RETROSPECTIVE ANALYSIS OF COVID-19 HOSPITALIZATION MODELLING SCENARIOS WHICH GUIDED POLICY RESPONSE IN FRANCE

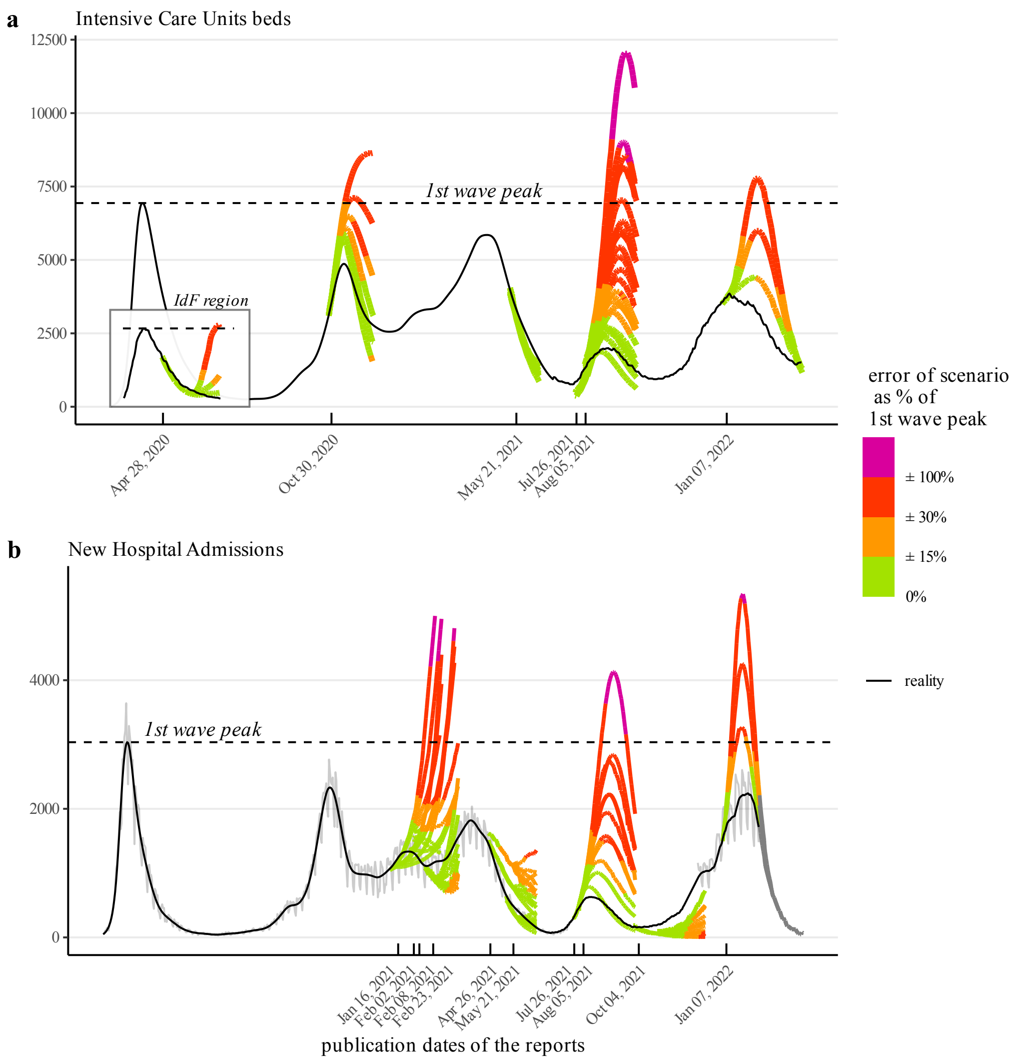
Thomas Starck[1] and Maxime Langevin[2]

[1,2]Independent researcher

December 16, 2023

ABSTRACT

Epidemiological modelling has played a key role in proposing, analyzing and justifying non-pharmaceuticals interventions in response to the COVID-19 pandemic. Despite its importance, evaluations of models' ability to accurately anticipate the evolution of the disease remain scarce. Thus, robust, systematic, and pre-specified evaluation criteria are needed to assess the relevance of modelling scenarios that guided policy response during the pandemic. We conduct a retrospective assessment of modelling reports which guided policy response in France from April 2020 to April 2022. After systematically verifying the scenarios hypotheses (e.g., exclusion of no-lockdown scenarios when a lockdown was effectively in place), we find that epidemiological models were (a) uncertain, (b) unaccurate, and (c) biased towards an overestimation of predicted COVID-19 related hospitalizations. In more than half of the reports, reality is below or equal to even the best-case scenario. To our knowledge, this is the only national systematic retrospective assessment of COVID-19 pandemic scenarios; such an approach should be reproduced in other countries whenever possible.



Graphical Abstract: Comparison of Pasteur Institute's scenarios to reality. Forecasting errors of scenarios (colors) compared to reality (black line) are expressed as a percentage of the maximum Intensive Care Units occupancy reached during the covid-19 pandemic in France (horizontal dashed line).

1 Introduction

The COVID-19 pandemic, as well as restrictions imposed to limit the spread of the virus, have had an unprecedented impact on global health, economy, and society. The COVID-19 pandemic still remains a public health emergency of international concern McVernon and Liberman [2023].

Models have been central to inform decision-making. For instance, two thirds of the $\approx$ 150 studies mapped by the UK Health Security Agency to assess the effectiveness of non pharmaceutical interventions are based on modelling UKHSA [2023]. Models are also used to retrospectively review the effectiveness of non pharmaceutical interventions with counterfactual scenarios Roux et al. [2023], Flaxman et al. [2020], and are known to influence the societal debate through their influence on policy-makers Sanchez [2021].

While intended to slow virus spread, the policies influenced by these models can also have harmful consequences such as global hunger FAO et al. [2023], routine child immunization disruption Chakrabarti et al. [2023], WHO and UNICEF [2020] and mental health issues Léon et al. [2023], disproportionately affecting the most vulnerable groups Li et al. [2023] and the young UNICEF [2021]. It is therefore essential that models correctly anticipate epidemic spread and accurately assess non pharmaceutical interventions effectiveness in order not to bias the political trade-offs: overestimating epidemic spread is not devoid of harms.

In most European countries, modelling teams have been set up to inform policy making during the COVID-19 pandemic Jit et al. [2023], Eker [2020]. Models used include statistical models, compartmental models, meta-population models, individual-based models, and geospatial models. In the context of France, INSERM and Pasteur Institute have used aged-structured compartmental models (such as SIR models) to produce prospective scenarios Di Domenico et al. [2021], Kiem et al. [2021].

Despite its unprecedented use to support large-scale policy decisions in the COVID-19 pandemic Eker [2020], modelling has usually been considered to offer a relatively "low to very low" level of evidence for pandemic preparedness WHO [2019]. This has led to arguments for cautiousness when dealing with modelling results Holmdahl and Buckee [2020], as well as calls for greater model transparency, reproducibility, and validity assessments Jin et al. [2020], Barton et al. [2020].

Epidemiological models used throughout the pandemic Gnanvi et al. [2021], such as SIR and compartmental models Kermack and McKendrick [1927], are especially known to be limited in their capacity to account for heterogeneity in population structure Zachreson et al. [2022], which can sometimes result in overestimation of disease incidence Merler et al. [2015].

Besides theoretical grounds, empirical evaluations of COVID-19 epidemiological modelling has been scarce. Some of the available empirical analysis points towards models being unable to significantly outperform simple baselines Chharia et al. [2022], Antulov-Fantulin and Böttcher [2022] or failing at predicting COVID-19 outcomes Ioannidis et al. [2022], Moein et al. [2021]. The critical role they have played thus calls for further rigorous assessment of models' ability to accurately forecast pandemic trajectories and the impact of non pharmaceutical interventions.

Empirical evaluations of COVID-19 models face several challenges. First, as explained by Nina Fefferman "in an ideal world, every epidemiological prediction of an outbreak would end up failing", as predictions would influence policy actions that would then mitigate the outbreak and falsify the key hypotheses behind the initial model, leading to predictive failure Jit et al. [2023], Holmdahl and Buckee [2020]. However, honest and careful checking of models hypotheses can ensure that the comparison of models output to reality is valid.

Furthermore, a comprehensive and systematic analysis is required in order to avoid biased cherry-picking of modelling scenarios according to whether they predicted accurately or not reality. In a related field, selective reporting of results of clinical trials has long been identified as a key source of bias. For instance, in 2000 the US Food and Drug Administration made mandatory to preregister clinical trials. This resulted in a large drop in randomized controlled trials reporting drugs benefit, from 60% to 10% Kaplan and Irvin [2015], Dickersin and Rennie [2012]. Similar requirements are needed to ensure valid analysis of COVID-19 models.

We set out to perform such an extensive systematic retrospective evaluation of epidemiological models that have informed policy-making in France. We analyze the scenarios based on these models, mainly from Pasteur Institute reports. To ensure the relevance of our retrospective, we define clear inclusion criteria and systematically check the scenarios assumptions to confirm comparability between them with reality. We assess the scenarios accuracy, uncertainty and bias. Finally, we compare the results of our systematic evaluation with the self-assessments made by the modelers.

2 Methods

The workflow for retrieving and selecting reports for our retrospective are detailed in Figure 1 and in the following subsections (sections 2.1 and 2.2). More details can be found in Supplementary Materials (Tables S1, S5 and S6). Since the numerical data underlying scenarios were not public, we explain how we extracted it from the reports' figures (sections 2.3). We finally describe our method to compare reality to scenarios and quantitatively evaluate them (section 2.2).

2.1 Scope of the retrospective

In this retrospective, we systematically focus on reports from the Pasteur Institute epidemiological modelling group, the most prominent source of modelling reports during the covid-19 pandemic in France, with major influence on policy-making. Members of the research group were also part of the French scientific council, and the modelling reports were frequently cited to support policies.

We did not systematically assess the other two main French modelling team, INSERM and ETE. They published fewer prospective scenarios, and had less political and media impact. Nonetheless, two of INSERM reports are included in our retrospective, because they were cited in joint reports on Pasteur Institute's website.

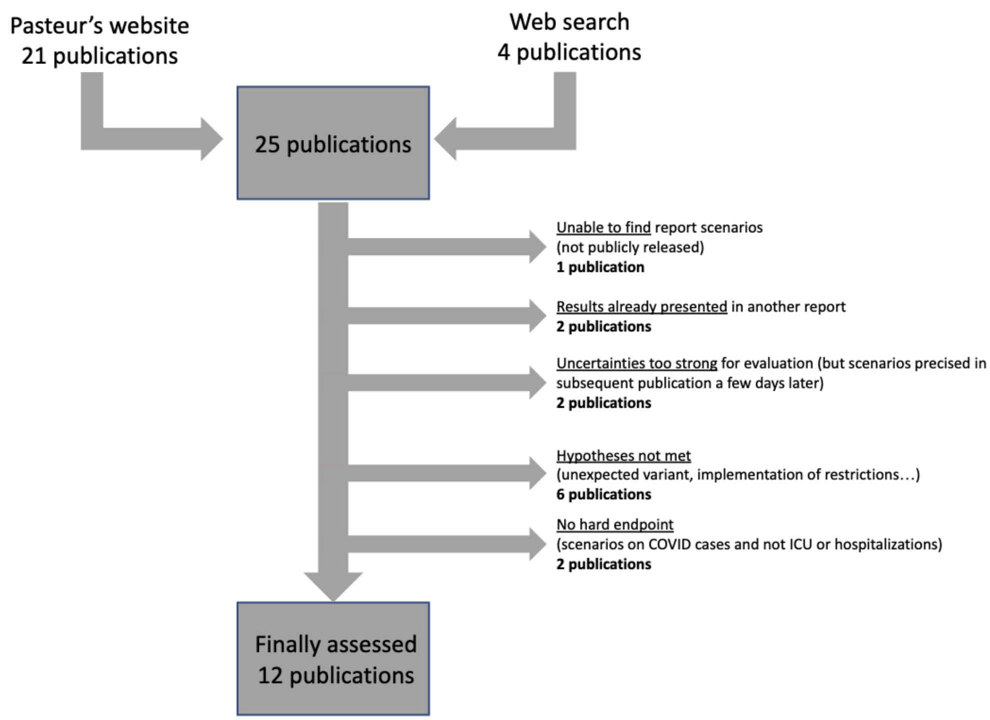


Figure 1: Workflow for the selection of the reports assessed in our retrospective.

To systematically identify all of Pasteur Institute's scenarios, we screened their website dedicated to Covid-19 modelling. We focused on medium to long term scenarios, but excluded short-term (2 weeks) projections. We distinguish between the two types of modelling because retrospective evaluation of short-term projections is more common. For instance, Paireau et al. [2022] already self-assessed Pasteur Institute short-term projections. Retrospective assessment of medium to long term scenarios is much less frequent because, as reminded on top of each of Pasteur Institute reports, "scenarios are not predictions". In this case, comparison to reality requires checking that scenarios hypotheses are verified.

All URL hyperlinks and citations of Pasteur's website were assessed. We identified 21 publications dealing with prospective scenarios. Most of them are official Pasteur Institute's reports. Three publications in 2021 (Jan 12, Jan 29 and March 11) are reports from the French Scientific Council, whose composition includes members of the modelling team. Two other publications in 2021 (Jan 16 and Feb 2) are INSERM reports cited in the Scientific Council notes and on Pasteur Institute's website.

Finally, we extended our systematic review with a web search. We searched "Pasteur scénario" and "Pasteur modélisation" on Google for each month from March 2020 to April 2022. We identified 4 other reports which were not publicly published, but whose main graphs were sometimes reported in news articles.

The total number of identified reports is thus 25.

2.2 Selection and exclusion criteria

All excluded reports and the reason underlying their exclusion are featured in the Table S5.

Of the 25 reports, one was mentioned by the press but we were unable to recover it since the French ministry of Health did not release it publicly (Oct 2020). Two of the Scientific Council reports (Jan 29, 2021 and March 11, 2021) feature results presented in reports already included in our retrospective analysis (Jan 16, 2021 and Feb 23, 2021), we therefore did not evaluate them a second time. Two reports (Jan 12, 2021 and Dec 27, 2021) have too strong uncertainties in the underlying hypotheses to verify them, precluding any assessment, and are thus not analysed; however, following reports covering their scope (Jan 16, 2021 and and Jan 7, 2022) considerably reduce those uncertainties, and are included in our analysis. This excluded 5 reports.

To perform a fair retrospective, we only select scenarios for which the hypotheses were verified in reality. For instance, we would discard a scenario which assumes a "no lockdown" situation, but where a lockdown was put in place shortly after the publication of the report. This is done for the different major non pharmaceutical interventions that could impact the modeled endpoints, such as lockdowns or curfews, but also for emergence of a new variant or vaccination rates. This alleviates a common form of circular reasoning which claims that epidemiological models were unaccurate precisely because they lead to measures that changed the underlying hypotheses. All details concerning our hypotheses verification are reported in Tables S1 and S5, and particular examples are given in Table1. Verifying these hypotheses excluded 6 publications.

Table 1: Example of hypotheses verification for scenarios inclusion or exclusion in our retrospective analysis.

Report Date	Included ?	Justification
Oct 26, 2020	no	national lockdown implemented 2 days later not considered in scenarios
Oct 30, 2020	yes	scenarios consider national lockdown announced 2 days earlier
Jul 9, 2021	no	implementation of health pass announced on July 12 not considered
Jul 26, 2021	yes	health pass implementation considered through estimations of R decrease

This yields 14 reports, whose scenarios endpoints are reported in Table 2. We restrict our retrospective study to reports modelling strong endpoints, i.e. hospitalizations and Intensive Care Units. Deaths would be another strong endpoint but is not modelled in the reports. We discard positive cases as it is an endpoint which depends heavily on the testing rate, and is less important for decision-making than Intensive Care Units and hospital caseloads. This criteria excluded 2 further publications (Feb 21, 2022 and Mar 10, 2022).

Among the remaining 12 reports, we focus on the most commonly reported endpoints, i.e. Intensive Care Units beds occupancy and hospital admissions (Table 2).

Table 2: Scenarios endpoints reported in the 14 reports kept after hypotheses validation. The last 2 reports were not included as our review focuses on hospital and Intensive Care Units (ICU), and not on positive Covid-19 cases.

Report Date	ICU beds	ICU admissions	Hospital beds	Hospital admissions	Positive Cases
April 28, 2020	x		x		
Oct 30, 2020	x	x			
Jan 16, 2021				x	
Feb 2, 2021				x	x
Feb 8, 2021				x	x
Feb 23, 2021				x	x
Apr 26, 2021				x	
May 21, 2021	x		x	x	
Jul 26, 2021	x	x		x	x
Aug 5, 2021	x	x	x	x	
Oct 4, 2021				x	
Jan 7, 2022	x		x	x	
Feb 21, 2022					x
Mar 10, 2022					x
Total	6	3	4	10	6

2.3 Data extraction and preparation

None of the reports provided their scenarios data in open data. We contacted the corresponding authors but received no reply. We finally chose to manually extract the data from the original figures, using WebPlotDigitizer `https://automeris.io/WebPlotDigitizer/`.

For each report, we first exclude the scenarios where hypotheses were not met, and then extract the remaining scenarios of interest. We also extracted the reality data available up to the report publication date (see Figure 2). This allows us to check if our manual extraction was carried out correctly, by comparing the reports reality data to French official hospitalization and Intensive Care Units data.

Most of our official reality data comes from Paireau et al. [2022], but this source stops on July 2021. This is the most reliable source for scenarios comparison to reality, since it comes directly from Pasteur Institute modelling team, and includes their own pre-processing procedure (see Paireau et al. [2022] for details). For the rest of the period, we use either French official data data.gouv [2023] or reality data manually extracted from Pasteur's reports.

2.4 Evaluating the scenarios

As illustrated in Figure 2, each report provides multiple scenarios, and no confidence interval is provided. While it is likely that some scenarios were favored as more probable during interactions between modelers and policy-makers, this information is not available. Therefore, our retrospective evaluation considers each scenario reported as equally probable, and focus on 3 particular scenarios in each report: the worst-case, the median and the best-case scenarios.

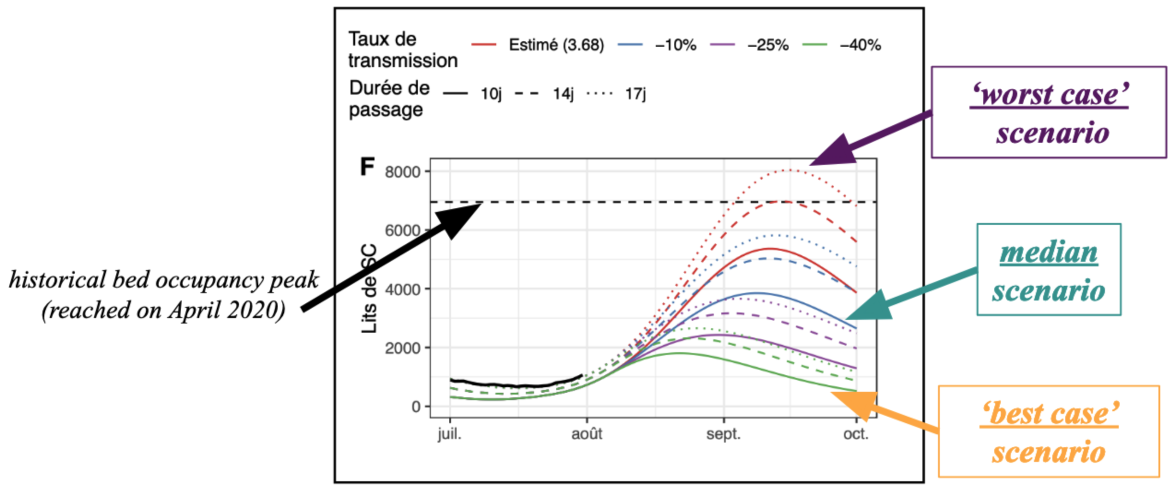


Figure 2: Screenshot example of the original scenarios published in the Aug 5^{th}, 2021 report. Some of the scenarios hypotheses are detailed in the top legend. Black solid line: real data available at publication date. Colored lines: scenarios. We indicate how are defined the median, best-case and worst-case scenarios for our study. We use the historical peak (horizontal dashed line) to normalize

We use standard quantitative metrics to asses scenarios uncertainty (section 2.4.1), accuracy (section 2.4.2) and bias (section 2.4.3). Each metric is computed over the whole report period and by 2-week periods.

2.4.1 Uncertainty

As multiple scenarios are proposed, a key aspect is the uncertainty (or conversely, the informativeness) of a report. We define the uncertainty as the difference between the values predicted by the worst-case and best-case scenarios. For each report, we compute the average uncertainty over the considered time period.

2.4.2 Accuracy

We assess accuracy for 3 representative individual scenarios in each report: the worst-case scenarios (with the highest predicted values), the best-case scenario (with the lowest predicted values), as well as the median scenario (see Figure 2).

For each of those scenario, we assess their accuracy with the following the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE).

- The Mean Absolute Error $MAE = \frac{1}{n} * \sum_{i=1}^{n} |y_{real_i} - y_{pred_i}|$ is a widely used metric for evaluating the accuracy of models. It measures the average absolute difference between the predicted values and the real values, where y_{real_i} is the actual value at the time step i, y_{pred_i} is the predicted value, and n is the number of time steps. In our case, it is expressed in number of beds per day.
- The Mean Absolute Percentage Error $MAPE = \frac{1}{n} * \sum_{i=1}^{n} 100 * \frac{|y_{real_i} - y_{pred_i}|}{y_{real_i}}$ is similar to Mean Absolute Error but is normalized with respect to the real values. It measures the average absolute difference between the predicted values and the real values divided by the real values, where y_{real_i} is the actual value at the time step i, y_{pred_i} is the predicted value, and n is the number of time steps. Unlike the Mean Absolute Error, it is expressed in percentage of the real value and is comparable across different endpoints (i.e. Intensive Care Units and new hospitalizations).

2.4.3 Bias

Previous accuracy metrics express the error in absolute terms, but do not indicate the direction of the error.

To address this limitation, we use the mean error. It indicates whether on average the scenario tended to overestimate (values >0) or to underestimate (values <0) the reality, and can thus be used to identify systematic errors in a forecast.

- The Mean Error $ME = \frac{1}{n} * \sum_{i=1}^{n} y_{real_i} - y_{pred_i}$ is a metric that measures the average difference between the predicted values and the real values. The Mean Error is positive when the scenario overestimates the reality on average and negative when it underestimates the reality on average. In our case, it is expressed in number of beds per day.

2.4.4 Normalization by historical peak

All the above metrics except Mean Absolute Percentage Error are expressed in terms of number of Intensive Care Units beds or new hospitalizations per day. We normalize these metrics by comparing them the maximum historical values reached by these two endpoints during the pandemic, based on Pasteur data : 6937 for Intensive Care Units beds occupancy and 3036 for daily new hospitalizations.

Normalizing allows to express metrics in a scale that is relevant for policy-making, as a percentage of historical peak, and to compare scenarios with different endpoints (Intensive Care Units and new hospitalizations).

All metrics are summarized in Table 3.

Table 3: Summary of metrics used for evaluating scenarios.

Assessment Goal	Metric	Computed over	Interpretation
Uncertainty	Average uncertainty	all scenarios from one report	What is the possible span of values according to the scenarios
Accuracy	Mean Absolute Error	one specific scenario (e.g. best-case, median or worst-case scenario)	Average error between the scenario and the reality
	Mean Absolute Percentage Error	one specific scenario	Average error between the scenario and the reality, with the error expressed as a percentage of real values
Bias	Mean Error	one specific scenario	Whether and how much the scenario, on average, overestimated or underestimated reality

3 Results

3.1 Qualitative comparison of scenarios to reality

3.1.1 Intensive Care Units beds

Figure 3a compares reality to the scenarios of the 6 reports anticipating Intensive Care Units beds occupancy.

In 4 reports (Apr 29, 2020; Oct 30, 2020; Jul 26, 2021 and Jan 07, 2022) reality was below the best-case scenario. For the Aug 5, 2021 report, reality corresponded to the best-case scenarios. Only one report (May 21, 2021) has a median scenario close to reality. This indicates a general systematic bias towards too pessimistic scenarios.

The May 21, 2021 report is also the only one with a low range between the best-case and the worst cas scenarios. Half of the reports (Apr 29, 2020 ; Jul 26, 2021 ; Aug 5, 2021) feature scenarios whose Intensive Care Units maximum occupancy range from close to 0 to higher than the 1st wave peak, providing little information. The two remaining reports (Oct 30, 2020 and Jan 07, 2022) have a scenarios range of about 50% of the historical peak.

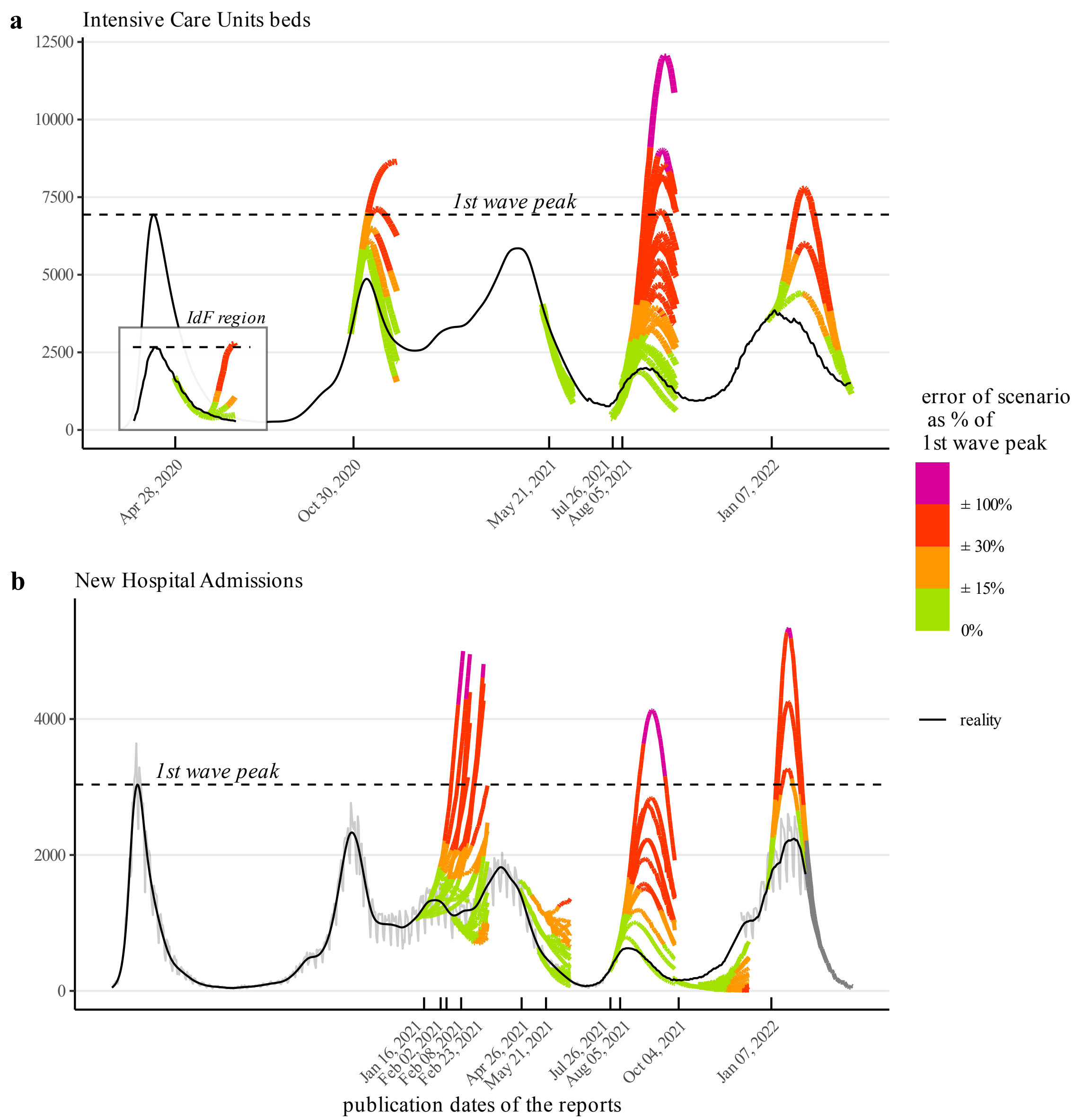


Figure 3: Comparison of Pasteur Institute's scenarios to reality during the COVID-19 pandemic for **a)** Intensive Care Units and **b)** New Hospital Admissions. Colors indicate the error between reality and scenarios, relative to 1st wave peak (horizontal dashed line). Note that an error of $\pm 15\%$ (green) means a confidence interval of 30% of the historical peak.

3.1.2 Hospital Admissions

Figure 3b compares reality to the scenarios of the 10 reports anticipating hospital admissions.

Out of these 10 reports, 5 have reality well below their best-case scenario (Jan 16, 20201; Feb 2, 2021; Apr 26, 2021; Jul 26, 2021; Jan 7, 2022) and 2 have reality slightly below the best-case scenario (Feb 8, 2021 : Aug 5, 2021). For 1 report (Oct 4, 2021), reality is above the worst-case scenario. This leaves only 2 reports (Feb 23, 2022 ; May 21, 2021) where reality is within the range of featured scenarios.

Thus, 60% of reports on hospital admissions have a scenarios range failing to capture reality. Moreover, the fact that 70% of reports have their best-case scenario equal or above reality points to a systematic bias towards too pessimistic scenarios.

3.1.3 Short-term change of the results

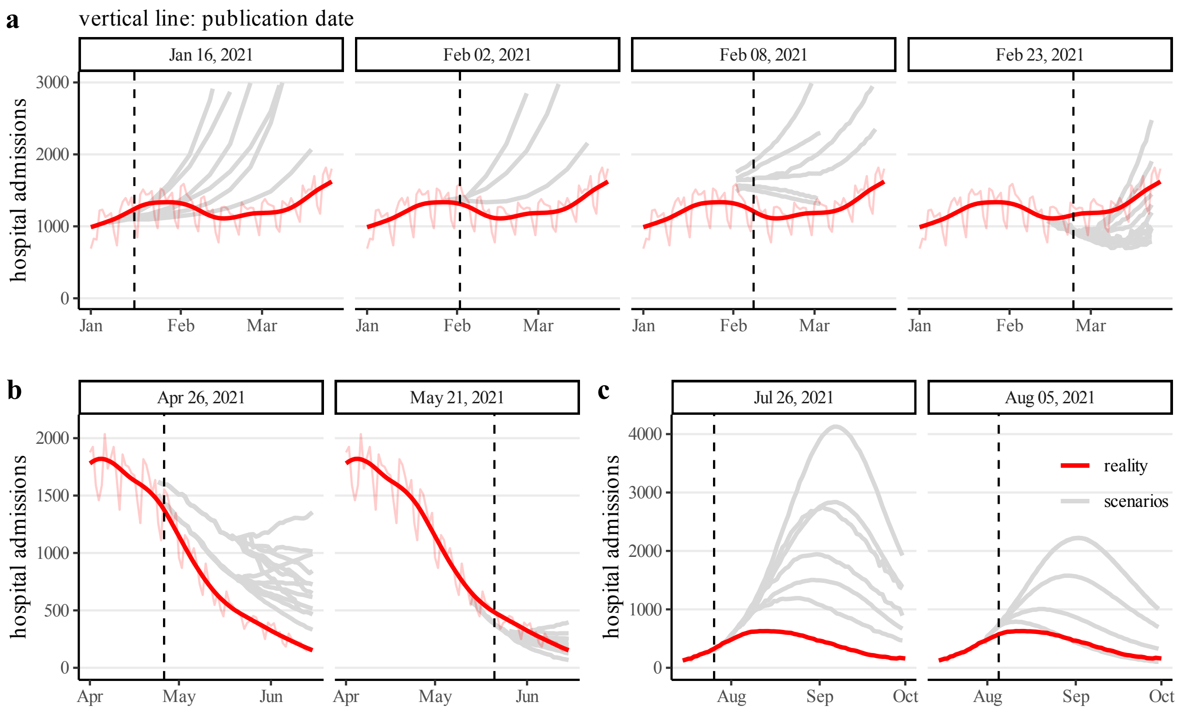


Figure 4: Comparison of scenarios variations for reports published a few weeks apart. Vertical lines indicate publication date. Red line: reality. Grey lines: scenarios. **a)** scenarios of the 4 reports during winter 2021 **b)** scenarios of the 2 reports during spring 2021 **c)** scenarios of the 2 reports during summer 2021.

In 3 instances (Jan-Feb 2021 ; Apr-May 2021 ; Jul-Aug 2021) several reports are published within a short time period, about 1 month or less. This allows to visualize how sensitive the modelers' output is to the epidemic short-term dynamic (figure 4).

For the Jan-Feb 2021 period (figure 4a). The first 2 reports (Jan 16 and Feb 2) point to unmitigated exponential growth overestimating the stagnating reality. In the 3rd report (Feb 8), a sensitivity analysis performed by the modelers suggests the possibility of both small decline to high increase, even though the main figure features an exponential growth. Only in the final re-assessment of Feb 23 does the reality fall within the scenarios range.

For the Apr-May 2021 period (figure 4b), the first report (Apr 26) presents a range of scenarios which, after a first decline, features dynamics ranging from downwards to upward trends. The range of these scenarios represent 60% of the historical 1st wave peak. Reality is below the best-case scenario. The update (May 21) features scenarios which are almost all below the best-case scenario of the previous report. Reality falls within the scenarios range of this updated report.

Finally, for Jul-Aug 2021 (figure 4c), 2 reports were published within 10 days. While the historical 1st wave peak was about 3000 daily hospital admissions, the Jul 26 scenarios ranged from 1000 to 4000; reality was half the best-case scenario. The Aug 5 update presents scenarios about 2 times smaller than the original report, and reality corresponds to the best-case scenario.

In all 3 instances, results are progressively adapted towards more optimistic results compared to the first publications. The difference between the original scenarios and the updates is never mentioned by the modelers in the following reports.

3.2 Quantitative performance assessment

In this section we present a systematic analysis of the modelling reports using the quantitative metrics presented in Table 3.

3.2.1 Uncertainty

Figure 5 shows the distribution of scenarios average uncertainty for each report published.

For Intensive Care Units beds, the median of the average uncertainty across reports is above 2000 beds at 2-4 weeks (out of a total capacity of 7000 beds). Uncertainty increases through time. After one month, half of the reports have an uncertainty close to the country ICU beds capacity, thus providing little information about the potential epidemic peak (see for instance Apr 28, July 26 and Aug 5 scenarios in 3a).

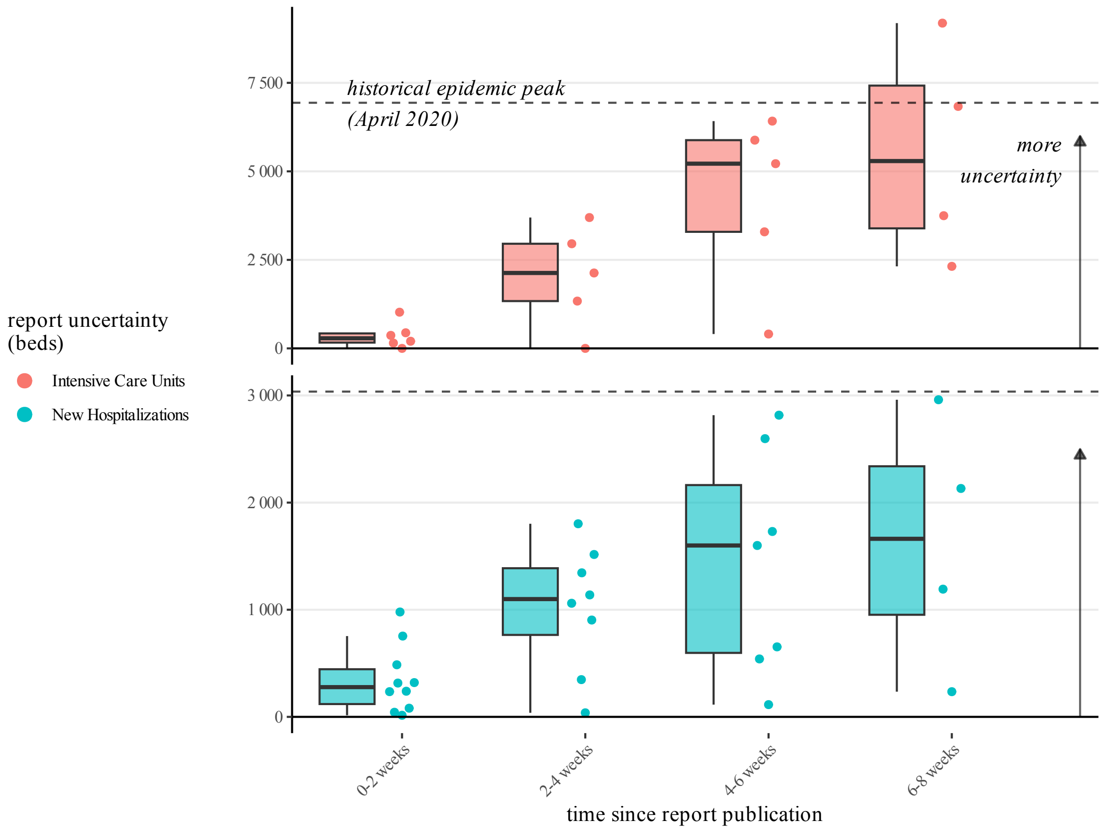


Figure 5: Average uncertainty of scenarios range in the reports published by Pasteur Institute, for each 2-week period since report publication. Top: Intensive Care Units. Bottom: New Hospitalizations.

3.2.2 Accuracy

Figure 6 shows the Mean Absolute Error of the median, worst-case and best-case scenario of each report. A complementary measure is presented in 8b as the Mean Absolute Percentage Error.

For worst-case and median scenarios, the error increases during the first month and then stagnates. The median error is then above half the 1st wave peak for worst-case scenarios, and about one third for median scenarios.

All 3 scenarios types (best-case, median, worst-case) have relatively small error for the first 2 weeks. Contrary to worst-case and median scenarios, the best-case scenarios are much closer to reality, and the error seldom increases through time.

However this must be seen in a context of high uncertainty for most reports (see previous section). At the time the reports are published, policy-makers cannot know which scenario type will be the closest to reality. Producing numerous scenarios with a large span guarantees one will be close to reality, but is of little use. That is why we analyse both uncertainty and accuracy together in the following section.

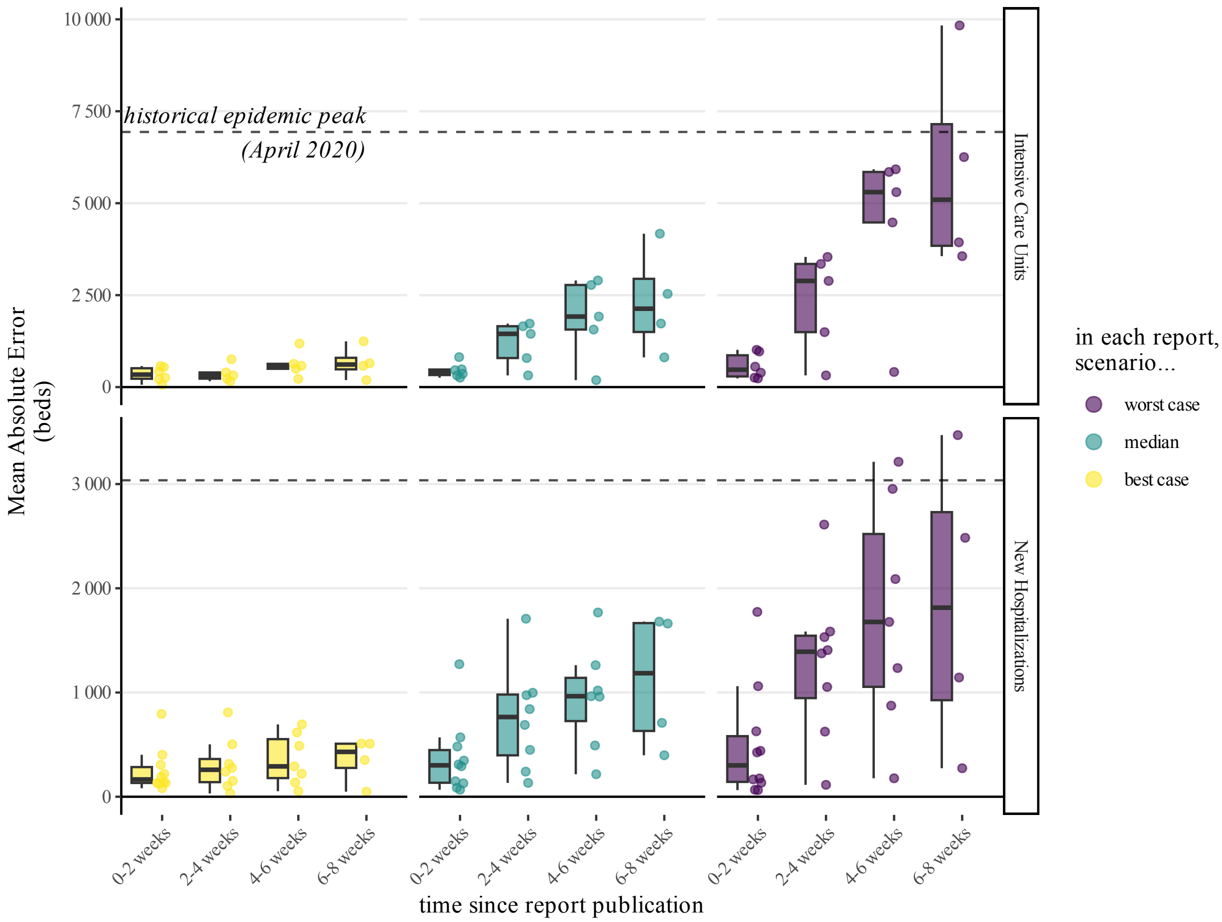


Figure 6: Accuracy of Pasteur Institute reports for the worst-case scenarios (purple), median scenarios (blue) and best-case scenarios (yellow) for intensive care units (top) and new hospitalizations (bottom). The best-case scenarios are the most accurate.

3.2.3 Uncertainty vs Accuracy

The scenarios from Jul 26 2021 in Figure 3 exemplify the dichotomy between accuracy and uncertainty. The report scenarios go from a peak smaller than any seen before to way higher than the first wave peak. Reality finally came close to the best case scenario, but a catastrophic outcome could also have been "anticipated" had it followed the worst case scenario.

Figure 7 synthesizes this dichotomy across all reports. For each 2-weeks period since report publication, it compares the accuracy of the scenarios (vertical axis) to the uncertainty of the report they are issued from (horizontal axis). For the first 2 weeks, most scenarios are both accurate and with a low uncertainty.

After the first 2 weeks, we find again the high accuracy of the best-case scenarios from previous section, but we also see that the uncertainty becomes substantive: mostly around 50% of the historical peak at 2-4 weeks, or above 50% after 4 weeks.

3.2.4 Bias

Previous error metrics focus on absolute errors, which does not indicate whether scenarios under or overestimate reality. From visual inspection (Figures 3), we see that the majority of reports overestimate the real epidemic activity, and that the errors discussed above corresponded to overestimation. To quantitatively evaluate this, we compute the Mean Error for median, best-case and worst-case scenarios.

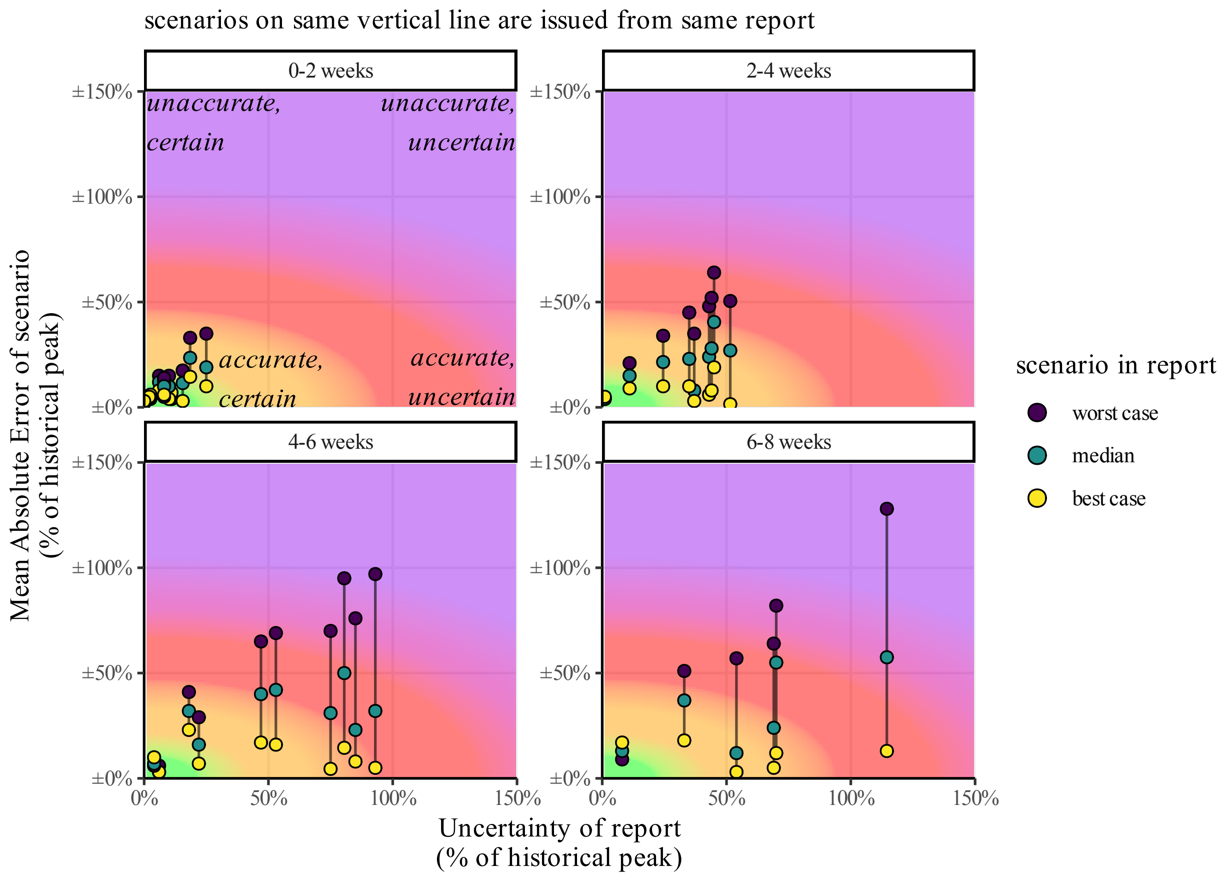


Figure 7: Uncertainty and Accuracy of scenarios. For each scenario (point), and for each 2-weeks period since report publication (panes), x-axis represents the mean uncertainty of the report and y-axis represent the Mean Absolute Error of the scenario. All values are expressed as percentage of historical peak. Scenarios on a same vertical line are issued from the same report. Note that an error of $\pm 50\%$ means a confidence interval of 100% of the historical peak: predicting a value of 50% the historical peak with $\pm 50\%$ error means reality can be anywhere between 0% and 100%.

In Figure 8a, the distribution of Mean Error across all reports is presented for the best-case, median, and worst-case scenarios. Unbiased reports, which do not consistently overestimate or underestimate the modeled endpoints, would exhibit a distribution of mean errors centered around zero for the median scenarios, while best-case and worst-case scenarios would respectively be centered around negative and positive values.

However, in our retrospective case, the median scenarios displays a bias towards overestimation. To have an unbiased assessment, one would have to focus on the best-case scenarios.

3.2.5 Modelers' own self-assessment

Out of the 14 reports where comparison to reality is appropriate after hypotheses verification (Table 2), the modelers performed a public retrospective assessment in 4 of them (figure8b). However in 2 of them the comparison is illegitimate (Figure 9).

The first illegitimate comparison concerns the Feb 8, 2021 report (performed in the Apr 26, 2021 report). A figure compares reality with a projection allegedly made on Feb 8, with an excellent matching (Figure 9a). However, the URL link pointing towards the Feb 8 projection is a dead-end, and the curve presented does not correspond to any published scenario in the actual Feb 8 report, which has more mixed results (Figure 9b). The modelers make no mention of these latter scenarios.

The second illegitimate comparison concerns the Jan 7, 2022 report (performed in a Feb 15, 2022 report). In their original report, the modelers present 2 sets of scenarios regarding vaccine efficacy against the Omicron variant. The optimistic scenarios assume a protection against Omicron infections of respectively 55% and 85% for 2 and 3 doses.

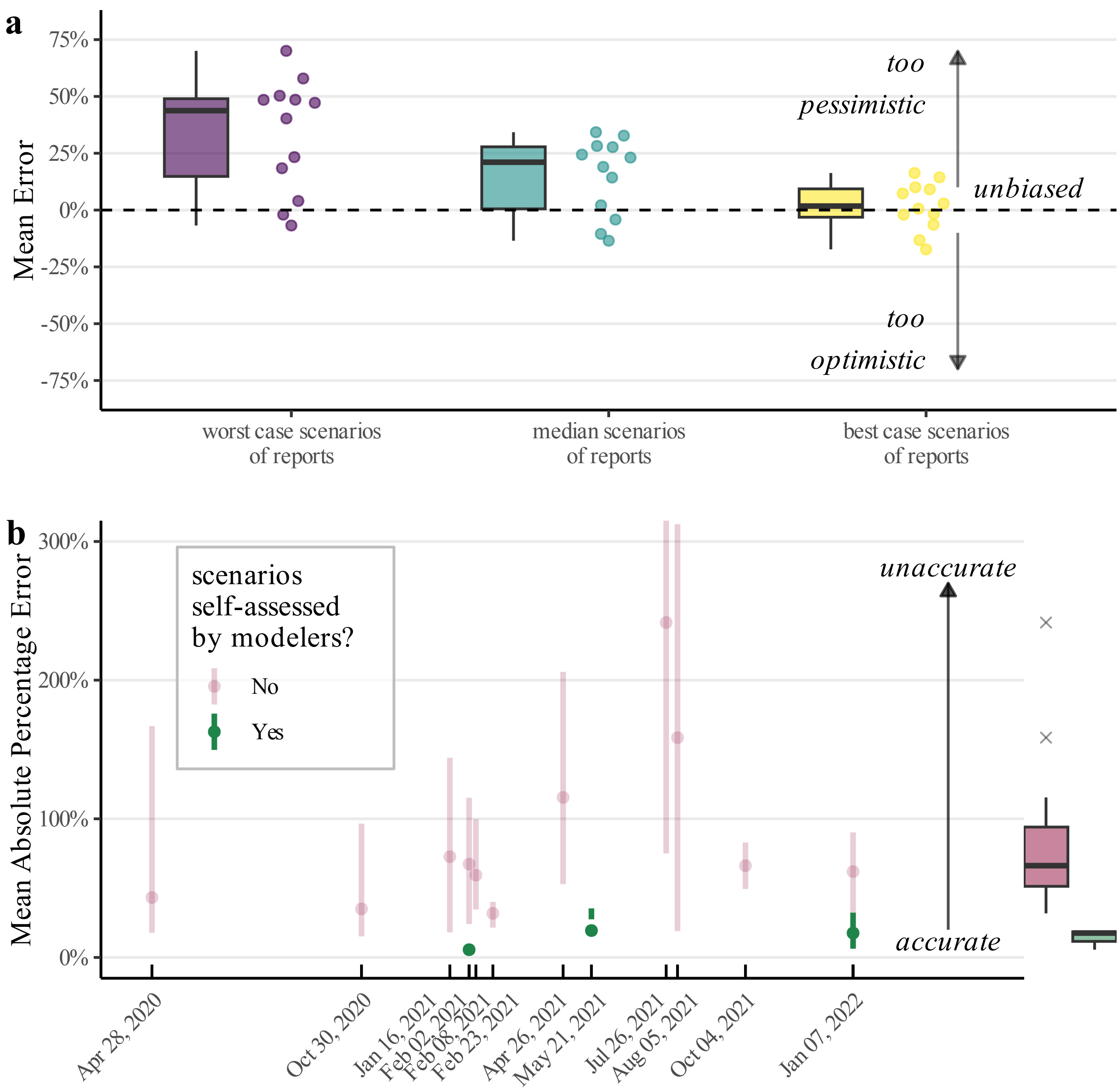


Figure 8: **a)** bias of best-case (yellow), median (blue) and worst-case (purple) scenarios of each report, assessed with the Mean Error. Compared to the real epidemic activity, values close to 0% are unbiased, negative values are too optimistic and positive values are too pessimistic. The best-case scenarios are unbiased, while median and worst-case scenarios are biased towards pessimism. **b)** for each report, accuracy of scenarios (assessed by mean absolute error) and public self-assessment by modelers. Boxplots on the right refers to median scenarios of each report. Reports self-assessed by modelers are more accurate than the complete distribution.

The pessimistic ones assume 40% and 60%. In their retrospective assessment, the modelers only present the optimistic scenarios, and do not discuss the hypotheses (Figure 9c). However, at the time of the retrospective, data from UK Health Security Agency UKHSA [2023] already indicated that the pessimistic assumptions were more correct, which was confirmed in later reports (see supplementary materials). The comparison to the scenarios with the more correct pessimistic assumptions, not evaluated by the modelers, is presented in Figure 9d.

This leaves 2 legitimate comparisons. The May 21, 2021 report was informally assessed on the modeler's twitter feed Cauchemez [2021]. Such a comparison was not performed on the less favorable report published just a few weeks earlier (Apr 26, Figure 4b). The second legitimate comparison concerns the Feb 21, 2022 report, not included in our

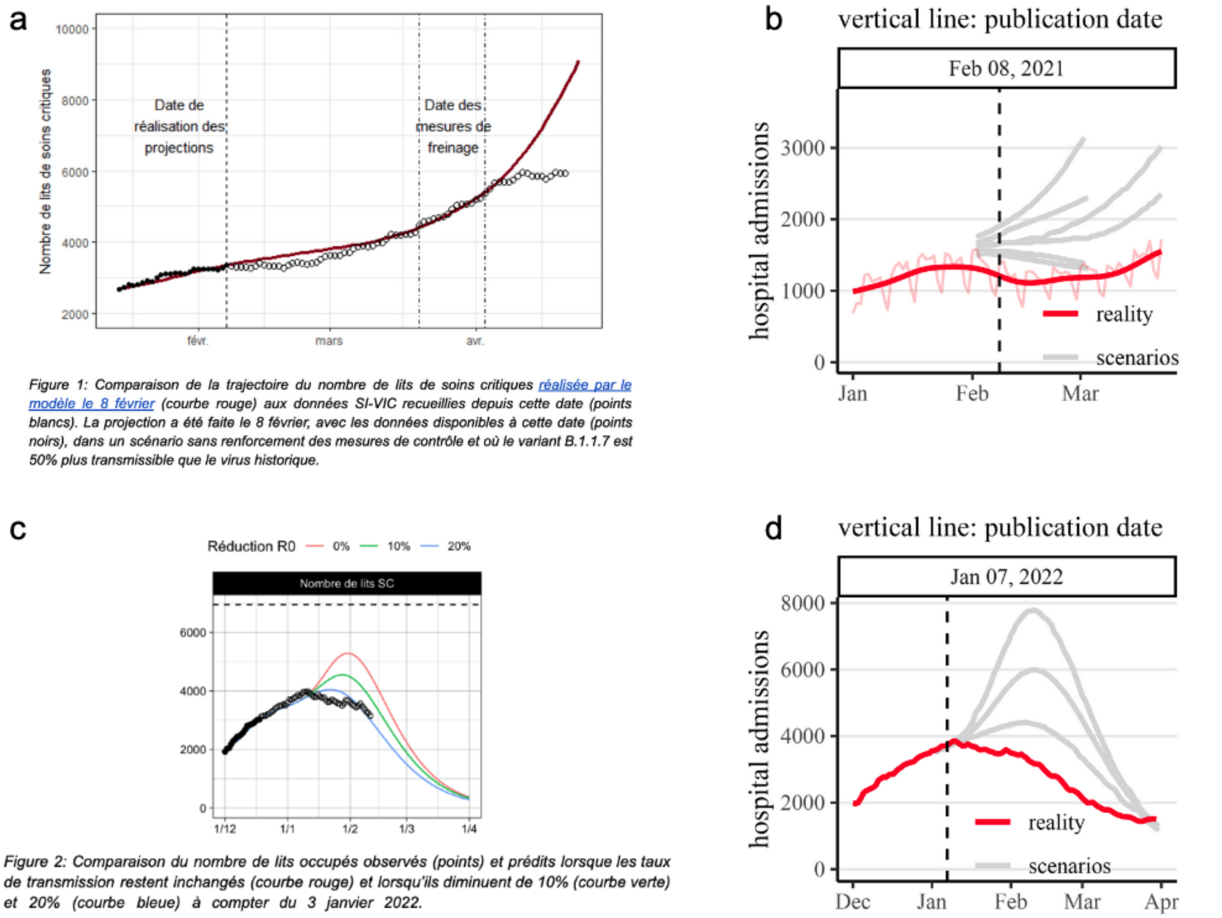


Figure 9: Improper self-assessments by modelers. Top: for February 8, 2021 report. Bottom: for Jan 7, 2022 report. **a)** self-assessment made by modelers, based on a curve absent from the original report. **b)** reality vs scenarios actually present in the original report. **c)** self-assessment made by modelers, based on the subset of scenarios with wrong vaccine efficacy assumptions (see text). **d)** reality vs scenarios with correct vaccine efficacy assumptions

retrospective since it only focuses on positive cases and not on hard endpoints such as hospitalizations or Intensive Care Units.

Figure8b compares the accuracy of the scenarios in the reports of our retrospective and in the subset of reports self-assessed by the modelers. Reports self-assessed by the modelers have a much lower Mean Absolute Percentage Error than the ones they did not assessed. The result remains robust whether or not the two illegitimate comparisons are included.

4 Discussion

4.1 Focus on worst-case scenarios

While this retrospective deals simultaneously with worst-case, best-case and median scenarios, it is usually worst-case scenarios which plays the major role in policy-making. For instance, the worst-case scenario from Pasteur Institute's modelling report of October 26th, 2020 was claimed as "unavoidable" by the head of the French government when announcing a national lockdown Macron [2020]. This exemplifies the fact that the public and policy-makers often focus on worst-case scenarios in decision making.

Yet, in our retrospective, worst-case scenarios appear to be the most unaccurate and biased. For Intensive Care Units beds, their median error at one moth is close to 6000 beds, considering that the maximum bed capacity itself was only slightly above 7000 beds (Figure 6). On the other hand, best-case scenarios were less discussed by policy-makers during the pandemic, but were on average much more accurate than median and worst-case scenarios (Figure6, Figure8a).

Interestingly, while best-case scenarios are consistently more accurate than median scenarios, indicating a trend to overestimate the endpoints in the reports, this trend was neither picked on nor modified by modelers throughout the span of this retrospective. This can be explained by a lack of retrospective work such as the one presented here to spot such a trend. It could also stem from a sense of urgency during the pandemic, preventing modelers from looking back at their previous scenarios to assess them thoroughly.

4.2 Common critics to scenarios evaluation

4.2.1 Scenarios trigger social distancing measures, resulting in lower epidemic spread

A common critic of retrospective analysis from modelers Jit et al. [2023], Holmdahl and Buckee [2020] is that good models lead to policy change, avoiding the dire predictions made by the models. This argument almost claims that predictive failure is an inherent feature of epidemiological modelling.

This is why we repeatedly stress the fact that we only assess scenarios of which the underlying hypotheses were met in reality, and exclude other scenarios. For instance, if a policy change takes place (e.g. curfew or national lockdown) which was not explicitly modeled, we would not run our retrospective on this specific scenario. After this exclusion, the remaining scenarios variability often lies in different effective reproduction number R, displaying uncertainty in the effectiveness of the non pharmaceutical interventions.

Careful checking of all modelling hypotheses, as well as detailed justification of every included and excluded scenario, is given in the supplementary materials.

4.2.2 Scenarios are not forecast

Another common objection to the evaluation of models is that modelling scenarios are not forecasts Holmdahl and Buckee [2020]. Or, as stated by Neil Ferguson Adam [2020], "Models are not crystal balls", referring to the models released by his team at Imperial College. Yet, the term "predict" (applied to the models' ability to anticipate the situation) appeared 20 times Ferguson et al. [2020] in his report presenting the models' forecasts.

While being more prudent, French reports from INSERM Sabbatini et al. [2021] focus on the "*predicted weekly number of hospitalizations*". The terminology "predict" is also used by Pasteur Institute team when performing retrospective on their own forecasts (Feb 15, 2022 report): *"predicted peak"*, *"the predicted and observed dynamics of daily admissions are also close*. This can explain the confusion among policy-makers and the public: while told that scenarios are not predictions, the term predict happens to be used by the modelers themselves.

4.2.3 Better safe than sorry

Finally, some may claim that it is better to anticipate worst-case scenarios event at the expense of overestimating future epidemic spread, because it allows to take measure avoiding the most dire consequences.

However, on must consider that non pharmaceutical interventions can be associated with numerous harms (including, but not restricted to : loss of civil liberties, rise in global hunger FAO et al. [2023], medical care and routine child immunization disruption Chakrabarti et al. [2023], WHO and UNICEF [2020], rise in mental health issues Léon et al. [2023] and economic disruption), we emphasize that overestimation of epidemic spread by models is not devoid of harms. Therefore, accuracy is critical for modelling, and overestimation should be viewed as a failure of modelling to the same extent as underestimation is.

4.3 Consequences for Non Pharmaceutical Interventions evaluation

4.3.1 Overestimating interventions effectiveness

Some models are used to assess the effectiveness of non pharmaceutical interventions, by providing counterfactual scenarios of what would have happened in the absence of the implemented measures Roux et al. [2020]. The difference between this counterfactual and reality is then deducted as the effect of the non pharmaceutical intervention.

However, if a model has a systematic bias toward overestimating epidemic activity, as suggested by our results (Figure 8a), this leads to overestimating the effect size of the intervention. This overestimation is consistent with studies reporting that the homogeneity hypothesis in compartmental SIR models leads to inflating the epidemic spread Merler et al. [2015], Zachreson et al. [2022].

Besides bias, it has been shown that too much flexibility in models can lead to implausible results Soltesz et al. [2020]. Flaxman et al. [2020] claimed that 3 million lives were saved due to NPIs —and lockdowns in particular—in 11 countries in Europe. However it also found that large gathering ban was 35 times more effective in Sweden compared to other European countries (reduction of reproductive number by 70% vs 2%). This makes the Swedish ban similar to a complete lockdown in the other European countries.

4.3.2 Disentangling hypotheses from results

Not challenging models relevancy can lead to mistaking assumptions ("any difference between reality and model output is due to an intervention") for conclusions ("the model shows the effect of the intervention").

For instance, in Pasteur Institute self-assessment of the Omicron wave (Feb 15, 2022 report), modelers simply observe that *“the observed trajectory is close to the scenario in which it was assumed that the French would reduce their contacts by 20% in January”*. Besides that careful verification of the vaccine efficacy assumptions does not support this claim (table S2 and Figure 9c and d), falsifying the 20% contact reduction claim would require empirically measuring contact in the French population, which is difficult. If anything, google mobility data indicates no major change in France during the period (Jan 7 to Feb 15, 2022) Google [2022].

Globally envisioning all previous scenarios rather points towards a systematic overestimation of pandemic activity by the model and not a 20% contact reduction by French people; but such a conclusion cannot be drawn if one does not consider the possibility that the model might be wrong.

4.4 Policy Influence

4.4.1 Stakeholders embracing models output

Models have had large political influence in first implementing social distancing measures, and then justifying them.

The first case is exemplified by the audition of the head of the French scientific council in the Senate, during hearings regarding the implementation of “health pass” in July 2021. He stated that *"the model clearly shows that we are going to find ourselves in a complicated situation at the end of August"* Delfraissy [2021], even though these models featured a range of possible outcomes from a negligible impact to a dire situation (see figure 4c). The French Conseil d’Etat then rejected legal challenges to this health pass implementation, citing these same models Conseil d’Etat [2021].

The second case is exemplified by Spain, when the first lockdown was judged unconstitutional by the Supreme Court, canceling all the fines that had been distributed. The Spanish president then justified the decision by stating that *“That is not me talking, but established by independent scientific studies - as a result of this lockdown, 450,000 lives have been saved”* Sanchez [2021]. While the source of the figure is not specified, it corresponds to the retrospective modelling study by Flaxman et al. [2020] , which had already been cited by Spanish newspapers El Paìs [2020].

4.4.2 Influence by the modelers themselves

In some cases the modelers themselves make explicit policy recommendations. In INSERM report, Sabbatini et al. [2021] summary for policy makers states *"These results show the need to reinforce social distancing measures"*. The original Imperial College report described "suppression" (i.e., lockdown) as the *"preferred policy option"* to avoid overwhelming many times over intensive care units Ferguson et al. [2020].

The self-assessment by modelers can also have political influence. During the French senatorial hearings of January 2022 regarding the extension of the health pass towards a vaccine pass, questions were raised about models’ relevance. The final senatorial report answers this question based on Pasteur Institute self-assessment of their Omicron wave models, concluding *“the Pasteur Institute’s projections proved to be within the expected ranges"* Deseyne et al. [2022]. There is no mention of the numerous times during the 2 pandemic years where the reality was outside Pasteur Institute’s scenarios range (figure 3).

4.5 Relevance for policy making

Some scenarios closely match reality, in most cases the best-case scenario. However these accurate scenarios are almost always part of reports displaying large uncertainty, sometimes as large as the historical epidemic peak of spring 2020 (see section 3.2.3 and figure 7). Anticipated outcomes ranging from a negligible to dire epidemic pressure is of little practical use for policy-makers.

One might think that this large range of scenarios could offer insights regarding the effectiveness of different policies. However we insist on the fact that the residual scenarios after our hypotheses verification only display irreducible uncertainty.

For instance the July 2021 report features more than 100 scenarios regarding daily hospital admissions. After checking for vaccine uptake hypotheses, the uncertainty regarding epidemic peak remains substantive: 1000-3500 daily admissions in the best vaccine uptake configuration, and 1500-5000 in the worst one. For context, the historical first wave peak was 3000. These ranges are entirely due to sheer uncertainties regarding modeled social distancing measures effect, on which the policy-maker has no grasp.

4.6 Good practices for modelling scenario evaluation

In France, prominent modelers were also members of the scientific council designed to advise governments on policy-making. This creates a conflict of interest where scientific advisers that help shape recommendations on non pharmaceutical interventions are also producers of epidemiological models used in hindsight to assess the usefulness of the same interventions. A good practice is therefore to have independent scientists perform critical assessment of epidemiological models accuracy; even though this doesn't prevent modelers from also taking critical look at their own modelling.

Besides conflict of interest, a risk faced when assessing models' predictive power is to perform cherry-picking and to focus only on the models that turned out to be accurate. We highlight this in figure 8b

Sofonea and Alizon [2021] also exemplifies this risk. In a retrospective assessment of the predictive power of their models, they describe their retrospective self-evaluation methodology with the following: “Among the pool of scenarios investigated for each period, only the closest to reality is shown”, without any other inclusion criteria or hypotheses verification. The shortcoming is obvious: at the time of the scenarios publication, they did not know which of the (sometimes very different) scenarios would turn out to be the closest to reality. With such a criteria, simply producing numerous scenarios almost guarantees that one will be close to reality. Since this scenario will be the only one displayed in the retrospective, this will give a false sense of predictive power, to the point of claiming that the model has the "ability to anticipate critical care activity more than a month in advance" ETE [2022].

A good practice to avoid this would be preregistration of modelling scenarios, similar to what has been done for clinical trials Dickersin and Rennie [2012], Kaplan and Irvin [2015]. Preregistration ensures that no cherry-picking is done, and minimizes the risk of false positives.

4.7 Transparency : reports and data accessibility

modelling reports were published systematically on Pasteur Institute's website only starting 2021 (figure S2). To assess previous reports, we had to rely on newspaper figures, when journalists had access to the data.

Part of the reason for not releasing the data throughout 2020 seems to be governmental decisions, which was bypassed when the modelling team started to publish their reports on their website. For instance, in April 2020, the Ministry of Health stated that modelling results "were not yet finalized" and "could not be publicly released for now" Les Echos [2020], but to our knowledge the results were not published afterwards. In another instance, in October 2021, the ministry chose not to communicate updated and less pessimistic projections. “The Minister of Health chose not to mention them [the models]. The concern to keep citizens on their toes has taken precedence over transparency, which would nevertheless require the scientific data to be made public, even if they are less catastrophic than what had been announced.” Franceinfo [2020]

Besides the reports not available for political reasons, even in the case of published reports, the underlying code is not available and numerical data is not available either. This prevents evaluation of the models, since the corresponding author did not reply when contacted for data and code availability. The only way to perform a retrospective assessment was thus to undertake manual extraction of the numerical data from the reports figures. Model reproducibility is however impossible, since the code used to produce the figures is not publicly available.

5 Conclusion

In this systematic retrospective assessment of COVID scenarios in France, we show that the prospective scenarios proposed during the pandemic were a) biased towards pessimistic results, b) unaccurate and c) uncertain.

The systematic bias suggests that using counterfactual scenarios to assess non pharmaceutical interventions may overestimate their efficacy. WHO pandemic preparedness plan rated modelling studies as providing "low to very low" level of evidence. Our review suggests that such caution is still warranted.

Despite their deep policy influence, some of these reports were never publicly released, and the underlying scenarios data were never public, rendering retrospective evaluation difficult. The modelers publicly self-assessed only a small subset of their published scenarios, making it prone to publication bias.

Our work highlights the need for more systematic evaluation of the relevancy scenarios guiding policy makers in times of pandemic, and more transparency.

5.1 Code availability

Code supporting the results presented are available at the following github repository.

References

Jodie McVernon and Jonathan Liberman. Who keeps covid-19 a public health emergency of international concern. *bmj*, 380, 2023.

UKHSA. Effectiveness of non-pharmaceutical interventions to reduce transmission of COVID-19 in the UK - A rapid mapping review. Technical report, UK Health Security Agency, 2023. URL `https://assets.publishing.service.gov.uk/media/65144556b1bad400144fd910/NPI_and_COVID_mapping_review.pdf`.

Jonathan Roux, Clément R Massonnaud, Vittoria Colizza, Simon Cauchemez, and Pascal Crépey. Modeling the impact of national and regional lockdowns on the 2020 spring wave of covid-19 in france. *Scientific Reports*, 13(1):1834, 2023.

Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261, 2020.

Pedro Sanchez. Pedro Sánchez argues that the unity of citizens and institutions led to defeat of COVID-19, 2021. URL `https://www.lamoncloa.gob.es/lang/en/presidente/news/Paginas/2020/20200610control-session.aspx`.

FAO et al. The state of food security and nutrition in the world 2023. Technical report, FAO, IFAD, UNICEF, WFP and WHO, 2023.

Averi Chakrabarti, Elizabeth F Bair, and Harsha Thirumurthy. Routine child immunizations in india during the covid-19 pandemic. *SSM-Population Health*, 22:101383, 2023.

WHO and UNICEF. At least 80 million children under one at risk of diseases such as diphtheria, measles and polio as COVID-19 disrupts routine vaccination efforts, warn Gavi, WHO and UNICEF, 2020. URL `https://www.who.int/news/item/22-05-2020-at-least-80-million-children-under-one-at-risk-of-diseases-such-as-diphtheria-measles-and`

Christophe Léon, Enguerrand du Roscoät, and François Beck. Prévalence des épisodes dépressifs en france chez les 18-85 ans: résultats du baromètre santé 2021. *Bull Épidemiol Hebd*, 2:28–40, 2023.

Lili Li, Araz Taeihagh, and Si Ying Tan. A scoping review of the impacts of covid-19 physical distancing measures on vulnerable population groups. *Nature communications*, 14(1):599, 2023.

UNICEF. Preventing a lost decade. Technical report, UNICEF, 2021.

Mark Jit, Kylie Ainslie, Christian Althaus, Constantino Caetano, Vittoria Colizza, Daniela Paolotti, Philippe Beutels, Lander Willem, John Edmunds, Baltazar Nunes, et al. Reflections on epidemiological modeling to inform policy during the covid-19 pandemic in western europe, 2020–23: Commentary reflects on epidemiological modeling during the covid-19 pandemic in western europe, 2020–23. *Health Affairs*, 42(12):1630–1636, 2023.

Sibel Eker. Validity and usefulness of covid-19 models. *Humanities and Social Sciences Communications*, 7(1), 2020.

Laura Di Domenico, Giulia Pullano, Chiara E Sabbatini, Pierre-Yves Boëlle, and Vittoria Colizza. Modelling safe protocols for reopening schools during the covid-19 pandemic in france. *Nature communications*, 12(1):1073, 2021.

Cécile Tran Kiem, Clément R Massonnaud, Daniel Levy-Bruhl, Chiara Poletto, Vittoria Colizza, Paolo Bosetti, Arnaud Fontanet, Amélie Gabet, Valérie Olié, Laura Zanetti, et al. A modelling study investigating short and medium-term challenges for covid-19 vaccination: From prioritisation to the relaxation of measures. *EClinicalMedicine*, 38, 2021.

WHO. Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza: annex: report of systematic literature reviews. Technical report, World Health Organization, 2019.

Inga Holmdahl and Caroline Buckee. Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*, 383(4):303–305, 2020.

J Jin, N Agarwala, P Kundu, Y Wang, R Zhao, and N Chatterjee. *Transparency, Reproducibility, and Validation of COVID-19 Projection Models*. Johns Hopkins Bloomberg School of Public Health Expert Insights, 2020.

C Michael Barton, Marina Alberti, Daniel Ames, Jo-An Atkinson, Jerad Bales, Edmund Burke, Min Chen, Saikou Y Diallo, David JD Earn, Brian Fath, et al. Call for transparency of covid-19 models. *Science*, 368(6490):482–483, 2020.

Janyce Eunice Gnanvi, Kolawolé Valère Salako, Gaëtan Brezesky Kotanmi, and Romain Glèlè Kakaï. On the reliability of predictions on covid-19 dynamics: A systematic and critical review of modelling techniques. *Infectious Disease Modelling*, 6:258–272, 2021.

William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

Cameron Zachreson, Sheryl Chang, Nathan Harding, and Mikhail Prokopenko. The effects of local homogeneity assumptions in metapopulation models of infectious disease. *Royal Society Open Science*, 9(7):211919, 2022.

Stefano Merler, Marco Ajelli, Laura Fumanelli, Marcelo F C Gomes, Ana Pastore y Piontti, Luca Rossi, Dennis L Chao, Ira M Longini, M Elizabeth Halloran, and Alessandro Vespignani. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases*, 15(2):204–211, February 2015. ISSN 1473-3099. doi:10.1016/s1473-3099(14)71074-6. URL `http://dx.doi.org/10.1016/S1473-3099(14)71074-6`.

Aviral Chharia, Govind Jeevan, Rajat Aayush Jha, Meng Liu, Jonathan M Berman, and Christin Glorioso. Accuracy of us cdc covid-19 forecasting models. *medRxiv*, pages 2022–04, 2022.

Nino Antulov-Fantulin and Lucas Böttcher. On the accuracy of short-term covid-19 fatality forecasts. *BMC Infectious Diseases*, 22(1):1–7, 2022.

John PA Ioannidis, Sally Cripps, and Martin A Tanner. Forecasting for covid-19 has failed. *International journal of forecasting*, 38(2):423–438, 2022.

Shiva Moein, Niloofar Nickaeen, Amir Roointan, Niloofar Borhani, Zarifeh Heidary, Shaghayegh Haghjooy Javanmard, Jafar Ghaisari, and Yousof Gheisari. Inefficiency of sir models in forecasting covid-19 epidemic: a case study of isfahan. *Scientific reports*, 11(1):4725, 2021.

Robert M Kaplan and Veronica L Irvin. Likelihood of null effects of large nhlbi clinical trials has increased over time. *PloS one*, 10(8):e0132382, 2015.

Kay Dickersin and Drummond Rennie. The evolution of trial registries and their use to assess the clinical trial enterprise. *Jama*, 307(17):1861–1864, 2012.

Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Maylis Layan, Pascal Crepey, Alix Roumagnac, Marc Lavielle, Pierre-Yves Boëlle, and Simon Cauchemez. An ensemble model based on early predictors to forecast covid-19 health care demand in france. *Proceedings of the National Academy of Sciences*, 119(18):e2103302119, 2022.

data.gouv. Données hospitalières relatives à l'épidémie de covid-19, 2023. URL `https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/`.

Simon Cauchemez. Twitter account, 2021. URL `https://twitter.com/SCauchemez/status/1405129313721241603?cxt%3DHHwWhsC-_dDqgoAnAAAA&sa=D&source=docs&ust=1692634113403346&usg=AOvVaw2OxLhP8X-6xTiVCOL1wAAL`.

Emmanuel Macron. Allocution télévisée de m. emmanuel macron, président de la république, sur l'annonce d'un nouveau confinement d'une durée d'un mois minimum pour lutter contre l'épidémie de covid-19, 2020. URL `https://www.vie-publique.fr/discours/276951-emmanuel-macron-28102020-covid-19`.

David Adam. Special report: The simulations driving the world's response to covid-19. *Nature*, 580(7802):316–319, 2020.

Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. Technical report, Imperial College COVID-19 Response Team, 2020.

Chiara E Sabbatini, LD Domenico, Giulia Pullano, and Vittoria Colizza. Estimated date of dominance of voc-202012/01 strain in france and projected scenarios. *Paris: EPIcx lab*, 2021.

Jonathan Roux, Clément Massonnaud, and Pascal Crépey. Covid-19: One-month impact of the french lockdown on the epidemic burden. *MedRxiv*, pages 2020–04, 2020.

Kristian Soltesz, Fredrik Gustafsson, Toomas Timpka, Joakim Jaldén, Carl Jidling, Albin Heimerson, Thomas B Schön, Armin Spreco, Joakim Ekberg, Örjan Dahlström, et al. The effect of interventions on covid-19. *Nature*, 588(7839): E26–E28, 2020.

Google. How did the number of visitors change since the beginning of the pandemic?, 2022. URL https://ourworldindata.org/grapher/changes-visitors-covid?time=2022-01-10..2022-02-15&country=~FRA.

Jean-François Delfraissy. Covid-19 : Jean-françois delfraissy prévoit « une situation très complexe » dans les hôpitaux fin août, 2021. URL https://www.publicsenat.fr/actualites/politique/covid-19-jean-francois-delfraissy-prevoit-une-situation-tres-complexe-dans-les.

Conseil d'Etat. Ordonnance du 26 juillet 2021, 2021. URL https://www.conseil-etat.fr/Media/actualites/documents/2021/07-juillet/454754.pdf.

El Pais. Coronavirus confinement measures may have saved 450,000 lives in spain, 2020. URL https://english.elpais.com/science_tech/2020-06-09/coronavirus-confinement-measures-may-have-saved-450000-lives-in-spain.html.

Chantal Deseyne, Olivier Henno, and Michelle Meunier. Rapport d'information fait au nom de la commission des affaires sociales sur l'adéquation du passe vaccinal à l'évolution de l'épidémie de covid 19, 2022. URL https://www.senat.fr/rap/r21-537/r21-5371.pdf.

Mircea T Sofonea and Samuel Alizon. Anticipating covid-19 intensive care unit capacity strain: A look back at epidemiological projections in france. *Anaesthesia, Critical Care & Pain Medicine*, 40(4):100943, 2021.

ETE. Twitter account, 2022. URL https://twitter.com/ete_fr/status/1484978921926008833.

Les Echos. Malades : les premières prévisions pour fin juin, 2020. URL https://www.lesechos.fr/idees-debats/editos-analyses/pourquoi-philippe-a-douche-les-francais-1199309.

Franceinfo. Désintox. covid-19 : les prévisions pas si alarmantes de l'institut pasteur, 2020. URL https://www.francetvinfo.fr/sante/maladie/coronavirus/desintox-covid-19-les-previsions-pas-si-inquietantes-de-l-institut-pasteur_4141179.html.