondary infections; there are many virology labs, and most have never experienced a leak. I cannot put the probability above 1/50; I think it is lower, and only put it so high as sars-cov-2 is uncommonly contagious.

Putting these together, we get

$$P(LL) = \frac{1}{50} \cdot \frac{1}{1.7} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{50} = \frac{1}{17000}.$$

Compared to $P(Z)$, we see $P(LL)$ is twice as high. While zoonotic pandemics are much more likely than GoF lab leak pandemics globally, the latter tends to happen in Wuhan

more often than the former. My intuition is that the Bayes factor $P(LL)/P(Z) = 2$ I calculated is grossly too high.

7.3 First cluster in HSM

Let H be the observation that the index case is a resident in the west half of HSM (or, more precisely, that the first significant covid cluster is centered among the residents in the west half of HSM).[70] Following the simple model of section 4.1, the probability of observing H equals 1 divided by the population of the smallest epidemiological circle that contains both the primary case and the index case. For Z this is simply the west half of HSM, and for LL this is all of Wuhan, so we get

$$P(H|Z) = \frac{1}{600}$$
$$P(H|LL) = \frac{1}{1.2 \cdot 10^7}$$

for a Bayes factor of 1/20000. As I do think a market vendor is more likely to catch an illness than the average resident of Wuhan, I will give LL an extra factor of 2.

As per previous discussion, we are using a flat prior for the number of transmission between the primary case and index case, so both $P(H|Z)$ and $P(H|LL)$ should be smaller by the same factor, which has no influence on the Bayes factor. I think the assumption of a flat prior here is somewhat favorable to LL.

7.4 Environmental samples in HSM

As discussed in section 4.6.1, the distribution of positive sars-cov-2 samples within HSM appears to generally point to the region where live animals were kept, or more specifically point to shop 6:29 which had been identified as a potential pandemic spillover location.

Of the 15 shops which were sampled 10 or more times, 12 had no positives, 3 had one positive, and shop 6:29 had the other 5 positives. Arguably, this observation has a Bayes factor of 15 in favor of Z, as LL has no viable explanation for this coincidence. While this observation is certainly very post hoc, more careful statistical analysis places 6:29 at the geographical center of the sample positivity ratio, and the positive drainage samples likewise implicate alley 6.

However the HSM environmental samples will not contribute to my Bayesian calculation, as will be explained.

7.5 Genetic evidence

Most of the genetic evidence is either already incorporated into the prior or too weak to compel anything; it's hard to find definitive proof of a virus's origin by examining its genetic sequence. What we are left with is two observations: the strange 12 nucleotide insert creating

[70] The two halves have similar areas, so I assume workers are roughly evenly split between them.

a furin cleavage site, and the strong affinity of sars-cov-2 to bind to human ACE2 over other species.

I give these Bayes factors of 20 and 2 respectively; the latter cannot go above 14 as that was the number of species in the study.

7.6 Conclusion

We've collected the most pertinent evidence and invented Bayes factors for each of them, so it is time to make a grand total.

evidence	Bayes	log-odds
prior	1.88235	0.63
outbreak at HSM	1/10000	−9.2
12 nct insert at FCS	20	3
ACE2 affinity	2	0.69
secret doesn't leak	1/10	−2.3
total	0.001255	−7.19

which gives a conditional probability of

$$P(LL|O) = 0.07529\%$$

Note that in this table I have combined observations differently from how either of the debate participants did; in particular most of the evidence involving DEFUSE, some of the genetic evidence, and the fact that the outbreak started in Wuhan are all baked into the prior. This is why the prior is so favorable to LL even though most pandemics are zoonotic in origin. I prefer this to using a prior for outbreaks globally and narrowing to Wuhan-specific, as the latter would introduce unnecessary errors in calculating the global GoF lab leak rate when all I need is the rate in Wuhan.

Do I believe the probability of a lab leak is only 1 in 1300? No, due to the significant uncertainties in my analysis I would give it some higher probability. Many of the components in my calculation should be dampened towards neutrality to account for the uncertainty in them.[71] However I was at times quite conservative, by my estimation, in favor of lab leak, so I would not shift the final probability much more than I have already; and no amount of dampening towards neutrality can make the probability cross the 50% line.

Moreover I am not convinced that this Bayesian calculation is even an appropriate way to estimate the relative posterior probability of Z and LL; it just seemed fair that after criticizing Rootclaim's calculations at length I should make an attempt at it myself.

I think the numerical calculation gives an undeserved illusion of certainty, so I prefer perhaps a more qualitative summary:

[71] Most obviously the 1/10000 factor sticks out as needing significant dampening; however it is the only factor that is actually based on calculating something, has much higher confidence than the others, and has already been dampened by a factor of 2.

evidence	strength
prior	lean zoonotic
location	strong zoonotic
genetic / DEFUSE	weak lab leak
secret doesn't leak	weak zoonotic
total	strong zoonotic

Here I have broken down the observations differently, better according to how they align with my intuition; some of the evidence previously included in the prior has been moved to location and genetic / DEFUSE. Notice this qualitative assessment is barely consistent with the numbers I gave before, as my intuitive feel for the evidence favors Z more strongly than the numbers did.

It is worthwhile to at least briefly consider whether my selection of evidence is sensible or motivated. I have excluded quite a few items mentioned by Rootclaim; not that this would impact the result, as I evaluate each of the excluded items as indistinguishable from neutral.

My basis for including evidence is loosely if it is either an essential component of the zoonotic / lab leak hypotheses, or if it is quite strong. The location of the outbreak qualifies on all three counts (really, it'd be crazy to ignore where the outbreak started when trying to evaluate how it started). The genetic information and DEFUSE proposal are included as they are inseparably tied to the lab leak hypothesis. Finally, I included the failure of the secret to escape the lab as it is strong evidence (as I gave it 1 / 64 before conservatively adjusting the Bayes factor to 1 / 10), and arguably an essential "step" in the lab leak theory.[72]

One piece of evidence I excluded was the distribution of positive environmental samples within HSM. While it appears to identify the origin of the HSM outbreak to be specifically the area where wild animals were kept, this particular evidence was not an essential component of the zoonotic hypothesis, nor was it so convincingly strong in isolation as to not possibly be a coincidence.

I have an alternative, less Bayesian, approach to thinking about the lab leak hypothesis. The centerpiece of the lab leak explanation is the Project DEFUSE proposal, and the observation that a number of features from that proposal match features of sars-cov-2:

- Sars-cov-2 first appeared in Wuhan, the same city as one of the labs involved in DEFUSE
- Sars-cov-2 has an FCS that looks artificially inserted when compared to banal-52
- Sars-cov-2 has a strong affinity to human ACE2
- Something about N-linked glycans

[72] Conversely, the lack of a known intermediate host for sars-cov-2 could be analogously considered an essential "step" to the zoonotic theory; I don't mention it as I find this evidence indistinguishable from neutral.

This is sufficient to raise suspicion: how odd to see so many specific details identified in advance of the pandemic![73] But the zoonotic hypothesis has an equivalent to Project DEFUSE, namely Eddie Holmes's 2014 photos [26]. It is a remarkable coincidence that, when searching for potential zoonotic spillover sites, the Wuhan CDC took him to HSM, specifically to the very shop 6:29 which had the greatest number of positive environmental samples (and whose owner was fined for selling illegal wildlife). This coincidence of location is likewise odd, and has no explanation under LL.

Figure 11: Photographs of shop 6:29 in HSM taken by Eddie Holmes in 2014. Photographs are from [39].

Neither the existence of the Project DEFUSE proposal nor the Eddie Holmes photos are definitive, or even strong, evidence as to the origin of sars-cov-2; they rise only to the standard of being odd, possibly just coincidences. However, Project DEFUSE is the *only* significant evidence suggesting the possibility of a GoF lab leak. Looking through the Bayesian calculation presented by Rootclaim, the remaining pieces of evidence they give a Bayes factor larger than 2 are `CGGCGG` , no known intermediate animal host, and "adjust for BSL-2", none of which I think are distinguishable from a Bayes factor of 1.

While the lab leak hypothesis has only loose circumstantial evidence in its favor, the physical evidence of the covid pandemic starting in the immediate vicinity of live animals susceptible to sars-like coronaviruses and the precedent of previous coronavirus epidemics starting in an analogous manner together point strongly in favor of the hypothesis of zoonotic spillover.

[73] Though, the zoonotic hypothesis has a neat explanation: these details were identified because they were known to be relevant to pandemics.

A Assorted arguments presented by Rootclaim

So far I have only directly discussed a small minority of the many points brought up by Rootclaim in the 20 hours of debate. It is fair for the reader to wonder why I was not persuaded by those other points, so I address a few of them here, focusing largely on the arguments I think others are most likely to have found convincing.

A.1 Lineage B outcompeted lineage A

Rootclaim argued that some of the observed distribution of lineages A and B in 2019 December were due to lineage B having greater evolutionary fitness than lineage A. However, at this time the two lineages disagreed in only two nucleotides, one of which was a synonymous mutation, and the other having unclear relevance.

The D614G mutation greatly increased the evolutionary fitness of sars-cov-2; a strain within lineage B that had this mutation grew around 2020 February or March and outcompeted all other sars-cov-2 strains of either lineage at that time. This created the appearance that lineage B could outcompete lineage A, but it was only due to a mutation that was not significantly present in Wuhan in 2019 December.

A.2 Connor Reed

Connor Reed was a 25 year old British man living in Wuhan who claims he became ill with covid on 2019 November 25. In interviews, he says he went to the hospital and was tested on December 6, and he was told that the test came back positive for covid on 2020 January 16.

If I believed every anecdotal story I saw related on the internet of people contracting covid, I could point to a swathe of covid cases worldwide that preceded the lockdown of Wuhan. There are far more people who believe they acquired covid in 2019 than ever did have covid in 2019. Nevertheless, Reed's story is more credible[74] than others', as he was willing to put his name to it, gave substantive details in interviews with the press, and claimed to have received a positive test result.

It is unlikely for a 25-year-old with covid to be hospitalized. Furthermore, Reed described his illness as undergoing multiple recoveries and relapses, so even if his story is accurate we only know that he had covid as early as December 6, which could have been the last of a succession of two or more unrelated illnesses.

It is also unlikely for the index case of covid to be a British man. English-speaking residents in Wuhan are much more likely to be interviewed by Western journalists; this introduces a strong selection bias as to which cases are ascertained by Western newspapers. While this selection bias is equally strong for true and falsely believed cases of covid, the fact that British newspapers were able to locate a British index case should raise our prior belief that a significant fraction of people are willing to falsely believe that they had covid.

[74] or less non-credible

If Reed is telling the truth that his samples tested positive in a retrospective search conducted by the hospital he went to, then why does no retrospective search contain his case? This necessitates the existence of unpublished retrospective searches.

I do not believe Connor Reed acquired covid on November 25.

A.3 Mahjong hall

Rootclaim identified the possibility that the mahjong room in HSM may have been a nexus of early covid cases. I think there are two facets to their tentative argument; one, anecdotal observations linking early covid cases to the mahjong room; and two, a supposition that the mahjong room would have an unusually high rate of disease transmission.

I think neither facet is sufficiently substantiated. Rootclaim claims that at least 35 early covid cases have been linked to the mahjong room or mahjong players in some way ([1] at 3:17:30). As far as I can tell, this number is reached by adding up each of the listed anecdotes. 27 of these come from a single source, a Chinese-language video that I cannot evaluate, or even verify the quote is in the video. None of the sources are reliable, being indirect anecdotes with no identifying details and few specifics. I do not see any attempt to de-duplicate individuals that might have appeared in multiple anecdotes, nor compare against the background rate of mahjong players.

While the HSM mahjong room would be more conducive to disease transmission than the shops, I don't see a reason it would be *especially* outstanding, nor in comparison to mahjong rooms (or bars, restaurants, etc.) elsewhere in Wuhan, which are numerous. And regardless of how infective the room may be, its effect is limited to people physically in the room, which appears not to have been many:

> The shopkeeper mentioned an illicit mah-jongg parlor in the seafood market. "I heard that four people were playing at one table, and all four got sick," he said.
>
> [...] By then, the journalist had heard about a possible virus, and also about some mah-jongg infections. He found the players unfazed by the rumors, still gambling for small stakes. "It was kind of a secret room," the journalist said. [...] "It was next to the public toilet, and you had to climb a ladder to get there. You wouldn't find it unless you were looking for it. Everybody was smoking and there was no ventilation." [18]

Out of thoroughness I read each of the six sources provided in so far as I could:

- the New Yorker article quoted above; it contains fourth-hand accounts which were related in 2020 August.
- a Chinese-language video
- an Australian reporter who met a Wuhan resident who said their friend is (as of 2020 January 22) in the hospital with covid and thinks he may have gotten it while playing

mahjong, with no further detail. This is well over a month after the time relevant to the HSM outbreak.

- a contemporaneous Chinese-language news report with many interviews with employees at HSM; based on a machine-translation, it contains two references to shopkeepers playing cards but does not mention mahjong. I do not think it is fair to translate the characters "打牌" ("play cards") as "play mahjong" without context to indicate that it is specifically mahjong being referred to. Even still, one instance refers to "three shop owners who often played cards at the stalls near tenth street", which is not near the mahjong room on fifth or sixth street.
- two Chinese-language news reports which machine translation failed on

The only source I could read that referred to illnesses that could even be compatible with the HSM mahjong room and in a relevant time period is the New Yorker article; I have quoted the relevant portion of it in full. It is also the only source that makes specific reference to the HSM mahjong room.

While it is certainly possible that the HSM mahjong room played a role in the HSM outbreak, there are no first-hand accounts of mahjong players with covid, no non-anecdotal accounts, and no known specific individuals who played mahjong and acquired covid. The only evidence is a single sentence published almost a year after the events in question with a fourth-hand anecdotal account of non-specific people.

A.4 Missing N-glycans

I lacked sufficient context to evaluate whether the missing N-glycans were strange, and if so, how unlikely they were to arise from random chance. Rootclaim only gave it a Bayes factor of 1.5 anyhow.

A.5 Early mutation rate

Naively one expects that the mutation rate of a virus would be greatest immediately after jumping species. In principle this could be used to distinguish a zoonotic origin from a virus that had been pre-adapted to a human host in the lab.

In the proposal for Project DEFUSE, viruses would be evaluated based on disease they cause in "humanized mice", ie mice with human ACE2 receptors [14]. While the result of this filtering process would be *better* adapted to humans than wild viruses, the effect would presumably be confined to the receptor binding domain of sars-cov-2; the rest of the virus would not be selected for being specialized for human hosts. Therefore I do not expect the early mutation rate to be useful in distinguishing the hypotheses without significant and very technical investigation.

B Inventory of earliest known covid cases

To be the best of my knowledge, there does not exist any publicly available list of the earliest known covid cases. The best source, and nearly the *only* source, on early covid cases is the 2021 WHO report [32] and its annex. This report does not have a complete list of early cases, but does give aggregate data on known 2019 cases broken down by date and other features, as well as blurry maps of cases, which with some effort can be used to infer some of the individual case information. This aggregate data is derived from extensive searching conducted by the WHO team, including interviews with many early patients, which cannot be independently verified or duplicated.

While the data in the WHO report is irreplaceable, it is not error free, nor is it complete. Notably other researchers have found that the WHO's primary case actually had on onset of 2019 December 16, not 8; and the report omits Wei Guixian with an onset of December 10, making her the primary covid case.

We begin by delineating the information in the WHO report, before incorporating other sources. This separation is useful for several reasons:

- I am not aware of any other place that presents the WHO data in tabular form.
- The WHO report is the highest quality source, so mixing other sources may introduce quality issues.
- It is relatively easy to understand the causes of ascertainment bias in the WHO report, but once a diversity of sources are mixed it is much harder to control ascertainment bias, as even the selection of sources[75] can be a cause of bias.

B.1 WHO cases by date and connection to HSM

Aggregate case data is presented as bar graphs found in the WHO report [32], annex pages 146, 152, and 179 (and others). In total the WHO identifies 174 known cases in 2019 December. For each date in December, I have listed the number of cases with that onset date according to whether WHO says they have an established connection to HSM, with those with connections broken down into vendors who have a fixed shop at HSM and visitors who do not. ("Visitor" includes 23 people who became infected at HSM and 2 people who became infected indirectly. "Unknown" presumably means that the patient was unavailable for interview.)

[75] including publication bias, which is largely invisible

date (2019 Dec)	vendor	visitor	non-HSM	unknown
8			1	
9				
10				
11		1		
12		1	2	
13	1			
14				
15			3	
16			2	
17	1	1	3	1
18	1		4	
19	2		1	
20	2	4	5	
21	2		2	
22		1	4	
23	6	1	7	
24		2	6	1
25	7	2	8	
26	2		6	1
27	2	2	12	
28	3		11	
29		2	10	2
30		4	16	1
31	1	4	10	
total	30	25	113	6

B.2 Map of HSM, west half

I have only found one map of HSM which labels the individual stalls; I have included the west half in figure 12.

Shown in figure 12 is the west half of the HSM market. The market is organized into streets or alleys which are numbered 1 to 15; each alley has many individual shops. Here the numbers are given on a regular grid, but in the WHO report some shops span multiple grid cells, and thus presumably have multiple numbers. I have used the locations of cases as marked on the WHO map and compared those locations with this map to identify the stall numbers of those cases; the results were in quite good agreement with other sources that reported early cases by stall numbers. I think this map is erroneous in prefixing an 'X' on stall numbers in alley 13. I think this map is not to scale. Orange-red shading are locations with environmental samples positive for covid; sky blue for those with environmental samples which were all negative. Unshaded locations had no environmental samples. This is taken from figure S1B of [12].

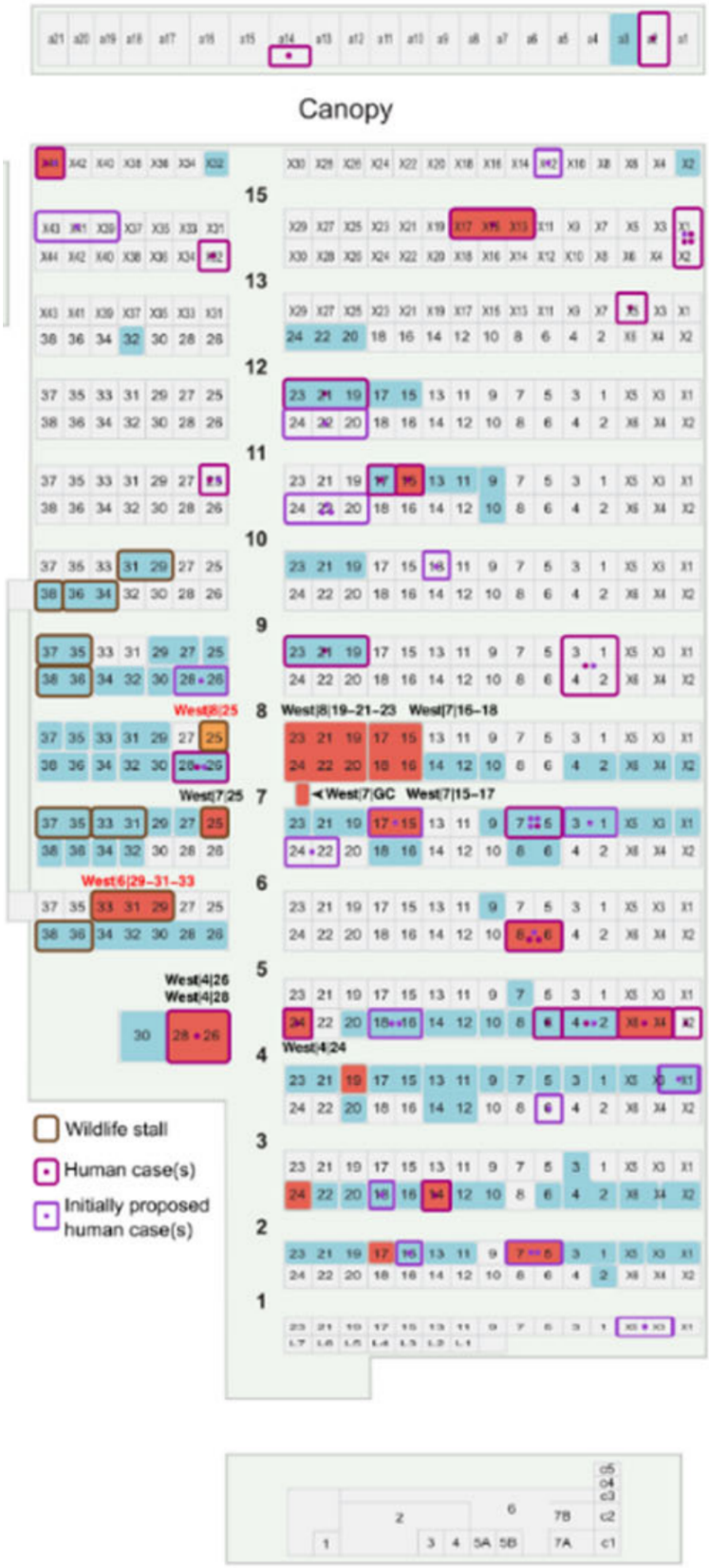

Figure 12: See text of section B.2.

B.3 Vendor cases

Of the 55 HSM-linked cases in the WHO report, 53 were acquired at HSM and 2 acquired indirectly; 30 of these were vendors who work at fixed shops in HSM (annex page 177). (Some of the other 23 were also HSM employees or regular visitors.) With much squinting at the map on annex page 181, and comparing to figure S2 of [12], the locations of each of the 30 vendor cases can be identified. I have consolidated this with the HSM vendor cases found by Dr Marion Koopmans from the data file of [12], located at `https://github.com/sars-cov-2-origins/huanan-market-environment/blob/main/metadata/cases.csv`. (She presumably had access to this data due to being on the 2021 WHO investigative team, specifically as the head of the molecular epidemiology subteam.) Date ranges can be narrowed in some cases by cross referencing information in other parts of the WHO report.

Note that the case data from Dr Koopmans is simply an unannotated csv file with no context or discussion that I found in a github repository linked to from a paper she co-authored. Thus the data has no implied warranty of validity – it could easily be some draft or working data that was accidentally uploaded. None-the-less it does match with case maps she has made in public presentations.

In the following table, each shop location is given as the alley followed by the number within that alley. Most shops are in the west half of HSM; shops in east half are indicated with an "E". Some shops occupy multiple numbers, in which case the numbers will be comma separated. A superscript indicates multiple cases at the same shop in the same date range.

date (2019 Dec)	stall number	source	notes
13	12:19,21,23	WHO+MK	annex page 179; WHO has this more specifically at 12:21
15 or earlier	2:14	MK	Wei Guixian, Dec 10, first known case
15 or earlier	6:22,24	MK	
15 or earlier	7:5,7^2	MK	
15 or earlier	7:15,17	MK	
15 or earlier	7:26,28^2	WHO+MK	WHO has only one case, in Dec 21 to 27
15 or earlier	8:26,28	MK	
15 or earlier	10:22	MK	
17	13:X1,15:2	WHO+MK	cluster 2, annex page 157
18 to 20	7:5,7	WHO+MK	
18 to 20	15:13,15,17	WHO+MK	MK has this more specifically at 15:15
18 to 20	back street:2	WHO+MK	
19	13:X1,15:2	WHO+MK	cluster 2, annex page 157
20	5:6,8	WHO+MK	cluster 4, annex page 158
21 to 27	4:6	WHO+MK	
21 to 27	4:24	WHO+MK	
21 to 27	4:26,28	WHO+MK	
21 to 27	4:X4,X6	WHO+MK	
21 to 27	9:19,21,23	WHO+MK	MK has this more specifically at 9:21
21 to 27	11:15	WHO+MK	
21 to 27	11:17	WHO+MK	
21 to 27	11:25	WHO+MK	
21 to 27	13:X5	WHO+MK	
21 to 27	13:X32	WHO	according to [12], the WHO has this in 13:26 instead
21 to 27	E2:1	WHO+MK	
21 to 27	E3:5,7,9	WHO+MK	MK has this more specifically at E3:7
21 to 27	E7:3	WHO	
21 to 27	E7:15,17	WHO+MK	
21 to 27	E11:42	WHO	
25	13:X1,15:2	WHO+MK	cluster 2, annex page 157
26	5:6,8	WHO+MK	cluster 4, annex page 158

(continued)

date (2019 Dec)	stall number	source	notes
16 to 31	E8:19	MK	
16 to 31	E9:46	MK	
16 to 31	1:X3,X5	MK	
16 to 31	$2{:}5{,}7^2$	MK	
16 to 31	2:15	MK	
16 to 31	2:18	MK	
16 to 31	3:6	MK	
16 to 31	$4{:}2{,}4^2$	WHO+MK	WHO has only one case, in Dec 28 to 31
16 to 31	$4{:}16{,}18^2$	MK	
16 to 31	4:X1	MK	
16 to 31	5:6,8	MK	presumably also part of cluster 4
16 to 31	7:1,3	MK	
16 to 31	7:5,7	MK	
16 to 31	$8{:}2{,}8{:}4{,}9{:}1{,}9{:}3^2$	WHO+MK	WHO has only one case, in Dec 28 to 31
16 to 31	10:13	MK	
16 to 31	$10{:}22^2$	MK	
16 to 31	11:22	MK	
16 to 31	13:26	MK	possibly same as WHO case in 13:X32
16 to 31	13:X1,15:2	MK	presumably also part of cluster 2
16 to 31	15:12	MK	
16 to 31	15:35	MK	
16 to 31	$15{:}44^2$	WHO+MK	WHO has only one case, in Dec 21 to 27
28 to 31	4:X2	WHO+MK	
28 to 31	back street:14	WHO+MK	

In total this constitutes 62 cases, assuming the 13:X32 and 13:26 cases are the same.

C Excluded information

Crucially for fairness of the debate, I was not permitted to base my decision on independent research beyond what the participants of the debate did. This gives the participants opportunity to identify flaws or rebut any information presented against their side. However for the purposes of the exposition within this report I have not so limited myself, mostly for the purposes of filling in minor details that do not influence the debate outcome.[76] For example, while both sides of the debate estimated there were about 1000 vendors at HSM, the WHO

[76]For example, at times when the debate participants cited secondary sources for a statement I would follow references back to a primary source to validate it before including and citing it in this document.

report annex page 182 [32] gives a more exact figure of 1162 vendors, which I freely use instead of 1000.

However some of the information I encountered does appear to directly bear on the central topic of the debate; I did not want to exclude it from the report, but the reader should realize it has been subjected to a lower level of critical scrutiny than the rest of the report, and may exhibit selection bias.

(What does it mean to "exclude" information? I have generally been using any material found in any reference from either party, so by that standard something in the annexes of the WHO report should be included. However, amongst the thousands of pages of reference material, it seems likely that one or both parties could have overlooked something 100 pages deep into a supplementary materials section. What is important to me is the balance between the level of scrutiny a claim received in the debate and how contentious that claim is; highly contentious claims that faced no scrutiny will be treated with extreme skepticism. This is not a binary distinction where evidence is treated with full credulity just because it lies inside some fixed boundary line.)

C.1 Anecdotal reports of the conditions at WIV

Annex D7 of the WHO 2021 report presents second-hand information reported from members of the WIV ([32], page 130 to 133 of annex). I will quote the relevant passages at length.

Professor Shi Zhengli (director of WIV Center for Emerging Infectious Diseases) was asked about the laboratory work performed at WIV:

> -her laboratory used recombinant viruses to test whether bat CoVs could use ACE2 to bind but used bat spike protein on a bat-CoV backbone, not human SARS. It is important to use this approach because it is difficult to isolate these viruses and envelope protein is most important part to understand function. For example, other researchers engineered SHC014 spike so did not need isolates for mouse experiments. WIV began recombinant work in 2015 with WIV-1. It received ACE2 mice in 2016 and started recombinant experiments with WIV +SHC014 in 2018 but did not finish them owing to the COVID outbreak.

Note how this compares to the work listed in the Project DEFUSE proposal, in which chimeric viruses would install experimental spike proteins into non-pandemic virus backbones, and test them with mice bearing human ACE2 receptors.

The WIV director (Professor Yuan Zhiming?) discussed the possibility of sars-cov-2 escaping from WIV:

> Staff had to report any symptoms every day after the outbreak of COVID began. Serum samples were preserved annually for laboratory staff. There was extra testing during COVID outbreak according to the Yuang Zhiming (laboratory director). The Institute did not

> respond to conspiracy theories but understood why the WHO team needed to ask. There had been no reports of unusual diseases, none diagnosed, and all staff tested negative for SARS-CoV-2 antibodies.

It is unclear to me if "Yuang Zhiming" is a typo, a different spelling of the same name, or a different person. If it is true that *all* staff at WIV were tested and none had sars-cov-2 antibodies, that would make a lab leak unlikely.

Members of the WIV insist on proper adherence to laboratory safety:

> With regards to questions about laboratory workers, all underwent a strict training regime that includes three levels with strict rules on number of hours training and in-laboratory experience prior to being allowed on own in lab, or to supervise others. P4 staff also undergo psychological evaluation before being allowed to work in the laboratory. Physical and mental health was monitored; no unusual respiratory infections had been noted in the previous year. Good compliance with mask use and hand hygiene was observed. Surveillance during the outbreak had been stringent; no suspected or confirmed case of COVID-19 was seen by PCR and antibody testing of all staff was negative. (If any worker had been infected, it would have been likely that close contacts would have shown signs of infection.) Sera were tested twice a year, and all had been negative. There had been no turnover of staff in the coronavirus team.

A rule of thumb I personally suggest when critically interpretting a passage such as the above is to simply ignore so-called 'weasel'-words like "good" or "strict"; eg this yields "Compliance with mask use and hand hygiene was observed. Surveillance during the outbreak existed...". People, even when trying to twist things to their advantage, generally don't deliberately lie about unambiguous things like whether an antibody test was positive or negative, but even well-intentioned people will frequently imagine any training regime to be a strict training regime if that is favorable to them. It is also easy for vague qualifiers to be added as information is passed along like a game of telephone; a first-hand account of a "training regime" becomes a second-hand "strict training regime", etc.

Everyone believes they have high safety standards; what is important in the above are any specific, tangible claims which are more likely than vague platitudes to have a basis in truth.

Yuan Zhiming denied the possibility of a lab leak from Wuhan:

> The rumours of a leak from the laboratory were refuted categorically by the laboratory director for the following reasons:
>
> - among the three SARS-like viruses cultured in the laboratory, none are closely related to SARS-CoV-2. The only SARS-CoV-2-like virus found by this group is RaTG13, which is neither a live (cultured) virus nor the progenitor of SARS-CoV-2
> - [...] The reserved sera in April 2019 and March 2020 from all

the workers and students in research group led by Professor Shi Zhengli were seronegative for SARS-CoV-2 antibodies.

- The laboratory director, responding to laboratory-leak theories, commented that from 2010, including the P3 laboratory, WIV has conducted experiments with more than 10,000 entries, and the P4 laboratory has conducted experiments with more than 3,000 entries in the last 3 years. No infection was ever reported. Close contacts would have been infected if there had been a laboratory leak. But serum samples from Professor Shi's team were all negative.

The second bullet point above is quite specific, and thus unlikely to be simple fabrication. However fabrication seems like the best explanation for how the text of the WHO report could be compatible with a lab leak from the WIV; other possibilities are a false negative for antibodies, the screening omitted some lab workers, the primary case was not in Shi's group, or perhaps some explanation I cannot think of.

Prior to the release of the WHO report, Peter Daszak gave a presentation on the WHO mission which included a few more tiny details not in the above. On slide 16 of `https://downloads.vanityfair.com/ecohealth-alliance/peter-daszak-powerpoint.pdf` he writes, regarding the WIV, "Illnesses (4/700 during COVID; no cases in Zhengli's group)". As for relevant work conducted at the WIV, he has "Recombinant work (ACE2 mice began 2016; WIV/SHC014 began in 2018)", compatible with the text of the WHO report.

C.2 To play cards or to play mahjong

After the debate decision was finalized, I asked several native Chinese speakers about the translation of "打牌". One said that it could refer to playing any of a variety of game pieces including cards or mahjong tiles, and context would be required to distinguish them. The others (separately) said that they have exclusively heard it in reference to playing cards, but several found that according to a dictionary it could also mean to play mahjong. Note that there does exist a Chinese word for mahjong (where the English word comes from, of course), which as far as I could tell was not in the relevant Chinese news articles.

D Glossary

- ACE2: angiotensin-converting enzyme 2, a receptor found in lungs, intestines, and throughout the body in humans and other animals. The spike protein of sars-cov-2 binds to ACE2 to enter human cells.
- amino acid: the smallest indivisible component of a protein. There are 20 standard amino acids in humans, denoted by different capital letters.
- Bayes factor: how much more likely to make an observation if one hypothesis is true than if another hypothesis is true, ie, $P(O|A)/P(O|B)$

- codon: a sequence of three nucleotides. When a protein is built, each codon yields a single amino acid according to a fixed rule (eg `CGG` yields `R`, arginine)
- covid, or covid-19: coronavirus disease 2019, the disease caused by the sars-cov-2 virus
- D614G, etc: this notation describes a mutation; the number 614 is the location of the mutation. The letter(s) before the number gives the ancestral amino acid(s) at that location, and after the number gives the mutated amino acid(s). Similar notation can be used with nucleotides. The D614G mutation in particular occurred in early 2020 and may have greatly increased the infectivity of sars-cov-2.
- DEFUSE: a 2018 research proposal that involves modifying viruses related to sars-cov-1; one of the involved research groups is at WIV
- FCS: furin cleavage site, a sequence of four amino acids in a protein typically of the form RxxR that is cut by the furin protein
- GoF: gain-of-function research, for example modifying a virus by artificially adding an FCS
- HSM: Huanan Seafood Market
- index case: the first *detected* case of a disease, not to be confused with primary case
- nucleotide: the smallest indivisible component of DNA and RNA. There are five standard nucleotides, denoted by `A`, `C`, `G`, or `T`/`U`.
- primary case: the first case of a disease; often never detected
- RBD: receptor binding domain, the part of the coronavirus spike protein that binds to ACE2 (or other receptors)
- sars-cov-1: the virus that causes Severe Acute Respiratory Syndrome (SARS), which infected 8110 and killed 811 people in 2002 to 2004, mostly in China and Hong Kong. While much more lethal than covid-19, it does not have significant pre-symptomatic transmission.
- sars-cov-2: the virus that causes covid-19
- SSE: superspreader event, here usually referring to the initial cluster of cases associated with the HSM in 2019 December
- resident: Throughout the discussion, we use "resident" at HSM to mean not only vendors who work at a fixed shop in HSM but also any person who goes to HSM on a daily basis; this would include other staff and some shoppers. Such people return to HSM often enough that they might acquire an illness there and then later transmit it to others at HSM.
- WIV: Wuhan Institute of Virology
- zoonosis: a human disease transmitted by or from animals

References

[1] Covid origins debate 2023 session 1, video part A. At the time of this writing, the recordings of the debate are not publicly available.

[2] Covid origins debate 2023 session 1, video part B.

[3] Covid origins debate 2023 session 2, video part A.

[4] Covid origins debate 2023 session 2, video part B.

[5] Covid origins debate 2023 session 3, video part A.

[6] Covid origins debate 2023 session 3, video part B.

[7] Wilf, S. 2023. Written supplements to debate, private communication.

[8] Miller, P. 2023. Written supplements to debate, private communication.

[9] Allen, T., Murray, K.A., Zambrana-Torrelio, C. et al. Global hotspots and correlates of emerging zoonotic diseases. Nat Commun 8, 1124 (2017). https://doi.org/10.1038/s41467-017-00923-8

[10] Carleton, Tamma, et al. "Global evidence for ultraviolet radiation decreasing COVID-19 growth rates." Proceedings of the National Academy of Sciences 118.1 (2021): e2012370118. https://www.pnas.org/doi/full/10.1073/pnas.2012370118

[11] Cauchemez, Simon, et al. "Unraveling the drivers of MERS-CoV transmission." Proceedings of the national academy of sciences 113.32 (2016): 9081-9086.

[12] Crits-Christoph, Alex, Karthik Gangavarapu, Jonathan Pekar, Niema Moshiri, Reema Singh, Joshua Levy, Stephen Goldstein et al. "Genetic evidence of susceptible wildlife in SARS-CoV-2 positive samples at the Huanan Wholesale Seafood Market, Wuhan: Analysis and interpretation of data released by the Chinese Center for Disease Control." (2023). https://doi.org/10.5281/zenodo.7754299

[13] Centers for Disease Control and Prevention (CDC) (2003). Prevalence of IgG antibody to SARS-associated coronavirus in animal traders – Guangdong Province, China, 2003. MMWR Morb. Mortal. Wkly. Rep. 52, 986–987. https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5241a2.htm

[14] EcoHealth Alliance. Project DEFUSE: Defusing the Threat of Bat-borne Coronaviruses. (2018) https://s3.documentcloud.org/documents/21066966/defuse-proposal.pdf

[15] Feld, Scott L. "Why your friends have more friends than you do." American journal of sociology 96.6 (1991): 1464-1477.

[16] Feng, M., Ling, Q., Xiong, J., Manyande, A., Xu, W., Xiang, B. Geographical and Epidemiological Characteristics of Sporadic Coronavirus Disease 2019 Outbreaks From June to December 2020 in China: An Overview of Environment-To-Human Transmis-

sion Events. Frontiers in medicine, 8, 654422 (2021). https://doi.org/10.3389/fmed.2021.654422

[17] Geldenhuys M, Mortlock M, Weyer J, Bezuidt O, Seamark ECJ, et al. A metagenomic viral discovery approach identifies potential zoonotic and novel mammalian viruses in Neoromicia bats within South Africa. PLOS ONE 13(3): e0194527 (2018). https://doi.org/10.1371/journal.pone.0194527

[18] Hessler, Peter. "Nine Days in Wuhan, the Ground Zero of the Coronavirus Pandemic." The New Yorker (2020). https://www.newyorker.com/magazine/2020/10/12/nine-days-in-wuhan-the-ground-zero-of-the-coronavirus-pandemic

[19] Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLOS Pathogens 13(11): e1006698. (2017) https://doi.org/10.1371/journal.ppat.1006698

[20] Lerner, Sharon, and Hibbett, Maia. "Leaked Grant Proposal Details High-Risk Coronavirus Research." The Intercept (2021). https://theintercept.com/2021/09/23/coronavirus-research-grant-darpa/

[21] Li, Qun, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren et al. "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia." New England journal of medicine 382, no. 13 (2020): 1199-1207. https://doi.org/10.1056/NEJMoa2001316

[22] Liu, W.J., Liu, P., Lei, W. et al. Surveillance of SARS-CoV-2 at the Huanan Seafood Market. Nature (2023). https://doi.org/10.1038/s41586-023-06043-2

[23] Liu, W., Wang, D., Hua, S. et al. Spatiotemporal analysis of COVID-19 outbreaks in Wuhan, China. Sci Rep 11, 13648 (2021). https://doi.org/10.1038/s41598-021-93020-2

[24] Moshiri, Niema, et al. FAVITES: simultaneous simulation of transmission networks, phylogenetic trees and sequences. Bioinformatics 35.11 (2019): 1852-1861.

[25] NCBI staff. Sars-cov-2 Wuhan-Hu-1 complete genome. 2020. https://www.ncbi.nlm.nih.gov/nuccore/1798174254

[26] Newey, Sarah. "Wuhan officials identified Huanan market as a pandemic risk at least five years before Covid emerged." The Telegraph (2021). https://www.telegraph.co.uk/global-health/science-and-disease/wuhan-officials-identified-huanan-market-pandemic-risk-least/

[27] Pekar, J.E. et al. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science 377, 960 - 966 (2022). https://doi.org/10.1126/science.abp8337

[28] Piplani, S., Singh, P.K., Winkler, D.A. et al. In silico comparison of SARS-CoV-2 spike protein-ACE2 binding affinities across species and implications for virus origin. Sci Rep 11, 13063 (2021). https://doi.org/10.1038/s41598-021-92388-5

[29] Rambaut, A., Holmes, E.C., O'Toole, Á. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 5, 1403–1407 (2020). https://doi.org/10.1038/s41564-020-0770-5

[30] Temmam, S., Vongphayloth, K., Baquero, E. et al. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 604, 330–336 (2022). https://doi.org/10.1038/s41586-022-04532-4

[31] Twain, M. Life on the Mississippi. (1883) https://www.gutenberg.org/files/245/245-h/245-h.htm

[32] World Health Organization. "WHO-convened global study of origins of SARS-CoV-2: China Part." (2021). https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part

[33] Worobey, Michael. "Dissecting the early COVID-19 cases in Wuhan." Science 374.6572 (2021): 1202-1204. https://doi.org/10.1126/science.abm4454

[34] Worobey, M., Levy, J. I., Malpica Serrano, L., Crits-Christoph, A., Pekar, J. E., Goldstein, S. A., Rasmussen, A. L., Kraemer, M. U. G., Newman, C., Koopmans, M. P. G., Suchard, M. A., Wertheim, J. O., Lemey, P., Robertson, D. L., Garry, R. F., Holmes, E. C., Rambaut, A., & Andersen, K. G. (2022). The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. Science (New York, N.Y.), 377(6609), 951–959. https://doi.org/10.1126/science.abp8715

[35] Jiuyang Xu, Wenxu Jia, Pengfei Wang, Senyan Zhang, Xuanling Shi, Xinquan Wang, Linqi Zhang. Antibodies and vaccines against Middle East respiratory syndrome coronavirus, Emerging Microbes & Infections, 8:1, 841-856 (2019), https://doi.org/10.1080/22221751.2019.1624482

[36] Yang, C., Zhang, S., Lu, S., Yang, J., Cheng, Y., Liu, Y., Zhao, L., Gong, J., Xu, J. All five COVID-19 outbreaks during epidemic period of 2020/2021 in China were instigated by asymptomatic or pre-symptomatic individuals. Journal of biosafety and biosecurity, 3(1), 35–40. (2021). https://doi.org/10.1016/j.jobb.2021.04.001

[37] Zeng, Lei-Ping, et al. "Bat severe acute respiratory syndrome-like coronavirus WIV1 encodes an extra accessory protein, ORFX, involved in modulation of the host immune response." Journal of virology 90.14 (2016): 6573-6582. https://doi.org/10.1128/jvi.03079-15

[38] Zhang Xiaomin. "Dalian infections linked to produce markets, buses." China Daily (2022 November 2). https://www.chinadaily.com.cn/a/202209/02/WS6311c66ea310fd2b29e75b12.html

[39] Zhang, Y.-Z., and Holmes, E.C. (2020). A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. Cell 181, 223–227. https://doi.org/10.1016/j.cell.2020.03.035