Royal Society report: what would a comprehensive evaluation suggest about non-pharmaceutical interventions during COVID-19?

Elisabeth Paul, Garrett W. Brown, David Bell, Jean Merlin von Agris & Valéry Ridde

Published online: 22 May 2024.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

COMMENTARY

Royal Society report: what would a comprehensive evaluation suggest about non-pharmaceutical interventions during COVID-19?

Elisabeth Paul [a], Garrett W. Brown [b], David Bell [c], Jean Merlin von Agris [b] and Valéry Ridde [d]

aSchool of Public Health, Université libre de Bruxelles, Brussels, Belgium; bSchool of Politics and International Studies, University of Leeds, Leeds, UK; cIndependent Public Health Consultant, Lake Jackson, TX, USA; dUniversité Paris Cité, IRD, Ceped, Paris, France

ABSTRACT

A recent report by the Royal Society examined the literature on the effectiveness of non-pharmaceutical interventions (NPIs) aimed at reducing the transmission of SARS-CoV-2 during the COVID-19 pandemic. It emphatically concluded that they were 'effective', which we argue may be misleading to policymakers. This paper performs a critical analysis of the Royal Society report and explains what a proper evaluation of NPIs would mean, arguing that theory-based evaluations are required to approach complex issues and nurture democratic debates on societal choices. We argue that, first, testing the relationship between NPI interventions and their effectiveness in reducing SARS-CoV-2 transmission over a bounded period (a single outcome) is irrelevant for policymaking, because several criteria must be balanced to evaluate any intervention, including efficiency, equity, acceptability, long-term impacts, and sustainability. Second, ignoring the disruptive and unintended effects of NPIs leads to wrong conclusions regarding their overall value. Third, we question the mere objective of NPIs, that is, reducing viral transmission. Finally, we question the methodology used by the Royal Society Report regarding the literature selection and quality. To conclude, we argue that the Royal Society report is not only irrelevant and weak from a methodological point of view but also dangerously misleading in terms of policymaking. Four years after the start of the COVID-19 pandemic, it is high time for researchers, journals, and policymakers to stop debating over NPIs' mere effectiveness against a parameter that requires a PCR test to determine, but engage in public health-based evaluations that weigh all criteria of interest.

Introduction

On 24 August 2023, the online media Politico published an article entitled 'Top review says COVID lockdowns and masks worked, period' (Furlong, 2023). The article, like the Royal Society report that it discusses (The Royal Society, 2023), almost completely overlooks the orthodox principle of evaluating interventions based not only on their effectiveness but on several complementary criteria. Here, we explain what a proper evaluation of lockdowns would mean.

A proper evaluation of COVID-19 NPIs must adopt a theory-based evaluation approach, compare costs and benefits; a recent Royal Society report is not only irrelevant and weak from a methodological point of view, but also dangerously misleading in terms of policymaking.

Evaluation approaches

This paper relies on a critical analysis of the Royal Society Report, referring to contemporary evaluation approaches. A basic principle of orthodox public health and health economics consists in judging interventions on a comparison between overall costs versus benefits, which can be measured in various ways. Cost-benefit analysis measures benefit in monetary terms (which can be useful to compare healthcare programmes with other social programmes, but is challenging); cost-effectiveness analysis measures benefit in physical units (e.g. tuberculosis cases averted, which has the disadvantage of not allowing cross-programmatic comparisons); while cost-utility analysis intends to compare different health programmes based on a more comprehensive measure of benefit, such as quality-adjusted life years (QALY) (Drummond et al., 2015). Based on decades of evaluation of public programs, theory-based and realist evaluation approaches have progressively replaced ancestral 'methods-based' evaluation approaches. They aim to go beyond the elusive attempt of assessing whether an intervention 'works' or not and to rather evaluate 'what works in which circumstances and for whom' (Pawson & Tilley, 1997). These approaches recognize that traditional experimental quantitative methods alone (e.g. randomized control trials or RCTs), which have become very influential in certain circles (Donovan, 2018), are as a sole measure inappropriate in the study of complex systems such as a pandemic response (Sturmberg, 2023). This is because they are dependent on the *ex ante* definition of outcomes, which seeks to limit heterogeneity in results, thereby ignoring the more difficult examination of the mechanisms through which a program works or not. As a result, experimental methods fail to take contextual moderators into account (Craig et al., 2018) and thus fail to understand how an intervention (input) moves through various pathways to outcomes, how context moderates these pathways, and, importantly, how those moderated outcomes correspond to an overall goal (realist evaluations compare *ex ante* assumptions about how it should work to *ex post* findings of how it actually worked or not and why – visualizing effects via a theory of change). This allows realist evaluations to be empirically inductive, giving necessary measure to a wider set of contextual (external) and programme (internal) effects. Moreover, experimental methods fail to anticipate unintended consequences of the implemented measures, which are likely to be numerous in the case of COVID-19 response (Bavli et al., 2020; Diallo et al., 2023; Turcotte-Tremblay et al., 2021).

Results

Is the report relevant for policymaking?

The Royal Society recently published a report entitled 'COVID-19: examining the effectiveness of non-pharmaceutical interventions' whose objective was to 'assess what has been learnt about the effectiveness of the application of non-pharmaceutical interventions (NPIs) during the COVID-19 pandemic of 2020–2023 by assembling and examining evidence from researchers around the world' (The Royal Society, 2023). As it states, '[s]ix groups of researchers were commissioned to assemble evidence reviews for this report, examining the effectiveness of a range of NPIs that were applied with the aim of reducing the transmission of SARS-CoV-2'. In other words, the relationship being tested is between NPI interventions (inputs) and their effectiveness in reducing SARS-CoV-2 transmission over a bounded period (a single outcome).

Most researchers or policymakers trained in evaluation should see a concern, since the research has not been designed to address the necessary questions required for public health policymaking. This is because public interventions should be chosen – and thus assessed – based on several criteria, each providing one facet of importance regarding an intervention and may have quite opposing impacts. The key to appropriate public health policy is to weigh these criteria against one another, knowing that sometimes the costs of an intervention (or series of interventions) may be greater than its benefits.

This is not just conjecture; it is best policy practice. The Organisation for Economic Cooperation and Development (OECD) considers six evaluation criteria: effectiveness, efficiency, relevance, impact, sustainability, and coherence (OECD, 2020). The UK Medical Research Council developed a new framework that 'defines evaluation as going beyond asking whether an intervention works (in the sense of achieving its intended outcome), to a broader range of questions including identifying what other impact it has, theorising how it works, taking account of how it interacts with the context in which it is implemented, how it contributes to system change, and how the evidence can be used to support decision-making in the real world' (Skivington et al., 2021). As for the World Health Organization (WHO), its strategic planning handbook identifies 'criteria such as, but not limited to, the burden of the health issue at hand, fairness, cost of the intervention, responsiveness, the effectiveness of the intervention and the acceptability of the intervention' and recognizes that '[a] society may also include other criteria that it feels are essential and reflect its culture, history and objectives' (Schmets et al., 2016). As an example, a leaky nuclear power station will provide positive outcomes in reducing carbon emissions, and one could ignore future cancer and foetal malformations, reporting such a study as demonstrating 'nuclear power works, period'.

A narrow evaluation objective leads to incomplete conclusions

Contrary to pre-pandemic evidence which shows uncertain effects of containment measures (Campeau et al., 2018), the Royal Society report concludes that NPIs were overall effective, particularly lockdowns (p.12). By limiting the evaluation to the narrow criteria of effectiveness in terms of reducing transmission of a single virus, without also evaluating NPI costs, acceptability, and equity – that is, the distribution of costs and benefits, this outcome is valueless for policymaking. Its cloak of 'science' makes it even more dangerous. Indeed, the choice of another evaluation criteria tells a totally different story. WHO notes that closing workplaces will impoverish people, and poverty is associated with overall worse health outcomes, disproportionately affecting low-income people (World Health Organization, G. I. P, 2019). In 2020, global economic activity, as measured by GDP per capita, contracted more dramatically than in any other year since World War II. The disruptions of pandemic restrictions caused the first real global decline since World War II, including in the world's poorest regions (International Monetary Fund, 2023). More than one-third of households in low- and middle-income countries stopped working during lockdowns (Bundervoet et al., 2022). Governments worldwide attempted to mitigate the self-imposed economic crisis by increasing public debt. The resulting cost-of-living crisis is now further increasing the hardships poor people face globally.

There is now extensive evidence that NPIs have had disruptive effects on society and health services (Abe et al., 2023; Ahmed et al., 2022; Bardosh, 2023; Teglia et al., 2022), increased inequalities (Carillon et al., 2020; Li et al., 2023; Ost et al., 2022), aggravated the burden of non-communicable diseases (Lartey et al., 2023; Seal et al., 2022), and thereby had negative effects on populations' wellbeing and health (Violato et al., 2023). It is not unreasonable to surmise that these will result in more years of life lost than were those 'saved' from COVID-19, as they tend to disproportionately affect children and younger people (Christakis et al., 2020; Felfe et al., 2023). Other effects such as loss of trust, fractured societies, and missed opportunities caused by disruption of daily life are difficult to capture through statistics. Any legitimate cost-benefit analysis of NPIs must take these into account. Even if we assume that NPIs managed to substantially reduce harms from COVID-19 in wealthier countries, which is disputed by many (Joffe & Redman, 2021), their potential benefits are still smaller in lower-income countries with younger populations and lower resilience to economic harms.

Was the declared objective of NPIs appropriate?

Returning to the criteria for the Royal Society investigation, the objective of 'reducing transmission of SARS-CoV-2' is also of itself problematic, since SARS-CoV-2 is not equivalent to COVID-19 – the

disease it contributes to provoking (Paul et al., 2022). It has been known since the end of the first wave in June 2020 that COVID-19 severity was, to a large extent, dependent on people's age and comorbidities (Levin et al., 2020; Verity et al., 2020; Williamson et al., 2020) and was generally worse on the first infection (León et al., 2022). Yet, the imposition of NPIs ignored individual risk and acquired immunity, making it both inequitable and unsustainable (Kalk et al., 2022). One may also question whether a zoonosis (such as SARS-CoV-2) is at all controllable long-term via reducing the movements of human beings alone – as testified by significant rates of infections (and deaths) following the interruption of NPIs in China (Xiao et al., 2023), Australia, and New Zealand (Our World in Data, 2023). Furthermore, there is, to our knowledge, no evidence of a significant and lasting impact of lockdowns/NPIs on mortality itself (Gupta et al., 2021; Herby et al., 2023; Kapitsinis, 2021; Talic et al., 2021) – a health outcome that should matter to the Royal Society.

At best, NPIs were demonstrated to slow the transmission or 'flatten the curve'. This may have had positive effects, since it can allow health systems to better adapt and respond and to increase hospitalised patients' chances of survival (Kadri et al., 2021). Nevertheless, this does not mean that fewer people eventually undergo their first infection over the longer-term. Despite the early promises of a 'zero-Covid' policy (Horton, 2021), the evidence suggests that SARS-CoV-2 transmission has only been delayed – even when vaccines became available – as shown in China, for example (Xiao et al., 2023). While the report says NPIs remained in place 'into 2021', restrictions such as mask mandates in fact remained in place well into 2022 in most countries, and into 2023 in many, at least in some settings (e.g. health facilities) (Hale et al., 2021). Any gains also came at huge economic cost, which was estimated at US$24 trillion when considering COVID-19 stimulus packages, medical response, and welfare safety nets (« COVID response drives $24 trillion surge in global debt: IIF », 2021).

Questioning the quality of the selected literature

The studies selected by the Royal Society report are quite supportive of lockdowns, but the inclusion of other literatures would have given a more comprehensive picture of reality. Moreover, the authors have insufficiently assessed the quality of included studies. In their report on the effectiveness of social distancing measures, the researchers themselves ranked 334 of 338 included studies as of low or very low quality, and none of high quality. The four included studies of 'moderate' quality include two studies on social distancing interventions in very specific settings with very low-risk individuals, namely school children and gym members. A third study was classified as being of moderate quality only where it showed a significant effect (of restrictions on gathering) but as of low-quality where it failed to show a significant effect (of stay-at-home orders) (Murphy et al., 2023). The low-quality studies used by the Royal Society further include several modelling studies resting on unproven assumptions, and many studies only compare trends within one location over time. By contrast, Herby et al. (2023) recently produced a meta-analysis of lockdowns, applying more rigorous inclusion criteria. Based on studies using the Oxford COVID-19 Government Response Tracker panel database and its lockdown stringency index (Hale et al., 2021), they found that the average lockdown in Europe and the United States to have reduced deaths attributed to COVID-19 by merely 3.2% in the first half of 2020 (Herby et al., 2023).

Take also the example of masks, for which the authors conclude that '[t]he weight of evidence from all studies suggests that wearing masks, particularly higher quality masks (respirators), supported by mask mandates, generally reduced the transmission of SARS-CoV-2 infection'. Most included sources were observational studies that relied on self-reported mask wearing. Could it be that the Royal Society put greater faith in opinion than in analysis of evidence? Without cherry picking, the Royal Society would have included country evidence contradicting the view that face masks are beneficial overall, such as the one in Finnish schools concluding 'no additional effect was gained from mandating face masks' (Juutinen et al., 2023), which is curiously not included. The Royal Society conclusions contradicts peer reviewed evidence including the recent Cochrane systematic

review, which used only RCTs of virus-related outcomes and concluded that '[t]here is uncertainty about the effects of face masks' (Jefferson et al., 2023). Indeed, only two RCTs were conducted on community masking: one in Denmark which found no significant effect (Bundgaard et al., 2021), and one in Bangladeshi villages which found a small effect (10%) of combined masks and training on virus transmission on symptomatic seroprevalence (Abaluck et al., 2021). Another systematic review, which included observational studies, concluded that 'masks may be associated with a small reduction in risk for SARS-CoV-2 infection in community settings. Surgical masks and N95 respirators may be associated with similar infection risk in routine patient care settings, but a beneficial effect of N95 respirators cannot be ruled out' (Chou & Dana, 2023).

Most studies included in the Royal Society report analysed mask wearing, not the effect of mask mandates. Eighteen studies analysed the effectiveness of mask mandates for reducing transmission of SARS-CoV-2. The authors excluded studies that compared mask-wearing and COVID-19 infection in large groups, e.g. comparing different countries with or without mask mandates. In fact, eight of the studies on mask mandates were set in healthcare settings, and six were set in schools. Only two studies investigated the effects of mask mandates on the general population, both from the United States. The review effectively gathers the results of many low-quality studies covering short time periods in selected regions. It neglects the experience of the Nordic countries who weathered the pandemic with no (Sweden) or comparably less restrictive (e.g. Norway, Denmark) mask mandates.

Such attempts to summarize and meta-analyse results from different types of masks, different settings (communities, schools, care facilities), and different contexts are doomed to fail according to the principles of realist evaluation or any real-world evaluation (Bamberger et al., 2006). Even when limiting the study to given real-world settings – say, schools – the many confounders within observational studies prevent the drawing of firm conclusions (Russell et al., 2023). As other researchers argue, '[e]ven if mask wearing could potentially reduce the transmission of SARS-CoV-2 in individual cases, this needs to be balanced against the physical, psychological and social harms that may be associated with mandated mask wearing, not to mention the negative impact of innumerable disposed masks entering our fragile environment' (Beauchamp & Mayhew, 2023).

When different case studies tell different stories

Aside from six (poorly designed) evidence reviews, the Royal Society report includes three case studies of countries that maintained low rates of transmission over a prolonged period. These exceptional cases are Hong Kong, New Zealand, and South Korea. However, the timeframe of the study matters. After suppressing the virus for almost 2 years, Hong Kong's Zero COVID policy ended with an Omicron wave that crushed hospitals in a population with very little natural immunity, quite the opposite of 'flattening the curve'. Between March and April of 2022, more than 9,200 people – that is, 0.12% of Hong Kong's population – died with COVID-19 (Our World in Data, 2023). Hong Kong's peak in excess mortality was more than three times higher that of Sweden in the Spring of 2020 when the country received widespread criticism for their openness. As for South Korea, its excess mortality peak in 2022 also exceeded Sweden's in 2020, and both countries reached similar results in terms of cumulative excess mortality over the past years (Our World in Data, 2023). The main difference was therefore the cumulative harms of lockdowns. While Hong Kong, New Zealand, and South Korea prove that 'it was possible, in certain contexts, to control transmission of SARS-CoV-2 for over a year by implementing early, stringent border controls accompanied by other strict NPIs to prevent and control domestic transmission' (The Royal Society, 2023), this has not proven a sustainable strategy to reduce overall mortality. In other contexts, some countries showed that it was possible to subject an entire population to months-long house arrest, close schools for years, force people to wear masks even outside, and still have extraordinarily high numbers of deaths (e.g. Peru or Ecuador) (Our World in Data, 2023).

Moreover, the choice of other case studies could have told a different story. Instead of the three mentioned case studies, the authors could have chosen to portray Sweden as an example of how to weather a pandemic without lockdowns or mask mandates, ending with lower all-cause mortality than most European countries over the past 3 years (Andersson & Jonung, 2024). In Central America, Honduras had many months of strict stay-at-home orders and closed its schools for almost 2 years, while its neighbour Nicaragua ranks lowest in lockdown stringency worldwide. Public schools remained open, life went on without many restrictions. Nicaragua was not immune to the global economic downturn, yet its GDP per capita fell less than that of other countries in the region. There is no evidence that Nicaragua suffered more COVID-19 deaths than neighbouring countries (Our World in Data, 2023). The same can be said about no-lockdown Tanzania when compared to neighbours such as Uganda, which had the longest school closures worldwide (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2022). Reliable mortality statistics from Nicaragua, Tanzania, or Belarus are not available, but certainly, there is no evidence of any catastrophic spike in deaths. Even if some studies find a minimal positive effect of lockdowns on death rates (Herby et al., 2023) that hardly justifies such extreme measures. If lockdowns worked, we should expect differences between countries that locked down and those who did not be visible. Instead, over the period from January 2020 to December 2022, the correlation between lockdown stringency and excess mortality is positive (Our World in Data, 2023). By contrast, pandemic mortality was closely related to socio-economic vulnerability: countries with low income and high inequality and poverty had high excess deaths, whereas several less vulnerable countries suffered no excess mortality (Ioannidis et al., 2023). The latter group includes countries featured by the Royal Society for their successful NPI regimes (South Korea, New Zealand) but also the Nordic countries with their less restrictive policies (Andersson & Jonung, 2024).

Overall, as for any complex intervention, the real-life effectiveness of NPIs is not inherent to a given NPI but depends on complex interactions between the often multiple NPIs implemented and numerous intertwined factors pertaining to climate, geography, population structure and density, governance (including trust in public authorities) and culture (acceptability), thus preventing comparisons based on single indicators and calling for richer evaluation designs that allow this complexity to be taken into account (Hawe et al., 2009).

Conclusion: what a "top evaluation" says about lockdowns, masks and other NPIs

To conclude, we argue that the Royal Society report is not only irrelevant and weak from a methodological point of view but also dangerously misleading in terms of policymaking. This is how health misinformation occurs. Three and a half years after the start of the COVID-19 pandemic, it is high time for researchers, journals and policymakers to stop debating over NPIs' mere effectiveness against a parameter that requires a PCR test to determine, but engage in public health-based evaluations that weigh all criteria of interest. Undermining the principles of democracy and bodily autonomy is also clear societal harm that should not be ignored. These are complex issues, not well handled in simplistically designed studies. The media and institutions such as the Royal Society were once cognizant of this complexity and willing to investigate more than one side of an equation. They should become so again.

Disclosure statement

ORCID

Elisabeth Paul (iD) http://orcid.org/0000-0002-3928-4093
Garrett W. Brown (iD) http://orcid.org/0000-0002-6557-5353
David Bell (iD) http://orcid.org/0000-0002-7010-6340
Jean Merlin von Agris (iD) http://orcid.org/0009-0005-9478-5316
Valéry Ridde (iD) http://orcid.org/0000-0001-9299-8266

Author contribution

Elisabeth Paul: Conceptualization, Data curation, Formal analysis, Supervision, Validation, Writing – original draft.
Garrett W. Brown: Conceptualization, Data curation, Formal analysis, Supervision, Validation, Writing – original draft.
David Bell: Conceptualization, Data curation, Formal analysis, Validation, Writing – review & editing.
Jean Merlin von Agris: Conceptualization, Data curation, Formal analysis, Validation, Writing – review & editing.
Valéry Ridde: Conceptualization, Data curation, Formal analysis, Validation, Writing – review & editing.

Data availability statement

Data available within the article.

References

Abaluck, J., Kwong, L. H., Styczynski, A., Haque, A., Kabir, M. A., Bates-Jefferys, E., Crawford, E., Benjamin-Chung, J., Raihan, S., Rahman, S., Benhachmi, S., Bintee, N. Z., Winch, P. J., Hossain, M., Reza, H. M., Jaber, A. A., Momen, S. G., Rahman, A., Banti, F. L. . . . Mobarak, A. M. (2021). Impact of community masking on COVID-19: A cluster-randomized trial in Bangladesh. *Science*, *375*(6577), eabi9069. https://doi.org/10.1126/science.abi9069

Abe, K., Kawachi, I., Iba, A., & Miyawaki, A. (2023). In-hospital deaths from ambulatory care–sensitive conditions before and during the COVID-19 pandemic in Japan. *JAMA Network Open*, *6*(6), e2319583–e2319583. https://doi.org/10.1001/jamanetworkopen.2023.19583

Ahmed, T., Roberton, T., Vergeer, P., Hansen, P. M., Peters, M. A., Ofosu, A. A., Mwansambo, C., Nzelu, C., Wesseh, C. S., Smart, F., Alfred, J. P., Diabate, M., Baye, M., Yansane, M. L., Wendrad, N., Mohamud, N. A., Mbaka, P., Yuma, S., Ndiaye, Y. . . . Shapira, G. (2022). Healthcare utilization and maternal and child mortality during the COVID-19 pandemic in 18 low- and middle-income countries: An interrupted time-series analysis with mathematical modeling of administrative data. *PLOS Medicine*, *19*(8), e1004070. https://doi.org/10.1371/journal.pmed.1004070

Andersson, F. N. G., & Jonung, L. (2024). The covid-19 lesson from Sweden: Don't lock down. *Economic Affairs*, *44*(1), 3–16. https://doi.org/10.1111/ecaf.12611

Bamberger, M. J., Rugh, J., & Mabry, L. (2006). *RealWorld evaluation: Working under budget, time, data, and political constraints*. SAGE Publications.

Bardosh, K. (2023). How did the COVID pandemic response harm society? A global evaluation and state of knowledge review (2020-21). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4447806

Bavli, I., Sutton, B., & Galea, S. (2020). Harms of public health interventions against COVID-19 must not be ignored. *BMJ: British Medical Journal*, *371*, m4074. https://doi.org/10.1136/bmj.m4074

Beauchamp, J. D., & Mayhew, C. A. (2023). Revisiting the rationale of mandatory masking. *Journal of Breath Research*, *17*(4), 042001. https://doi.org/10.1088/1752-7163/acdf12

Bundervoet, T., Dávalos, M. E., & Garcia, N. (2022). The short-term impacts of COVID-19 on households in developing countries: An overview based on a harmonized dataset of high-frequency surveys. *World Development*, *153*, 105844. https://doi.org/10.1016/j.worlddev.2022.105844

Bundgaard, H., Bundgaard, J. S., Raaschou-Pedersen, D. E. T., von Buchwald, C., Todsen, T., Norsk, J. B., Pries-Heje, M. M., Vissing, C. R., Nielsen, P. B., Winsløw, U. C., Fogh, K., Hasselbalch, R., Kristensen, J. H., Ringgaard, A., Porsborg Andersen, M., Goecke, N. B., Trebbien, R., Skovgaard, K., Benfield, T. . . . Iversen, K. (2021). Effectiveness of adding a mask recommendation to other public health measures to prevent SARS-CoV-2 infection in Danish mask wearers. *Annals of Internal Medicine*, *174*(3), 335–343. https://doi.org/10.7326/M20-6817

Campeau, L., Degroote, S., Ridde, V., Carabali, M., & Zinszer, K. (2018). Containment measures for emerging and re-emerging vector-borne and other infectious diseases of poverty in urban settings: A scoping review. *Infectious Diseases of Poverty*, *7*(1), Article 1. https://doi.org/10.1186/s40249-018-0478-4

Carillon, S., Gosselin, A., Coulibaly, K., Ridde, V., & Desgrées du Loû, A. (2020). Immigrants facing covid 19 containment in France: An ordinary hardship of disaffiliation. *Journal of Migration and Health*, *1-2*, 100032. https://doi.org/10.1016/j.jmh.2020.100032

Chou, R., & Dana, T. (2023). Major update: Masks for prevention of SARS-CoV-2 in health care and community settings—final update of a living, rapid review. *Annals of Internal Medicine*, *176*(6), 827–835. https://doi.org/10.7326/M23-0570

Christakis, D. A., Van Cleve, W., & Zimmerman, F. J. (2020). Estimation of US children's educational attainment and years of life lost associated with primary school closures during the coronavirus disease 2019 pandemic. *JAMA Network Open*, *3*(11), Article 11. e2028786. https://doi.org/10.1001/jamanetworkopen.2020.28786

Craig, P., DiRuggiero, E., Frohlich, K., Mykhalovskiy, E., & White, M. (2018). *Taking account of context in population health intervention research : Guidance for producers, users and funders of research* [NIHR Evaluation, Trials and Studies Coordinating Centre]. on behalf of the Canadian Institutes of Health Research. CIHR–National Institute for Health Research (NIHR. https://www.ncbi.nlm.nih.gov/books/NBK498645/pdf/Bookshelf_NBK498645.pdf

Diallo, I., Ndejjo, R., Leye, M. M. M., Egbende, L., Tusubira, A., Bamgboye, E. A., Fall, M., Namuhani, N., Bosonkie, M., Salawu, M. M., Ndiaye, Y., Kabwama, S. N., Sougou, N. M., Bello, S., Bassoum, O., Babirye, Z., Afolabi, R. F., Gueye, T., Kizito, S. . . . Seck, I. (2023). Unintended consequences of implementing non-pharmaceutical interventions for the COVID-19 response in Africa: Experiences from DRC, Nigeria, Senegal, and Uganda. *Globalization and Health*, *19*(1), 36. https://doi.org/10.1186/s12992-023-00937-6

Donovan, K. P. (2018). The rise of the randomistas: On the experimental turn in international aid. *Economy and Society*, *47*(1), 27–58. https://doi.org/10.1080/03085147.2018.1432153

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (4th ed.). Oxford University Press.

Felfe, C., Saurer, J., Schneider, P., Vornberger, J., Erhart, M., Kaman, A., & Ravens-Sieberer, U. (2023). The youth mental health crisis: Quasi-experimental evidence on the role of school closures. *Science Advances*, *9*(33), eadh4030. https://doi.org/10.1126/sciadv.adh4030

Furlong, A. (2023, août 24). *Top review says COVID lockdowns and masks worked, period*. https://www.politico.eu/article/masks-lockdowns-social-distancing-evidence-reduce-coronavirus-spread/

Gupta, S., Georgiou, A., Sen, S., Simon, K., & Karaca-Mandic, P. (2021). US trends in COVID-19–associated hospitalization and mortality rates before and after reopening economies. *JAMA Health Forum*, *2*(6), e211262–e211262. https://doi.org/10.1001/jamahealthforum.2021.1262

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government response tracker). *Nature Human Behaviour*, *5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Hawe, P., Shiell, A., & Riley, T. (2009). Theorising interventions as events in systems. *American Journal of Community Psychology*, *43*(3-4), 267–276. https://doi.org/10.1007/s10464-009-9229-9

Herby, J., Jonung, L., & Hanke, S. H. (2023). *Did lockdowns work? The verdict on Covid restrictions*. Institute of Economic Affairs (IEA). https://iea.org.uk/publications/did-lockdowns-work-the-verdict-on-covid-restrictions/?fbclid=IwAR2B0GDS5POsc08F91O_-tNykE-Gjx7P4VLpxNwWji4_CuqAvbcVjnAK9II

Horton, R. (2021). Offline: The case for No-COVID. *The Lancet*, *397*(10272), 359. https://doi.org/10.1016/S0140-6736(21)00186-0

International Monetary Fund. (2023, avril). *World economic outlook (April 2023)*. https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/WEOWORLD

Ioannidis, J. P. A., Zonta, F., & Levitt, M. (2023). Variability in excess deaths across countries with different vulnerability during 2020–2023. *Proceedings of the National Academy of Sciences*, *120*(49), e2309557120. https://doi.org/10.1073/pnas.2309557120

Jefferson, T., Dooley, L., Ferroni, E., Al-Ansary, L., van Driel, M., Bawazeer, G., Jones, M., Hoffmann, T., Clark, J., Beller, E., Glasziou, P. P., & Conly, J. M. (2023). Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews*, *1*(4). https://doi.org/10.1002/14651858.CD006207.pub6

Joffe, A. R., & Redman, D. (2021). The SARS-CoV-2 pandemic in high income countries such as Canada: A better way forward without lockdowns. *Frontiers in Public Health*, *9*, 9. https://doi.org/10.3389/fpubh.2021.715904

Jones, M. (2021, February 17). *COVID response drives $24 trillion surge in global debt: IIF*. Reuters. https://www.reuters.com/article/us-global-debt-iif/covid-response-drives-24-trillion-surge-in-global-debt-iif-idUSKBN2AH285

Juutinen, A., Sarvikivi, E., Laukkanen-Nevala, P., & Helve, O. (2023). Face mask recommendations in schools did not impact COVID-19 incidence among 10–12-year-olds in Finland – joinpoint regression analysis. *BMC Public Health*, *23*(1), 730. https://doi.org/10.1186/s12889-023-15624-9

Kadri, S. S., Sun, J., Lawandi, A., Strich, J. R., Busch, L. M., Keller, M., Babiker, A., Yek, C., Malik, S., Krack, J., Dekker, J. P., Spaulding, A. B., Ricotta, E., Powers, J. H., Rhee, C., Klompas, M., Athale, J., Boehmer, T. K., Gundlapalli, A. V. . . . Warner, S. (2021). Association between caseload surge and COVID-19 survival in 558 U.S. Hospitals, March to August 2020. *Annals of Internal Medicine*, *174*(9), 1240–1251. https://doi.org/10.7326/M21-1213

Kalk, A., Sturmberg, J., Van Damme, W., Brown, G., Ridde, V., Zizi, M., & Paul, E. (2022). Surfing corona waves – instead of breaking them: Rethinking the role of natural immunity in COVID-19 policy. *F1000research*, *11*(337), Article 337. 337. https://doi.org/10.12688/f1000research.110593.1

Kapitsinis, N. (2021). The underlying factors of excess mortality in 2020: A cross-country analysis of pre-pandemic healthcare conditions and strategies to cope with covid-19. *BMC Health Services Research*, *21*(1), Article 1. https://doi.org/10.1186/s12913-021-07169-7

Lartey, S. T., Jayawardene, W. P., Dickinson, S. L., Chen, X., Gletsu-Miller, N., & Lohrmann, D. K. (2023). Evaluation of unintended consequences of COVID-19 pandemic restrictions and obesity prevalence among youths. *JAMA Network Open*, *6*(7), e2323596–e2323596. https://doi.org/10.1001/jamanetworkopen.2023.23596

León, T., Dorabawila, V., & Nelson, L., Lutterloh, E., Bauer, U. E., Backenson, B., Bassett, M. T., Henry, H., Bregman, B., Midgley, C. M., Myers, J. F., Plumb, I. D., Reese, H. E., Zhao, R., Briggs-Hagen, M., Hoefer, D., Watt, J. P., Silk, B. J., Jain, S., & Rosenberg, E. S. (2022). COVID-19 cases and hospitalizations by COVID-19 vaccination status and previous COVID-19 diagnosis—California and New York, May–November 2021. *MMWR Morb Mortal Wkly Rep 71: 125-131; Numéro 71: 25-131*. Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/71/wr/mm7104e1.htm?s_cid=mm7104e1_w

Levin, A. T., Hanage, W. P., Owusu-Boaitey, N., Cochran, K. B., Walsh, S. P., & Meyerowitz-Katz, G. (2020). Assessing the age specificity of infection fatality rates for COVID-19: Systematic review, meta-analysis, and public policy implications. *European Journal of Epidemiology*, *35*(12), 1123–1138. https://doi.org/10.1007/s10654-020-00698-1

Li, L., Taeihagh, A., & Tan, S. Y. (2023). A scoping review of the impacts of COVID-19 physical distancing measures on vulnerable population groups. *Nature Communications*, *14*(1), 599. https://doi.org/10.1038/s41467-023-36267-9

Murphy, C., Lim, W. W., Mills, C., Wong, J. Y., Chen, D., & Xie, Y., Li, M., Gould, S., Xin, H., Cheung, J. K., Bhatt, S., Cowling, B. J., & Donnelly, C. A. (2023). *Appendix B from effectiveness of social distancing measures and lockdowns for reducing transmission of COVID-19 in non-healthcare, community-based settings*. The Royal Society. https://doi.org/10.6084/m9.figshare.23283511.v1

OECD. (2020). *Evaluation criteria*. Organisation for Economic Cooperation and Development. https://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

Ost, K., Duquesne, L., Duguay, C., Traverson, L., Mathevet, I., Ridde, V., & Zinszer, K. (2022). Large-scale infectious disease testing programs have little consideration for equity: Findings from a scoping review. *Journal of Clinical Epidemiology*, *143*, 30–60. https://doi.org/10.1016/j.jclinepi.2021.11.030

Our World in Data. (2023). *COVID-19 data explorer*. https://ourworldindata.org/explorers/coronavirus-data-explorer?zoomToSelection=true&time=2020-03-01..latest&facet=none&country=USA~GBR~CAN~DEU~ITA~IND&pickerSort=asc&pickerMetric=location&Metric=Confirmed+cases&Interval=7-day+rolling+average&Relative+to+Population=true&Color+by+test+positivity=false

Paul, E., Brown, G. W., Kalk, A., Van Damme, W., Ridde, V., & Sturmberg, J. (2022). "When My Information Changes, I Alter My Conclusions." What can we learn from the failures to adaptively respond to the SARS-CoV-2 pandemic and the under preparedness of health systems to manage COVID-19? *International Journal of Health Policy and Management*, *11*(7), 1241–1245. https://doi.org/10.34172/ijhpm.2020.240

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

The Royal Society. (2023). *COVID-19: Examining the effectiveness of non-pharmaceutical interventions*. https://royalsociety.org/topics-policy/projects/impact-non-pharmaceutical-interventions-on-covid-19-transmission/

Russell, F. M., Curtis, N., & Carlin, J. (2023). Lifting universal school masking—Covid-19 incidence among students and staff. *New England Journal of Medicine*, *389*(6), 578–580. https://doi.org/10.1056/NEJMc2215560

Schmets, G., Rajan, D., & Kadandale, S. (2016). *Strategizing national health in the 21st century: A handbook*. World Health Organization.

Seal, A., Schaffner, A., Phelan, S., Brunner-Gaydos, H., Tseng, M., Keadle, S., Alber, J., Kiteck, I., & Hagobian, T. (2022). COVID-19 pandemic and stay-at-home mandates promote weight gain in US adults. *Obesity*, *30*(1), 240–248. https://doi.org/10.1002/oby.23293

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of medical research council guidance. *BMJ: British Medical Journal*, *374*, n2061. https://doi.org/10.1136/bmj.n2061

Sturmberg, J. P. (2023). Changing the paradigm of research. *Journal of Evaluation in Clinical Practice*, *29*(5), 726–729. https://doi.org/10.1111/jep.13828

Talic, S., Shah, S., Wild, H., Gasevic, D., Maharaj, A., Ademi, Z., Li, X., Xu, W., Mesa-Eguiagaray, I., Rostron, J., Theodoratou, E., Zhang, X., Motee, A., Liew, D., & Ilic, D. (2021). Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and COVID-19 mortality: Systematic review and meta-analysis. *BMJ: British Medical Journal*, *375*, e068302. https://doi.org/10.1136/bmj-2021-068302

Teglia, F., Angelini, M., Astolfi, L., Casolari, G., & Boffetta, P. (2022). Global association of COVID-19 pandemic measures with cancer screening: A systematic review and meta-analysis. *JAMA Oncology*, *8*(9), 1287–1293. https://doi.org/10.1001/jamaoncol.2022.2617

Turcotte-Tremblay, A.-M., Gali Gali, I. A., & Ridde, V. (2021). The unintended consequences of COVID-19 mitigation measures matter: Practical guidance for investigating them. *BMC Medical Research Methodology*, *21*(1), 28. https://doi.org/10.1186/s12874-020-01200-x

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2022). *Survey on National education responses to COVID-19 school closures*. https://covid19.uis.unesco.org/school-closures-survey/

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Griffin, J. T., Baguelin, M., Bhatia, S., Boonyasiri, A., Cori, A., Cucunubá, Z.,

FitzJohn, R., Gaythorpe, K. . . . Ferguson, N. M. (2020). Estimates of the severity of coronavirus disease 2019: A model-based analysis. *The Lancet Infectious Diseases*, *20*(6), 669–677. https://doi.org/10.1016/S1473-3099(20)30243-7

Violato, M., Pollard, J., Lloyd, A., Roope, L. S. J., Duch, R., Becerra, M. F., Clarke, P. M., & Kola, L. (2023). The COVID-19 pandemic and health-related quality of life across 13 high- and low-middle-income countries: A cross-sectional analysis. *PLOS Medicine*, *20*(4), 1–20. https://doi.org/10.1371/journal.pmed.1004146

Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., Mehrkar, A., Evans, D., Inglesby, P., Cockburn, J., McDonald, H. I., MacKenna, B., Tomlinson, L., Douglas, I. J., Rentsch, C. T., Mathur, R., Wong, A. Y. S., Grieve, R. . . . Goldacre, B. (2020). Factors associated with COVID-19-related death using Opensafely. *Nature*, *584*(7821), 430–436. https://doi.org/10.1038/s41586-020-2521-4

World Health Organization, G. I. P. (2019). *Non-pharmaceutical public health measures for mitigating the risk and impact of epidemic and pandemic influenza*. https://apps.who.int/iris/bitstream/handle/10665/329438/9789241516839-eng.pdf?ua=1

Xiao, H., Wang, Z., Liu, F., & Unger, J. M. (2023). Excess all-cause mortality in China after ending the zero COVID policy. *JAMA Network Open*, *6*(8), e2330877. https://doi.org/10.1001/jamanetworkopen.2023.30877