Exceptional Compared With What? Estimand Transfer and Background-Mortality Benchmarking in the Diamond Princess Passenger Cohort

Article type: Original Article

Running title: Background mortality in the Diamond Princess

Author: George Kujo

ORCID: 0009-0006-6864-0145

Affiliation: Independent Researcher, Japan

Correspondence: George Kujo, Independent Researcher, Japan. Email: george.kujo.irjp@gmail.com

Abstract

Background

The *Diamond Princess* became an influential early source for SARS-CoV-2 case- and infection-fatality estimates because it provided an unusually well-enumerated, older cohort. Those analyses addressed infection denominators, delay to death and age adjustment. We asked a different question: when the reported death count is interpreted at the passenger-population level, is age- and time-expected background mortality numerically negligible?

Methods

We conducted a descriptive benchmark analysis of 2,666 passengers. Expected all-cause deaths were calculated from official pre-pandemic life tables using passenger age structure and observation windows. The primary window was 3 February–15 April 2020 (72 days). The calculation used Japan 2019 mortality, 45% male/55% female allocation within each 10-year age band, mean single-year mortality within bands and no healthy-traveller reduction. Sex-specific annual probabilities were converted to interval risks before weighting. The reference expectation was compared with 13 passenger deaths reported in Japan as outbreak-associated/COVID-related; a sensitivity outcome added one death after Australian repatriation. Alternative start dates, national life tables, sex allocations, within-band age assumptions and healthy-traveller stress tests were examined.

Results

The primary calculation yielded 8.33 expected background deaths. The background-mortality benchmark ratio was 1.56 for 13 reported deaths and 1.68 for 14. With demographics fixed, four national life tables yielded 8.33–12.41 expected deaths and ratios of 1.05–1.56. Japan-only all-female and all-male structural bounds yielded 5.64 and 11.62 expected deaths. Across aligned 70-, 72- and 86-day anchors, ratios for 13 deaths ranged from 1.31 to 1.61.

Conclusions

Background mortality was not numerically negligible relative to the reported *Diamond Princess* death count. This does not reclassify deaths or estimate excess mortality. It identifies an estimand-transfer problem: the same outbreak death count can inform both infection-fatality and population-mortality questions, but these require different reference structures.

Keywords: Diamond Princess; COVID-19; background mortality; life table; infection fatality ratio; cruise ship; estimand

Introduction

The *Diamond Princess* outbreak became one of the most influential early datasets in COVID-19 epidemiology. The ship carried 2,666 passengers and 1,045 crew when it returned to Yokohama on 3 February 2020; passengers had a mean age of 66.1 years and 55% were female. Extensive testing and follow-up made the outbreak unusually informative when infection denominators were uncertain elsewhere. [1–3]

Early analyses therefore concentrated on infection-conditional questions. Mizumoto and colleagues estimated the asymptomatic proportion, while Russell and colleagues adjusted for delay to death and age structure to estimate a case-fatality ratio (CFR) of 2.6% and infection-fatality ratio (IFR) of 1.3%. [1,2] Later follow-up confirmed a steep age gradient in symptomatic disease, severity and death among infected persons. [4]

Those studies asked: conditional on infection or confirmed disease, what was the risk of death? A different estimand arises when the same reported death count is used to characterize mortality experienced by the passenger cohort: how large is that mortality signal relative to deaths expected from observing a population with the same age structure for the same period?

The distinction matters because background mortality is not zero. Russell and colleagues noted that some deaths could arise from other causes in an older cohort, but regarded natural mortality as slow relative to COVID-19 fatality for their CFR/IFR analysis. [2] That simplification can be

coherent for an infection-fatality estimand without being transferable to a population-mortality interpretation over several weeks.

Recent cruise-ship outbreak work has highlighted a related inferential problem. Chowell and Mizumoto argued in this journal that counting everyone onboard does not by itself identify the correct causal risk set when exposures differ before and during a voyage. [11] The *Diamond Princess* illustrates a distinct second problem: even when the population count is well defined, the reference structure needed to interpret a death count changes when the estimand changes. A correct denominator does not guarantee a correct baseline.

A contemporaneous non-peer-reviewed web analysis, cited here only as a historical precedent rather than as methodological evidence, raised the possibility that natural mortality might matter by applying UK mortality rates to older COVID-positive *Diamond Princess* passengers and deducting an allowance for expected non-COVID deaths. [5] We do not make such a subtraction. Instead, we benchmark the full passenger cohort against age- and time-structured all-cause mortality and keep that reference expectation separate from causal attribution.

We therefore estimated expected background mortality among the 2,666 passengers over observation windows aligned to available mortality ascertainment. The aim was not to revise published IFR estimates or determine how many deaths were caused by COVID-19, but to test whether the background term was small enough to ignore when the passenger cohort's mortality experience is interpreted at the population level.

Methods

Study design and estimand

We conducted a descriptive numerical-scale benchmark analysis using public aggregate data. The target estimand was the relation between reported outbreak-associated/COVID-related passenger deaths and the all-cause mortality expected from observing the full passenger cohort over the same fixed interval. Because the observed and reference quantities have different ascertainment structures, the analysis was not designed as an excess-mortality study, standardized mortality ratio (SMR), attributable-mortality analysis or causal effect estimate.

The analytic population comprised the manifest-based cohort of 2,666 passengers reported for 3 February, which is also the population for which the published age-band structure is available. [3] Crew were excluded because their age structure was substantially younger and would require a separate mortality baseline. Restricting the benchmark to infected passengers would instead answer an infection-conditional severity question and was therefore not used as the primary population reference.

Passenger structure and outcomes

Published passenger counts were available in 10-year age bands: 0–9 (n=16), 10–19 (22), 20–29 (49), 30–39 (59), 40–49 (73), 50–59 (302), 60–69 (911), 70–79 (1,008), 80–89 (215) and 90–99 (11). [3] Exact age-by-sex counts were unavailable. Applying band midpoints to these counts implied a mean age of 66.17 years, closely matching the reported mean of 66.1 years and supporting the overall location of the cohort's age distribution. This calibration does not identify

the within-band age distribution; lower- and upper-age assignments were therefore retained as structural sensitivity bounds.

Outcome A was the 13 deaths in the Japanese Ministry of Health, Labour and Welfare (MHLW) *Diamond Princess* series through 15 April 2020; the 15 April individual press release explicitly identified the event as the 13th death. [9] The MHLW cruise-ship press-release index separately notes one death after Australian repatriation; Outcome B included that death, giving 14. [10] These outcomes were not treated as complete all-cause vital-status follow-up, and no individual cause of death was reclassified.

Observation windows

The primary interval was 3 February–15 April 2020 (72 elapsed days). Three February is an independent epidemiological anchor: the ship was in Yokohama and the cohort was enumerated before formal quarantine began. [3] Fifteen April is the date of the thirteenth MHLW-listed ship-associated death. [9]

Two aligned sensitivity windows used the same outcome endpoint: 20 January–15 April (86 days), representing the broad voyage horizon, and 5 February–15 April (70 days), representing the start of quarantine. The 70-, 72- and 86-day analyses therefore alter the start date while preserving the same 15 April outcome-ascertainment endpoint. We did not extend the benchmark to 8 June because that would change the outcome horizon: later COVID-19 prognosis follow-up among infected persons does not provide correspondingly complete all-cause vital status for the original passenger cohort. Pairing a longer life-table window with the fixed 13- or 14-death series would therefore be an unaligned follow-up comparison. [4]

The choice of start date is analytically consequential because expected background deaths accumulate with elapsed time. An earlier start therefore lowers the benchmark ratio even if the reported death count is unchanged. We selected 3 February rather than the 20 January voyage departure as primary because it is independently tied to cohort enumeration and outbreak recognition, while preserving the broader voyage horizon as a transparent sensitivity analysis. Using 5 February as a later quarantine-start anchor tests the opposite direction. This hierarchy was specified before the final v1.4 numerical freeze and prevents the longest available interval from becoming the headline result merely because it produces the largest background expectation.

Life-table benchmark

The central analysis used the Japanese 2019 life table. [12] Official US 2019, UK 2017–2019 and Australian 2018–2020 life tables were used as reference-table sensitivities, not as a reconstruction of passenger nationality. [13–15]

For each single year of age and sex, the annual probability of death, (q_x), was converted to an interval risk over (w) days assuming a constant hazard:

$p_x(w) = 1 - (1 - q_x)^{(w/365.25)}$.

Within each 10-year passenger age band, single-year interval risks were averaged. Sex-specific interval risks were then weighted by the observed overall passenger composition (45% male, 55% female). Expected deaths were summed across age bands:

$E_{bg}(w) = \sum_a N_a [0.45\, p_{a,m}(w) + 0.55\, p_{a,f}(w)]$.

The primary benchmark ratio was

$R = D / E_{bg}(w)$,

where (D) was Outcome A or B. We call this the **background-mortality benchmark ratio** to avoid implying an SMR or causal excess-death estimate.

Sensitivity analysis

Sensitivity layers were separated by assumption. Japan-only sex-allocation bounds used all-female and all-male structures. Within-band lower, mean and upper single-year mortality assignments examined age-band uncertainty. Cross-national reference-table sensitivity changed the life table while fixing the 45%/55% allocation and band-mean ages. Finally, 75% and 50% mortality-retention scenarios were used only as one-direction healthy-traveller stress tests; they were not empirical corrections or probability intervals. Cruise-mortality studies were insufficiently comparable to identify a defensible passenger-specific multiplier. [6–8]

The final reproducibility grid contained 648 scenarios spanning anchors, life tables, sex structures, within-band assumptions and mortality-retention factors. All primary and sensitivity calculations were frozen under Analysis v1.4 / Protocol Amendment v1.4.

Ethics

This study used only publicly available aggregate data and did not involve identifiable private information or direct interaction with human participants. Institutional ethics review was therefore not required.

Results

Under the primary Japanese 2019 assumptions, expected all-cause background mortality over 72 days was **8.33 deaths** (unrounded, 8.329246). The benchmark ratio was **1.56** for Outcome A (13/8.33) and **1.68** for Outcome B (14/8.33). The background expectation was therefore **64% of the 13 reported deaths in numerical scale**, although the two quantities are not equivalent outcome constructs.

With demographics fixed, the four national life tables yielded **8.33–12.41 expected deaths** and Outcome A benchmark ratios of **1.05–1.56**. Japan 2019 produced the smallest background expectation of the four and therefore the largest ratio, making it the most conservative of these reference tables for the claim that background mortality is numerically material. These are alternative reference structures, not estimates of nationality-adjusted passenger mortality.

Under the Japanese life table alone, the all-female structural bound yielded **5.64 expected deaths** and a ratio of **2.31**; the all-male bound yielded **11.62** and **1.12**, so the reported count remained above the background benchmark even under the sex structure producing the largest Japanese expectation. Across national life tables, sex allocations and lower/mean/upper within-band age scenarios at 100% mortality retention, the wider structural envelope was **0.59–4.11**. These are stress bounds, not statistical intervals.

The illustrative healthy-traveller scenarios reduced the central expected count to 6.25 deaths at 75% mortality retention and 4.16 at 50%, increasing the corresponding ratios to 2.08 and 3.12.

Alternative aligned anchors did not change the qualitative result. The 86-day voyage-horizon analysis yielded **9.94 expected deaths** and an Outcome A ratio of **1.31**; the 70-day quarantine-start analysis yielded **8.10** and **1.61**.

Table 1 summarizes the three aligned-anchor results. Figure 1 displays the protocol-frozen sensitivity layers for Outcome A; the plotted ranges are structural sensitivity bounds rather than confidence intervals.

Discussion

The main result is straightforward: in this older cohort, 72 days of background mortality was not a negligible term. The central life-table expectation was 8.33 deaths, compared with 13 passenger deaths reported in Japan as outbreak-associated/COVID-related. The reported count was higher, but the background benchmark was 64% of that count in numerical scale, and the ratio remained 1.31–1.61 across three defensible aligned start dates.

The 64% comparison must not be read as meaning that 64% of the reported deaths "would have occurred anyway". It is a scale comparison between an expected all-cause reference process and an outbreak-associated death series with different ascertainment. The purpose is to show that the omitted background term is large enough to demand explicit treatment when the inferential target is population mortality. This is also why our result should not be interpreted as a correction to Russell and colleagues' IFR estimate: their denominator, follow-up logic and target quantity were different. [2] The methodological claim is narrower but more general—when an event count migrates from one estimand to another, the assumptions attached to its original use must be re-examined.

Estimand transfer rather than death reclassification

The methodological contribution is an **estimand-transfer diagnosis**. Early *Diamond Princess* analyses estimated infection-conditional severity. [1,2,4] When the same death count is later used to characterize mortality in the passenger population, the inferential target changes. A population denominator, observation horizon and background all-cause process become relevant. An assumption that is negligible for the original estimand cannot simply be inherited by the new one.

This complements, rather than duplicates, the recent risk-set argument by Chowell and Mizumoto. [11] Their cruise-ship example shows that an apparently complete onboard denominator may still be causally inappropriate if exposure pathways differ. Our analysis addresses the next layer: **even when the population count is accepted, the baseline required to interpret an outcome changes when the estimand changes**. Risk-set definition asks who belongs in the denominator; estimand transfer asks what reference process makes the resulting comparison meaningful.

Kobayashi and colleagues followed COVID-19 prognosis through 8 June among infected persons and strengthened evidence on age-dependent disease outcomes. [4] That longer disease-specific follow-up does not supply complete all-cause vital status for all 2,666 original passengers. It therefore improves the infection-severity estimand without resolving the population-mortality benchmark considered here.

Why 13 minus 8.33 is not a causal quantity

The life-table expectation cannot be subtracted from the reported outbreak-associated count to recover "true COVID-19 deaths". COVID-19 death and other-cause death are competing terminal events. A person dying from COVID-19 is no longer at risk of another death later in the same interval, while those who died with or from COVID-19 also carried non-zero baseline mortality risk before death. In addition, the available 13- or 14-death series is not a complete all-cause registry for the full passenger cohort. The joint counterfactual distribution of causes is therefore unidentified. The benchmark answers only whether the background term is small enough to ignore without checking.

Robustness and structural uncertainty

The conclusion does not depend on a single life table. Holding demographics fixed, four national references produced ratios between 1.05 and 1.56; Japan 2019 was the reference yielding the smallest background expectation. Nor does it depend on assuming exactly the observed overall sex mix within every age band: Japan-only all-female and all-male structural bounds produced ratios of 2.31 and 1.12. The midpoint-implied mean age of 66.17 years closely reproduces the reported mean of 66.1 and supports the overall age location, but it does not identify the within-band age distribution. The 0.59–4.11 full structural envelope therefore remains a set of stress bounds rather than a probability interval.

We deliberately do not attach a Poisson p value or similar significance test to 13 versus 8.33. Such a calculation would imply that both numbers arise from the same outcome-generating process, whereas one is an outbreak-associated series and the other an all-cause life-table expectation. The relevant uncertainty here is therefore structural—reference population, time

window, age allocation, sex allocation and baseline mortality—not sampling error around a single common stochastic model.

The healthy-traveller effect is less identifiable. General-population life tables may overstate baseline mortality among people healthy enough to cruise, but available cruise mortality studies do not provide the passenger exposure and post-disembarkation follow-up needed to estimate a transferable correction. [6–8] We therefore retain 75% and 50% only as illustrative one-direction stress tests.

A practical ignorability diagnostic

The *Diamond Princess* also suggests a general diagnostic. For age groups (a), let

$$E_{bg}(w) = \sum_a N_a \, p_a(w)$$

be expected background mortality over interval (w). If an age-specific outbreak model provides attack rates (AR_a) and infection-fatality risks (IFR_a), define

$$E_{out} = \sum_a N_a \, AR_a \, IFR_a.$$

Background mortality is plausibly ignorable for that inferential problem only when

$$E_{bg}(w) / E_{out} \ll 1.$$

This is a diagnostic, not a universal threshold. It makes explicit that ignorability is driven primarily by **baseline mortality × observation time**. The term may matter in older cruise populations, nursing homes, long-term-care settings and geriatric cohorts observed for weeks or months, while often being trivial in young cohorts over short windows.

For travel medicine, the implication extends beyond COVID-19. Cruise ships, expedition vessels, mass gatherings and repatriation cohorts often look attractive because the number of travellers is unusually visible. That visibility solves only part of the inferential problem. The denominator must still correspond to the exposure process, as recent cruise-ship work emphasizes, and the comparison baseline must correspond to the estimand. [11] These checks become particularly important when severe outcomes are counted in older travellers over observation periods long enough for ordinary mortality to accumulate.

Limitations

The principal limitation is outcome ascertainment. The MHLW outbreak-associated/COVID-related death series is not a complete all-cause registry for the 2,666 passengers, preventing causal decomposition. Individual person-time was also unavailable: passengers disembarked, were hospitalized, evacuated or repatriated on different dates. Fixed elapsed-time windows are therefore contextual benchmarks rather than exact person-time denominators.

Exact ages and age-by-sex counts were unavailable, so the central model applies the overall 45% male/55% female mix within each age band. Structural bounds quantify the direction of that uncertainty but do not reconstruct the true joint distribution. Nationality-by-age information was also insufficient for a multinational passenger-weighted life table; the four national tables are therefore sensitivity references only.

Finally, the constant-hazard conversion and fixed-window design are approximations, and general-population life tables may not represent healthy travellers perfectly. These limitations

widen uncertainty around the numerical benchmark but do not justify assuming the background term is zero.

Conclusions

Among 2,666 *Diamond Princess* passengers, the primary 72-day analysis yielded **8.33 expected background deaths** compared with 13 passenger deaths reported in Japan as outbreak-associated/COVID-related. The reported count was higher, but background mortality was not numerically negligible, and the benchmark ratio remained between 1.31 and 1.61 across aligned observation anchors.

The implication is methodological, not causal reclassification. Closed travel cohorts can offer unusually clean counts yet still invite inferential error when estimates are moved between questions. Infection-fatality and passenger-population mortality interpretations of the same outbreak death count require different reference structures. The *Diamond Princess* therefore illustrates an estimand-transfer problem that should be checked whenever outbreak counts from older cohorts are reused for population-level mortality claims.

Author statements

Author contributions

George Kujo: Conceptualization; methodology; formal analysis; investigation; data curation; writing—original draft; writing—review and editing; project administration.

Data availability

All aggregate cohort data were derived from publicly available sources cited in the article. A frozen reproducibility archive containing the passenger age-band dataset, corrected primary-results table, band-level Japanese 2019 reproducibility table, source manifest, complete 648-scenario sensitivity grid, analysis code and SHA-256 manifest is submitted as supplementary material with this manuscript. A Japanese-language author version of the work is publicly available on Zenodo at https://doi.org/10.5281/zenodo.22231591.

Code availability

The frozen analysis archive contains the inputs and code required to reproduce the primary calculation, the protocol-frozen sensitivity grid and Figure 1. Primary numerical implementation: **Analysis v1.4 / Protocol Amendment v1.4 — FINAL FREEZE**. Archive SHA-256 values are listed in the accompanying manifest.

Artificial intelligence assistance

Generative AI tools (OpenAI ChatGPT/Codex and Anthropic Claude) were used during manuscript development for language editing, code assistance, consistency checking and adversarial review. The author independently verified source claims, analytic inputs, calculations, methodological decisions and final wording, and takes full responsibility for the submitted work.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The author has declared no conflicts of interest.

References

1. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the *Diamond Princess* cruise ship, Yokohama, Japan, 2020. *Euro Surveill*. 2020;25(10):2000180. doi:10.2807/1560-7917.ES.2020.25.10.2000180.
2. Russell TW, Hellewell J, Jarvis CI, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the *Diamond Princess* cruise ship, February 2020. *Euro Surveill*. 2020;25(12):2000256. doi:10.2807/1560-7917.ES.2020.25.12.2000256.
3. Expert Taskforce for the COVID-19 Cruise Ship Outbreak. Epidemiology of COVID-19 outbreak on cruise ship quarantined at Yokohama, Japan, February 2020. *Emerg Infect Dis*. 2020;26(11):2591–2597. doi:10.3201/eid2611.201165.
4. Kobayashi T, Yoshii K, Linton NM, Suzuki M, Nishiura H. Age dependence of the natural history of infection with severe acute respiratory syndrome coronavirus 2

(SARS-CoV-2): an analysis of *Diamond Princess* data. *Int J Infect Dis*. 2022;115:109–115. doi:10.1016/j.ijid.2021.12.319.

5. Lewis N. COVID-19: Data from the Diamond Princess cruise ship implies that UK modelling hugely overestimates the expected death rates from infection. Web analysis. Originally published 25 March 2020; updated 27 March 2020. Non-peer-reviewed. https://nicholaslewis.org/covid-19-updated-data-implies-that-uk-modelling-hugely-overestimates-the-expected-death-rates-from-infection/. Accessed 30 August 2026.
6. Dahl E. Passenger mortalities aboard cruise ships. *Int Marit Health*. 2001;52(1–4):19–23. PMID: 11817837.
7. Oldenburg M, Herzog J, Püschel K, Harth V. Mortality of German travellers on passenger vessels. *J Travel Med*. 2016;23(1):tav003. doi:10.1093/jtm/tav003.
8. Lawson CJ, Dykewicz CA, Molinari NAM, Lipman H, Alvarado-Ramy F. Deaths in international travelers arriving in the United States, July 1, 2005 to June 30, 2008. *J Travel Med*. 2012;19(2):96–103. doi:10.1111/j.1708-8305.2011.00586.x.
9. Ministry of Health, Labour and Welfare, Japan. Death of a patient associated with the cruise ship quarantined at Yokohama (13th death). 15 April 2020. https://www.mhlw.go.jp/stf/newpage_10870.html. Accessed 1 September 2026.
10. Ministry of Health, Labour and Welfare, Japan. Press-release index for COVID-19 charter-flight and cruise-ship cases; cruise-ship section notes one death after Australian repatriation separately from the Japan-reported series. 2020. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000121431_00108.html. Accessed 1 September 2026.

11. Chowell G, Mizumoto K. Counting everyone onboard is not enough: modelling lessons from the MV Hondius Andes virus outbreak. *J Travel Med*. Published online 23 June 2026:taag036. doi:10.1093/jtm/taag036.
12. Ministry of Health, Labour and Welfare, Japan. Abridged Life Tables for Japan 2019. 2020. https://www.mhlw.go.jp/toukei/saikin/hw/life/life19/dl/life19-15.pdf. Accessed 1 September 2026.
13. Arias E, Xu J. United States Life Tables, 2019. *Natl Vital Stat Rep*. 2022. https://stacks.cdc.gov/view/cdc/231916. Accessed 1 September 2026.
14. Office for National Statistics. National life tables—life expectancy in the UK: 2017 to 2019. 2020. https://www.ons.gov.uk/releases/nationallifetablesuk2017to2019. Accessed 1 September 2026.
15. Australian Bureau of Statistics. Life tables, 2018–2020. 2021. https://www.abs.gov.au/statistics/people/population/life-expectancy/2018-2020. Accessed 1 September 2026.

Table 1. Central aligned-anchor results

Anchor	Expected background deaths	Outcome A	A benchmark ratio	Outcome B	B benchmark ratio
72 days: 3 Feb–15	**8.33**	13	**1.56**	14	**1.68**

Anchor	Expected background deaths	Outcome A	A benchmark ratio	Outcome B	B benchmark ratio
Apr (primary)					
86 days: 20 Jan–15 Apr	**9.94**	13	**1.31**	14	**1.41**
70 days: 5 Feb–15 Apr	**8.10**	13	**1.61**	14	**1.73**

Figure 1. Protocol-frozen sensitivity of the background-mortality benchmark ratio

Outcome A uses 13 reported passenger deaths. The point denotes the primary Japanese 2019, 72-day, 45% male/55% female, band-mean, 100%-mortality-retention scenario (ratio 1.56). Horizontal ranges show, respectively, aligned start-date anchors; four national life tables with demographics fixed; Japan-only sex-allocation bounds; illustrative healthy-traveller stress tests; and the full 100%-retention structural envelope across life table, sex allocation and within-band age scenario. The vertical reference at 1.0 denotes equality between the reported outbreak-associated count and the life-table background expectation. Ranges are sensitivity bounds, not statistical confidence intervals.