Quantifying the information in noisy epidemic curves

Kris V Parag*[,1,2], Christl A Donnelly[2,3], and Alexander E Zarebski[4]

[1]NIHR Health Protection Research Unit in Behavioural Science and Evaluation, University of Bristol, Bristol, BS8 2BN
[2]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, W2 1PG, UK
[3]Department of Statistics, University of Oxford, Oxford, OX1 3LB, UK
[4]Department of Zoology, University of Oxford, Oxford OX1 3SY, UK
*Email: kris.parag@bristol.ac.uk

Abstract

Reliably estimating the dynamics of transmissible diseases from noisy surveillance data is an enduring problem in modern epidemiology. Key parameters, such as the time-varying reproduction number, R_t at time t, are often inferred from incident time series, with the aim of informing policymakers on the growth rate of outbreaks or testing hypotheses about the effectiveness of public health interventions. However, the reliability of these inferences depends critically on reporting errors and latencies innate to those time series. While studies have proposed corrections for these issues, methodology for formally assessing how these noise sources degrade R_t estimate quality is lacking. By adapting Fisher information and experimental design theory, we develop an analytical framework to quantify the uncertainty induced by under-reporting and delays in reporting infections. This yields a novel metric, defined by the geometric means of reporting and cumulative delay probabilities, for ranking surveillance data informativeness. We apply this metric to two primary data sources for inferring R_t: epidemic case and death curves. We show that the assumption of death curves as more reliable, commonly made for acute infectious diseases such as COVID-19 and influenza, is not obvious and possibly untrue in many settings. Our framework clarifies and quantifies how actionable information about pathogen transmissibility is lost due to surveillance limitations.

Key-words: epidemic curves, noise, information, reproduction numbers, death counts, surveillance limits.

Introduction

The time-varying *effective reproduction number*, denoted R_t at time t, is an important and popular measure of the transmissibility of an unfolding infectious disease epidemic [1]. This parameter defines the average number of secondary infections generated by a primary case at t, providing a critical threshold for delineating growing epidemics ($R_t > 1$) from those likely to become controlled ($R_t < 1$). Estimates of R_t derived from surveillance data are widely used to evaluate the efficacies of interventions [2, 3] (e.g., lock-downs), forecast upcoming disease burden [4, 5] (e.g., hospitalisations), inform policymaking [1] and improve public situational awareness [6].

The reliability of these estimates depends fundamentally on the quality and timeliness of available surveillance data. Practical epidemic monitoring is subject to various errors or imperfections that can obscure or bias inferred transmission dynamics [7]. Prime among these are *under-reporting* and *reporting delays*, which can scale and smear R_t estimates, potentially misinforming public health authorities [8, 9]. The ideal data source for estimating R_t is the incident time series of infections, I_t. Unfortunately, as infections are rarely observed directly, the epidemic curve of reported cases, C_t, or that of death counts, D_t, is commonly used as a proxy [10, 11]. Both provide noisy approximations of the unknown I_t but with different and important relative advantages.

The epidemic case curve C_t records the most routinely available data i.e., counts of new cases [12], but is often limited by delays and under-reporting. Ascertainment delays smear or reorder case incidence and may emerge from fixed surveillance capacities, weekend effects and lags in diagnosing symptomatic patients [8, 13]. Delays may be classed as *occurred but not yet reported* (OBNR), when the source times of delayed cases eventually become known (i.e., delays basically cause right censoring of the case counts), or what we term as *never reported* (NEVR), when source times are not uncovered [14, 15].

Case under-reporting or under-ascertainment strongly distorts the true, but unknown, infection incidence curve, altering its size and shape [9, 16]. Temporal fluctuations in testing, behaviour-linked reporting (e.g., based on the severity of symptoms) [17], detection bottle-

necks and other surveillance biases can lead to under-ascertainment and inconsistent reporting [18]. *Constant reporting* (CONR) describes when the case detection fraction or probability is stable. We use the term *variable reporting* (VARR) for the more realistic scenario where this probability varies appreciably with time.

Death time series, D_t, count newly reported deaths attributable to the pathogen being studied and are also subject to under-reporting and reporting delays, but with two key differences [10]. First, death reporting delays incorporate an extra lag for the intrinsic time it takes an infection to culminate in mortality (this also subsumes hospitalisation periods). Second, apart from the under-reporting fraction of deaths, there is another scaling factor known as the infection fatality ratio, which defines the proportion of infections that result in mortality [2, 11]. We visualise how the noise types underlying epidemic and death curves distort infection incidence in Fig. 1.

Although the influences of surveillance latencies and under-ascertainment fractions on key parameters, such as R_t, are well known [8, 19–21] and much ongoing work attempts to compensate for these noise sources [10, 22–24], there exists no formal framework for assessing and exposing how they inherently limit information available for estimating epidemic dynamics. Most studies utilise simulation-based approaches (with some exceptions e.g., [9, 20]) to characterise surveillance defects, which while invaluable, preclude generalisable insights into how epidemic monitoring shapes parameter inference.

Here we develop one such analytic framework. Using *Fisher information* theory we derive a measure of how much usable information an epidemic time series contains for inferring R_t at every time. This yields metrics for cross-comparing different types of surveillance time series as we are able to explicitly quantify how under-reporting (both CONR and VARR) and reporting delays (exactly for OBNR with a tight upper bound for NEVR) degrade available information. As this metric only depends on the properties of surveillance (and not R_t or I_t) we extract simulation-agnostic insights into what are the least and most detrimental types of surveillance noise.

We prove for constrained mean reporting fractions and delays, that it is preferable to minimise variability among reporting fractions but to maximise the heterogeneity of the reporting delay distribution (such that a minority of infections face large delays but the majority possess short lags to notification). This proceeds from standard *experimental design* theory applied to our metric, which shows that the information embedded within an epidemic curve depends on the product of the geometric means of the reporting fractions and cumulative delay probabilities corrupting that curve. This central result also provides a

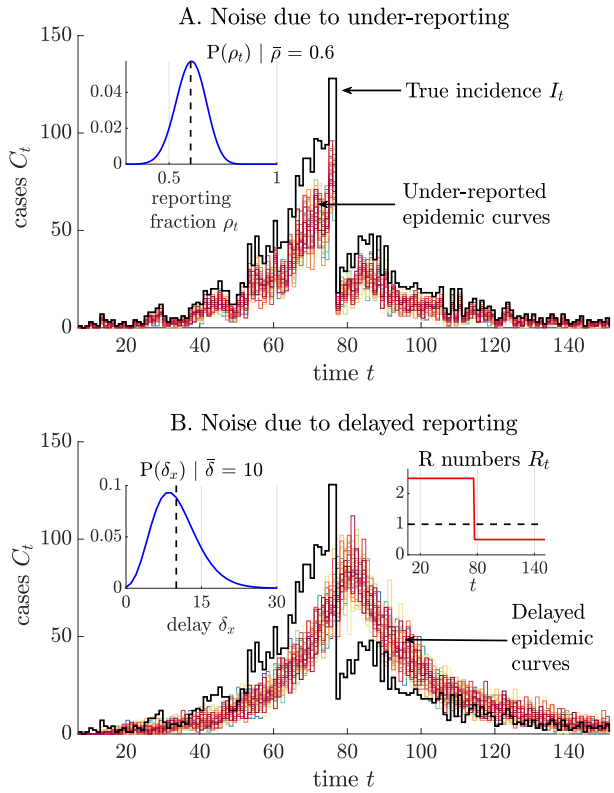

Fig. 1: **Under-reporting and delayed reporting noise.** We simulate true infection incidence I_t (black) from a renewal model (Eq. (1) with Ebola virus dynamics) with reproduction number R_t that switches from supercritical to subcritical spread due to some intervention. Panel A shows under-reported case curves (50 realisations, various colours) with reporting fractions sampled from the distribution in the inset. We observe stochastic trajectories and appreciable under-counting of peak incidence. Panel B considers delays in case reports (50 realisations, various colours) from the distribution plotted in the inset. We find variability and a smearing of the sharp change in incidence due to R_t (also provided as an inset). The main question of this study is how do we quantify which of these two scenarios incurs a larger loss of the information originally available from I_t, ideally without simulation.

non-dimensional score for summarising and ranking the reliability of (or uncertainty within) different surveillance data for inferring pathogen transmissibility.

Last, we apply this framework to explore and critique a common claim in the literature, which asserts that death curves are more robust for inferring transmissibility than case curves. This claim, which as far as we can tell has never been formally verified, is usually made for acute infectious diseases such as COVID-19 and pandemic in-

fluenza [2, 11], where cases are severely under-reported, with symptom-based fluctuations in reporting. In such settings it seems plausible to reason that deaths are less likely under-counted and more reliable for R_t inference.

However, we input COVID-19 reporting rate estimates [18, 25] within our metric and discover few instances in which death curves are definitively more informative than case counts. While this may not rule out that possibility, it elucidates and exposes how different noise terms within both data sources corrupt information, providing new methodology for exploring these types of questions more precisely. We also outline how other common data such as hospitalisations, prevalence and wastewater virus surveys conform to our framework. Hopefully the tools we developed here will improve quantification of noise and information and highlight the areas where enhanced surveillance strategies can maximise impact.

METHODS

Renewal models and Fisher information theory

The *renewal model* [26, 27] is a popular approach for describing how infections dynamically propagate during the course of an epidemic. The number of newly infected cases at time t, I_t, depends on the effective reproduction number, R_t, which counts the new infections generated per infected individual (on average) and the total infectiousness, Λ_t, which measures how many past infections (up to time $t-1$) will effectively produce new ones. This measurement weighs past case counts by the generation time distribution, $\boldsymbol{w}$, which we assume to be known. We define w_s as the probability that it takes s time units for a primary case to generate a secondary case. The distribution is then $\boldsymbol{w} = w_1^\infty := \{w_1, w_2, \ldots, w_\infty\}$.

The statistical relationship between these variables is often defined as in Eq. (1) with symbol Pois specifying a Poisson distribution [19]. This relationship strictly holds when I_t is perfectly observed both in size (no under-reporting) and in time (no delays in reporting).

$$I_t \sim \mathrm{Pois}\left(\Lambda_t R_t\right), \qquad \Lambda_t := \sum_{x=1}^{t-1} w_{t-x} I_x. \tag{1}$$

Eq. (1) has been widely used to model transmission dynamics of many infectious diseases, including COVID-19 [1], influenza [28] and Ebola virus disease [29].

A common and important problem in infectious disease epidemiology is the estimation of the latent variable R_t in real time from the observed incidence curve of infections. If this time series persists over $1 \leq t \leq \tau$, with τ as the present time, then we want to estimate the vector of parameters $R_1^\tau := \{R_t : 1 \leq t \leq \tau\}$ from the incidence time series $I_1^\tau := \{I_t : 1 \leq t \leq \tau\}$. We assume that time is measured in appropriate units such that R_t is expected to change (independently) at every time step. As illustrations, this time-step can be weekly for COVID-19 or Severe Acute Respiratory Syndrome (SARS) [2, 19] but monthly for diseases like rabies [30].

Following the development in [28, 31], we solve this inference problem by constructing the incidence log-likelihood function $\ell(R_1^\tau) = \log \mathbb{P}(I_1^\tau \mid R_1^\tau)$ as in Eq. (2) with K_τ as a constant that does not depend on any R_t.

$$\ell(R_1^\tau) = \sum_{t=1}^{\tau} I_t \log R_t - \Lambda_t R_t + K_\tau. \tag{2}$$

We compute the maximum likelihood estimate (MLE) of R_t as $\hat{R}_t$, which is the maximal solution of $\frac{\partial \ell(R_1^\tau)}{\partial R_t} = 0$. From Eq. (2) this gives $\hat{R}_t = I_t \Lambda_t^{-1}$ [26]. Repeating this for all t we obtain estimates of the complete vector of transmissibility parameters R_1^τ underlying I_1^τ.

To quantify the precision (the inverse of the variance, var) around these MLEs or any unbiased estimator of R_t we calculate the Fisher information (FI) that I_1^τ contains about R_t. This is $\mathbb{F}_I(R_t) := \mathbb{E}\left[-\frac{\partial^2 \ell(R_1^\tau)}{\partial R_t^2}\right]$, where expectation $\mathbb{E}[.]$ is taken across the data I_1^τ (hence the subscript I). The FI defines the best (smallest) possible uncertainty asymptotically achievable by any unbiased estimate, $\tilde{R}_t$. This follows from the Cramer-Rao bound [32], which states that $\mathrm{var}(\tilde{R}_t) \geq \mathbb{F}_I(R_t)^{-1}$. The confidence intervals around $\tilde{R}_t$ converge to $\tilde{R}_t \pm 1.96\, \mathbb{F}_I(R_t)^{-\frac{1}{2}}$. The FI also links to the Shannon mutual information that I_1^τ contains about R_t [33, 34] and is pivotal to describing both model identifiability and complexity [32, 35].

Using the Poisson renewal log-likelihood in Eq. (2) we compute the FI as the left equality in Eq. (3). Observe that this depends on the unknown 'true' R_t.

$$\mathbb{F}_I(R_t) = \Lambda_s R_t^{-1}, \qquad \mathbb{F}_I(2\sqrt{R_t}) = \Lambda_t. \tag{3}$$

This reflects the heteroscedasticity of Poisson models where the estimate mean and variance are co-dependent. We construct a square root transform that uncouples this dependence [31], yielding the right formula in Eq. (3). We can evaluate $\mathbb{F}_I(2\sqrt{R_t})$ purely from I_1^τ. The result follows from the Fisher information change of variables formula $\mathbb{F}_I(\mathcal{R}_t) = \mathbb{F}_I(R_t)\left(\frac{\partial R_t}{\partial \mathcal{R}_t}\right)^2$ [32]. This transformation has several optimal statistical properties [36, 37] and so we will commonly work with $\mathcal{R}_t := 2\sqrt{R_t}$.

As we are interested in evaluating the informativeness or reliability of the entire I_1^τ time series for inferring transmission dynamics we require the total FI it provides for all estimable reproduction numbers, R_1^τ. As we noted above, the inverse of the square root of the FI for a single R_t corresponds to an uncertainty (or confidence) interval. Generalising this to multiple dimensions yields an

uncertainty ellipsoid with volume inversely proportional to the square root of the determinant of the FI matrix [35, 36]. This matrix has diagonals given by $\mathbb{F}_I(R_t)$ and off-diagonals defined as $\mathbb{E}[-\frac{\partial^2 \ell(R_1^\tau)}{\partial R_t R_s}]$ for $1 \leq t,\, s \leq \tau$.

Maximising this non-negative determinant, which we denote the *total information* $\mathbb{T}(I_1^\tau)$ from the data I_1^τ, corresponds to what is known as a *D-optimal* design [38]. This design minimises the overall asymptotic uncertainty around estimates of the vector R_1^τ. As the renewal model in Eq. (1) treats every R_t as independent, off-diagonal terms are 0 and $\mathbb{T}(I_1^\tau)$ is a product of the diagonal FI terms. Transforming $R_t \to \mathcal{R}_t$ we then obtain Eq. (4).

$$\mathbb{T}(I_1^\tau) = \prod_{t=1}^{\tau} \sqrt{\mathbb{F}_I(\mathcal{R}_t)} = \prod_{t=1}^{\tau} \sqrt{\Lambda_t}. \tag{4}$$

If we work directly in R_t we get $\prod_{t=1}^{\tau} \Lambda_t^{\frac{1}{2}} R_t^{-\frac{1}{2}}$ instead. In two dimensions (i.e., $\tau = 2$) our ellipsoid becomes an ellipse and Eq. (4) intuitively means that its area is proportional to a product of lengths $\mathbb{F}_I(\mathcal{R}_1)^{-\frac{1}{2}} \mathbb{F}_I(\mathcal{R}_2)^{-\frac{1}{2}}$, which factors in the uncertainty from each estimate.

We will use this recipe of formulating a log-likelihood for R_1^τ given some data source and then computing the total information, $\mathbb{T}(.)$, it provides about these parameters to quantify the reliability of case, death and other time series for inferring transmissibility. Comparing data source quality will involve ratios of these total information terms. Metrics such as Eq. (4) are valuable because they measure the usable information within a time series and also delimit the possible distributions that a model can describe given that data (see [35, 39] for more on these ideas, which emerge from information geometry). Transforms like $\mathcal{R}_t = 2\sqrt{R_t}$ stabilise these metrics (i.e., maximise robustness) to unknown true values [36, 37].

Epidemic noise sources and surveillance models

We investigate two important and common sources of noise, under-reporting and reporting delay, which limit our ability to precisely monitor I_1^τ, the true time series of new infections. We quantify how much information is lost due to these noise processes by examining how these imperfections degrade $\mathbb{T}(I_1^\tau)$, the total information obtainable from I_1^τ under perfect (noiseless) surveillance for estimating parameter vector R_1^τ (see Eq. (4)). Fig. 1 illustrates how these two main noise sources individually alter the shape and size of incidence curves.

(i) Under-reporting or under-ascertainment. Practical surveillance systems generally detect some proportion of the true number of cases occurring at any given time t. If this proportion is $\rho_t \leq 1$ then the number of cases, C_t, observed is generally modelled as $C_t \sim \mathrm{Bin}\,(I_t,\, \rho_t)$ [22, 40], where Bin indicates the binomial distribution. The under-reported fraction is $1 - \rho_t$ and so the reported case count $C_t \sim \mathrm{Pois}\,(\rho_t \Lambda_t R_t)$. Reporting protocols are defined by choices of ρ_t. Constant reporting (CONR) is the simplest and most popular, assuming every $\rho_t = \rho$ [19]. Variable reporting (VARR) describes general time-varying protocols where every ρ_t can differ [41].

(ii) Reporting delays or latencies. There can be notable lags between the onset of an infected case and when it is reported [14]. If $\boldsymbol{\delta}$ defines the the distribution of these lags with δ_x as the probability of a delay of $x \geq 0$ time units, then the cases reported at t, C_t, sums the cases actually occurring at t but not delayed and those from previous days that were delayed [10]. This is commonly modelled as $C_t \sim \mathrm{Pois}\left(\sum_{x=0}^{t-1} \delta_x \Lambda_{t-x} R_{t-x}\right)$ [11, 41] and means that true incidence I_t splits over future times as $\sim \mathrm{Mult}(I_t,\, \boldsymbol{\delta})$, where Mult indicates multinomial [13]. The C_t time series is observed but not yet reported (OBNR) if we later learn about the past I_t splits (right-censoring), else we say data are never reported (NEVR).

We make some standard assumptions [8, 12, 19, 41] in incorporating the above noise sources within renewal model frameworks. We consider stationary generation time and reporting delay distributions only i.e., $\boldsymbol{w}$ and $\boldsymbol{\delta}$ do not vary with time and we neglect co-dependencies between reporting and transmissibility. Additionally, we assume these distributions and all reporting or ascertainment fractions, ρ_t, are inferred or known from other data (e.g., contact tracing studies) [13]. In the absence of these assumptions R_1^τ would be non-identifiable. We next examine how (i)-(ii) in combination limit the information available about epidemic transmissibility.

RESULTS

Renewal models with noisy observations

We denote the empirically observed or reported number of cases at time t, subject to noise from both under-reporting and reporting delays, as C_t with $C_1^\tau := \{C_t : 1 \leq t \leq \tau\}$ as the epidemic case curve. This curve is obtained from routine outbreak surveillance and is a corrupted version of the true incidence I_1^τ [10], modelled by Eq. (1). These noise sources (see Methods for statistical descriptions) are parametrised by reporting fractions ρ_1^τ and a delay distribution $\boldsymbol{\delta}$, which we assume to be known from other data (e.g. line-lists) [13, 42]. As a result, we can construct Eq. (5) as in [24] (see Methods).

$$C_t \sim \mathrm{Pois}\left(\sum_{x=1}^{t} \delta_{t-x} \rho_x \Lambda_x R_x\right). \tag{5}$$

This noisy renewal model suggests that C_t (unlike I_t) contains partial information about the entire time series of reproduction numbers for $x \leq t$ as mediated by delay

and reporting probabilities. Perfect reporting corresponds to $\rho_x = 1$ for all x as well as $\delta_0 = 1$ with $\delta_{x\neq 0} = 0$. The models in (i)-(ii) of the Methods are obtained by individually removing noise sources from Eq. (5).

Other practical epidemic surveillance data such as the time series of new deaths or hospitalisations conform to the framework in Eq. (5) either directly or with additional effective delay and under-reporting stages [11]. The main one we investigate here is the count of new deaths (due to infections) across time, which we denote $D_1^\tau := \{D_t : 1 \leq t \leq \tau\}$. The death curve involves a reporting delay that includes the intrinsic lag from infection to death. We let $\boldsymbol{\gamma}$ represent the distribution of time to observed death and use σ_1^τ for the fraction of deaths that are reported.

An important additional component when describing the chain from I_1^τ to D_1^τ is the infection fatality ratio, ifr_t, which is the probability at time t that an infection culminates in a death event [10]. Fusing these components yields Eq. (6) as a model for death counts D_t.

$$D_t \sim \mathrm{Pois}\left(\sum_{x=1}^{t} \mathrm{ifr}_x \gamma_{t-x} \sigma_x \Lambda_x R_x\right). \tag{6}$$

In a later section we also explain how this description fits other data streams such as hospitalisations and prevalence. Some studies [2, 43] replace this Pois formulation with a negative binomially (NB) distributed one to model extra variance in these data streams. In the Appendix we show that this does not disrupt our subsequent results on the relative informativeness of surveillance data (though the NB formulation is less tractable and unsuitable for extracting generalisable, simulation-free insights).

Fisher information derivations for practical data

We derive the FI of our vector of effective reproduction numbers R_1^τ given the case curve C_1^τ, which follows Eq. (5). We initially assume that reporting delays are OBNR i.e., that we eventually learn the source time of cases at a later date. This corresponds to a right censoring that can be compensated for using *nowcasting* techniques [14]. Later we prove that this not only defines a practical noise model but also serves as an upper bound on the information available from NEVR delays, where the true timestamps of cases are unresolvable. Mathematically, the OBNR assumption lets us to decompose the sum in Eq. (5). We can therefore identify the component of C_t that is informative about R_x. This follows from the statistical relationship $C_t \,|\, R_x \sim \mathrm{Pois}\,(\delta_{t-x}\rho_x\Lambda_x R_x)$.

As we are interested in the total information that C_1^τ contains about every R_t we collect and sum contributions from every C_t. We can better understand this process by constructing the matrix Q in Eq. (7), which expands the convolution of the reporting fractions with the delay probabilities over the entire observed time series.

$$Q = \begin{bmatrix} \delta_0\rho_\tau & \delta_1\rho_{\tau-1} & \delta_2\rho_{\tau-2} & \cdots & \delta_{\tau-1}\rho_1 \\ 0 & \delta_0\rho_{\tau-1} & \delta_1\rho_{\tau-2} & \cdots & \delta_{\tau-2}\rho_1 \\ \vdots & 0 & \delta_0\rho_{\tau-2} & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & \delta_1\rho_1 \\ 0 & 0 & \cdots & \cdots & \delta_0\rho_1 \end{bmatrix}. \tag{7}$$

We consider the vector $\mu = [\mu_\tau, \mu_{\tau-1}, \ldots, \mu_1]^\intercal$ with $\mu_t = \Lambda_t R_t$ and $\intercal$ denoting the transpose operation. Then $Q\mu = [\mathbb{E}[C_\tau], \mathbb{E}[C_{\tau-1}], \ldots, \mathbb{E}[C_1]]^\intercal$ with $\mathbb{E}[C_t]$ as the mean of the reported case incidence at time t.

The components of C_1^τ that contain information about every given reproduction number follow from $Q^\intercal\mu = [\delta_0\rho_\tau\mu_\tau, (\delta_0+\delta_1)\rho_{\tau-1}\mu_{\tau-1}, \ldots, (\delta_0+\ldots+\delta_{\tau-1})\rho_1\mu_1]$. The elements of this vector are Poisson means formed by collecting and summing the components of C_1^τ that inform about $[R_\tau, R_{\tau-1}, \ldots, R_1]$, respectively. Hence we obtain the key relationship in Eq. (8) with $F_{\tau-t} := \sum_{x=0}^{\tau-t}\delta_x$ as the cumulative probability delay distribution.

$$C_1^\tau \,|\, R_t \sim \mathrm{Pois}\,(\rho_t F_{\tau-t}\Lambda_t R_t). \tag{8}$$

The ability to decompose the row or column sums from Q into the Poisson relationships of Eq. (8) is a property of the independence properties of renewal models and the infinite divisibility of Poisson formulations.

Using Eq. (8) and analogues to Poisson log-likelihood definitions from the Methods we derive the Fisher information that C_1^τ contains about R_t as in Eq. (9).

$$\mathbb{F}_C(R_t) = \rho_t F_{\tau-t}\Lambda_t R_t^{-1}. \tag{9}$$

As in Eq. (3) we recompute the FI in Eq. (9) under the transform $\mathcal{R}_t = 2\sqrt{R_t}$ to obtain $\mathbb{F}_C(\mathcal{R}_t) = \rho_t F_{\tau-t}\Lambda_t$. It is clear that under-reporting and delays can substantially reduce our information about reproduction numbers. As we might expect, if $\rho_t = 0$ (no reports at time unit t) or $F_{\tau-t} = 0$ (all delays are larger than $\tau - t$) then we have no information on R_t at all from time series C_1^τ.

The MLE, $\hat{R}_t$, also follows from Eq. (8) (see Methods) as $(\sum_{x=t}^{\tau} C_x \,|\, R_t)(\rho_t F_{\tau-t}\Lambda_t)^{-1}$, with $C_x \,|\, R_t$ as the component of C_x containing information about R_t. By comparison with the MLE under perfect surveillance we see that $(\sum_{x=t}^{\tau} C_x \,|\, R_t)(\rho_t F_{\tau-t})^{-1}$ is equivalent to applying a nowcasting correction as in [13, 14]. An important point to make here is that while such corrections can remove bias, allowing inference despite these noise sources, they cannot improve on the information (in this case Eq. (9)) inherently available from the data. This is known as the data processing inequality [44, 45].

If we cannot resolve the components of every C_t from Eq. (5) as $\sum_{x=1}^{t}\mathrm{Pois}(\delta_{t-x}\rho_x\Lambda_x R_x)$, then our

reporting delay is classed as NEVR (i.e., we never uncover case source dates). Hence we know $Q\mu$ but not $Q^{\intercal}\mu$. Accordingly, we must use Eq. (5) to construct an aggregated log-likelihood $\ell(R_1^\tau) = \log \mathbb{P}(C_1^\tau \mid R_1^\tau) = \sum_{t=1}^{\tau} \log \mathbb{P}(C_1^\tau \mid R_t)$. This gives Eq. (10) with the aggregated term $h(R_1^t) := \sum_{x=1}^{t} \delta_{t-x}\rho_x\Lambda_x R_x$. We ignore constants that do not depend on any R_t in this likelihood.

$$\ell(R_1^\tau) = \sum_{t=1}^{\tau} C_t \log h(R_1^t) - h(R_1^t). \tag{10}$$

For every given R_t we decompose $h(R_1^s)$ for $s \geq t$ into the form $\delta_{s-t}\rho_t\Lambda_t R_t + a_t$, where a_t collects all terms that are not informative about that specific R_t. Here $s \geq t$ simply indicates that information about R_t is distributed across later times due to the reporting delays.

We can then obtain the FI contained in C_1^τ about R_t by computing $\mathbb{E}[-\frac{\partial^2 \ell(R_1^\tau)}{\partial R_t^2}]$, yielding Eq. (11) (see Appendix for derivation details), with $b_t := a_t(\delta_{s-t}\rho_t\Lambda_t R_t)^{-1}$.

$$\mathbb{F}_C(R_t) = \sum_{x=t}^{\tau} \delta_{x-t}\rho_t\Lambda_t(R_t + b_x)^{-1}. \tag{11}$$

If we could decouple the interactions among the reproduction numbers then the b_x terms would disappear and we would recover the expressions derived under OBNR delay types. Since b_x is a function of other reproduction numbers, the overall FI matrix for R_1^τ is not diagonal (there are non-zero terms from evaluating $\mathbb{E}[-\frac{\partial^2 \ell(R_1^\tau)}{\partial R_t R_x}]$).

However, we find that this matrix can be reduced to a triangular form with determinant equal to the product of terms (across t) in Eq. (11). We show this for the example case of $\tau = 3$ in the Appendix. As a result, the FI term for R_t in Eq. (11) does behave like and correspond to that in Eq. (9). Interestingly, as $b_x \geq 0$, Eq. (11) yields the revealing inequality $\mathbb{F}_C(R_t) \leq \rho_t F_{\tau-t}\Lambda_t R_t^{-1}$. This proves that OBNR delays upper bound the information available from NEVR delays. Last, we note that robust transforms cannot be applied to remove the dependence of Eq. (11) on the unknown R_t parameters. The best we can do is evaluate Eq. (11) at the MLEs $\hat{R}_t$ for all t.

These MLEs emerge as the joint maxima of the set of coupled differential equations $\frac{\partial \ell(R_1^\tau)}{\partial R_t} = \sum_{x=t}^{\tau} \frac{I_x}{b_x + R_t} - \delta_{x-t}\rho_t\Lambda_t$ i.e., numerical solutions of Eq. (12) for all t.

$$\sum_{x=t}^{\tau} C_x(\hat{R}_t + b_x)^{-1} = \rho_t F_{\tau-t}\Lambda_t. \tag{12}$$

Here sums start at t as they include only time points that contain information about R_t. Expectation-maximisation algorithms, such as *deconvolution* approaches outlined in [10], are viable means of computing these MLEs or equivalents. Note that the nowcasting methods used to correct for OBNR delays do not help here [13].

Reliability measures for surveillance data

Having derived the FI for each effective reproduction number above, we now provide a measure of the total information that C_1^τ provides about R_1^τ or the transformed $\mathcal{R}_1^\tau$. As detailed in the Methods, this total information, $\mathbb{T}(C_1^\tau)$, relates inversely to the smallest joint uncertainty around unbiased estimates of all our parameters [35]. As larger $\mathbb{T}(C_1^\tau)$ implies reduced overall uncertainty, this is a rigorous measure of the statistical reliability of noisy data sources for inferring pathogen transmissibility. Use of this or related metrics for quantifying the information in noisy epidemic data is novel (as far as we can tell).

We first consider the OBNR delay case under arbitrarily varying (VARR) reporting rates. Since the FI matrix under OBNR delays is diagonal, with each element given by Eq. (9), we can adapt Eq. (4) to derive Eq. (13).

$$\mathbb{T}(C_1^\tau) = \prod_{t=1}^{\tau} \sqrt{\mathbb{F}_C(\mathcal{R}_t)} = \prod_{t=1}^{\tau} \sqrt{\rho_t F_{\tau-t}\Lambda_t}. \tag{13}$$

Here we have applied the $\mathcal{R}_t = 2\sqrt{R_t}$ transformation to show that the total information in this noisy stream can be obtained without knowing R_t. In the absence of this transform we would have $\prod_{t=1}^{\tau} \sqrt{\rho_t F_{\tau-t}\Lambda_t R_t^{-1}}$.

Since C_1^τ is a distortion of the true infection incidence I_1^τ we normalise Eq. (13) by Eq. (4) to obtain a new reliability metric, $\eta(C_1^\tau) := \mathbb{T}(C_1^\tau)\mathbb{T}(I_1^\tau)^{-1}$. This is given in Eq. (14) and valid under both R_t and $\mathcal{R}_t$.

$$0 \leq \eta(C_1^\tau) = \prod_{t=1}^{\tau} \sqrt{\rho_t F_{\tau-t}} \leq 1. \tag{14}$$

We can relate this reliability measure to an effective (fixed) reporting fraction, $\theta(C_1^\tau)$, which causes an equivalent information loss. Applying Eq. (14) gives $\eta(C_1^\tau) = \sqrt{\theta(C_1^\tau)^\tau}$, which yields Eq. (15). Here $\mathbb{G}(.)$ indicates the *geometric mean* of its arguments over $1 \leq t \leq \tau$.

$$\theta(C_1^\tau) = \prod_{t=1}^{\tau} \sqrt[\tau]{\rho_t F_{\tau-t}} = \mathbb{G}(\rho_t)\mathbb{G}(F_{\tau-t}). \tag{15}$$

Eq. (15) is a central result of this work. It states that the total information content of a noisy epidemic curve is independently modulated by the geometric mean of its reporting fractions, $\mathbb{G}(\rho_t)$, and that of its cumulative delay probabilities, $\mathbb{G}(F_{\tau-t})$. Moreover, Eq. (15) provides a framework for gaining analytic insights into the influences of both noise sources from different surveillance data and for ranking the quality of those diverse data.

Eq. (15) applies to OBNR delays exactly and upper bounds the reliability of epidemic curves with NEVR delays (see previous section). Tractable results for NEVR delays are not possible and require numerical computation of Hessian matrices of $-\log \mathbb{P}(C_1^\tau \mid R_1^\tau)$ (we outline log-likelihoods and other equations for a $\tau = 3$ example

in the Appendix). However, we find that the Eq. (15) upper bound is tight for two fundamental settings. The first is under a constant or deterministic delay of d i.e., $\delta_{x=d} = 1$. Eq. (5) reduces to $C_t \sim \text{Pois}(\rho_{t-d}\Lambda_{t-d}R_{t-d})$. As each C_t only informs on R_{t-d} OBNR and NEVR delays are the same (and corrected by truncation).

The second occurs when transmissibility is constant or stable i.e., $R_t = R$ for all t. We can sum Eq. (9) to get $\mathbb{F}_C(R) = \sum_{t=1}^{\tau} \rho_t F_{\tau-t}\Lambda_t R^{-1}$ for OBNR delays. We can calculate the FI for NEVR delays from Eq. (10), which admits a derivative $\frac{\partial \ell(R)}{\partial R} = \sum_{t=1}^{\tau} C_t R^{-1} - F_{\tau-t}\rho_t\Lambda_t$ and hence a FI and MLE that are precisely equal to those for OBNR delays. This proves a convergence in the impact of two fundamentally different delay noise sources and emphasises that noise has to be contextualised with the complexity of the signal to be inferred. Simpler signals, such as a stationary R that remains robust to the shifts and reordering of I_1^τ due to delays, may be notably less susceptible to fluctuations in noise probabilities.

Ranking noise sources by their information loss

The metric proposed in Eq. (15) provides an original and general framework for scoring proxies of incidence (e.g., epidemic case curves, death counts, hospitalisations and others) using only their noise probabilities and without the need for simulations. We explore the implications of Eq. (15) both for understanding noise and ranking those proxies. The geometric mean decomposition allows us to separately dissect the influences of under-reporting and delays. We start by applying experimental design theory [36, 38], to characterise the best and worst noise types for inferring effective reproduction numbers.

We consider $\mathbb{G}(\rho_t)$, the geometric mean of the reporting probabilities across time. If we assume the average sampling fraction $\bar{\rho} = \frac{1}{\tau}\sum_{t=1}^{\tau}\rho_t$ is fixed (e.g., by some overall surveillance capacity) then we immediately know from design theory that $\bar{\rho} = \arg\max_{\boldsymbol{\rho}} \mathbb{G}(\rho_t)$. This means that of all the possible distributions of sampling fractions fitting that constraint, $\boldsymbol{\rho}$, CONR or constant reporting with probability $\bar{\rho}$ is the most informative [46]. This result is new but supports earlier studies recognising that CONR is preferred to VARR, although they investigate estimator bias and not information loss [9, 19].

Accordingly, we also discover that the worst sampling distribution is maximally variable. This involves setting $\rho_t \approx 1$ for some time subset $\mathbb{S}$ such that $\sum_{t\in\mathbb{S}}\rho_t = \tau\bar{\rho}$ with all other $\rho_t \approx 0$ (we use approximate signs as we assume non-zero sampling probabilities). Relaxing this constraint, Eq. (15) presents a framework for comparing different reporting protocols. We demonstrate these ideas in Fig. 2, where $\rho_t \sim \text{Beta}(a, b)$ i.e., each reporting fraction is a sample from a Beta distribution. Reporting protocols differ in (a, b) choices. We select 10^4 ρ_t samples each from 2000 distributions with $10^{-1} \leq b \leq 10^2$ and a computed to fulfil the mean constraint $\bar{\rho}$.

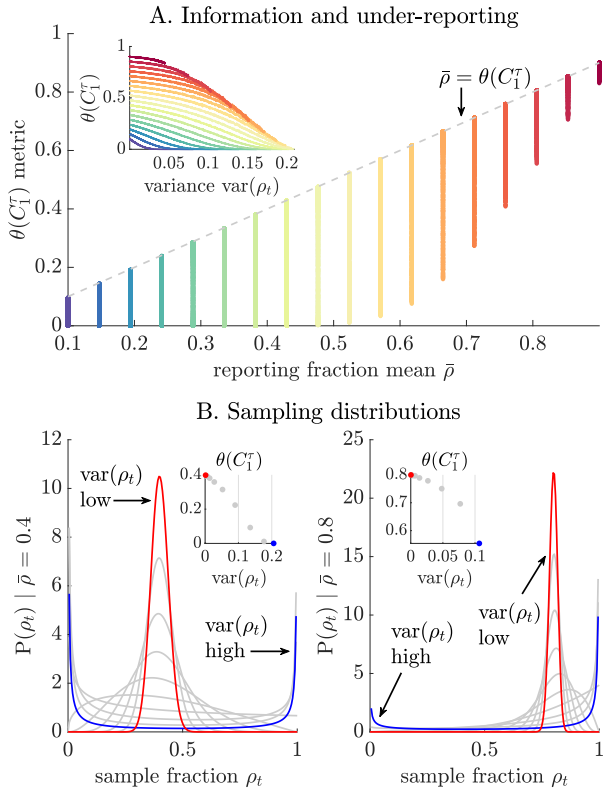

Fig. 2: **The information loss in under-reporting.** We investigate the effective information metric ($\theta(C_1^\tau)$) for variable reporting strategies (VAR) with reporting fraction ρ_t drawn from various Beta distributions. Panel A shows that while $\theta(C_1^\tau)$ depends on the mean reporting probability ($\bar{\rho}$), large fluctuations in the total information can emerge from the level of variability, controlled here by the Beta distribution shape. This is indicated by the overlap among metric values with different $\bar{\rho}$. The grey line (dashed) is the stable or constant reporting (CONR) protocol, which is optimal. Our metric decreases with the variance of the protocol (var(ρ_t)) as seen in the inset with colours indicating a given $\bar{\rho}$. Panel B illustrates the Beta sampling distributions and their resulting variance and metric scores (inset). The most variable reporting strategy (blue) is the worst protocol for a given $\bar{\rho}$.

Panel A of Fig. 2 shows that $\theta(C_1^\tau)$ generally increases with the mean reporting probability $\bar{\rho}$. However, this improvement can be denatured by the variance, var(ρ_t), of the reporting scheme (inset where each colour indicates the various schemes with a given $\bar{\rho}$). The CONR scheme

is outlined with a grey line (dashed) and, as derived, is the most informative. Panel B confirms our theoretical intuition on how $\mathrm{var}(\rho_t)$ reduces total information with the extreme (worst) sampling scheme outlined above in blue and the most stable protocol in red. There are many ways to construct ρ_t protocols. We chose Beta distributions because they can express diverse reporting probability shapes using only two parameters.

Similarly we examine reporting delays via $\mathbb{G}(F_{\tau-t})$, the geometric mean of the cumulative delay or latency distribution across time. Applying a mean delay constraint $\bar{\delta} = \sum_{x\geq 0} x\delta_x = \sum_{t=1}^{\tau}(1-F_{\tau-t})$ (e.g., reflecting operational limits on the speed of case notification), we adapt experimental design principles (by maximising a FI determinant our results are termed D-optimal) [46]. These suggest that $\max_{\boldsymbol{\delta}} \mathbb{G}(F_{\tau-t})$ is achieved by cumulative distributions possessing the largest δ_0 within this constraint. Delay distributions with significant dispersion (e.g., heavy tails) attain this optima while constant delays (where $\delta_{x\approx\bar{\delta}} = 1$ and 0 otherwise) lead to the largest information loss under this constraint.

This result may seem counter-intuitive as deterministic delays best preserve information outside of that delay and can be treated by truncating the observed epidemic time series (see previous section) e.g., for a fixed weekly lag we would ignore the latest week of data. However, this causes a bottleneck. No information is available for that truncated week eliminating any possibility of timely inference (and making epidemic control difficult [47]). In contrast, a maximally heterogeneous delay only slightly lags the majority of infected cases at the expense of large latencies for a few cases. This ensures that, overall, we gain more actionable information about the time series.

We illustrate this point (and relax the mean constraint) in Fig. 3, where we verify the usefulness of Eq. (15) as a framework for comparing the information loss induced by delay distributions of various shapes and forms. We model $\boldsymbol{\delta}$ as $\mathrm{NB}(k, \frac{\bar{\delta}}{\bar{\delta}+k})$ with k describing the dispersion of the delay. Panel A demonstrates how our $\theta(C_1^\tau)$ metric varies with k (30 values taken between 10^{-1} and 10^2) at various fixed mean constraints ($3 \leq \bar{\delta} \leq 30$, each given as a separate colour). As suggested by the theory, we see that decreasing k (increasing dispersion or delay heterogeneity) improves information at any given $\bar{\delta}$.

The importance of both the shape and mean of reporting delays is indicated in the inset as well as by the number of distributions (seen as intersects of the dashed black line) that result in the same $\theta(C_1^\tau)$. Panel B plots corresponding cumulative delay probability distributions, validating our assertion from design theory that the best delays (blue, with metric in inset) are heterogeneous and force $F_{\tau-t}$ high very early on (maximise δ_0), while the worst ones are more deterministic (red, larger k). These curves are for OBNR delays and upper bound the performance of NEVR delays except for the settings described in the previous section where both types coincide.

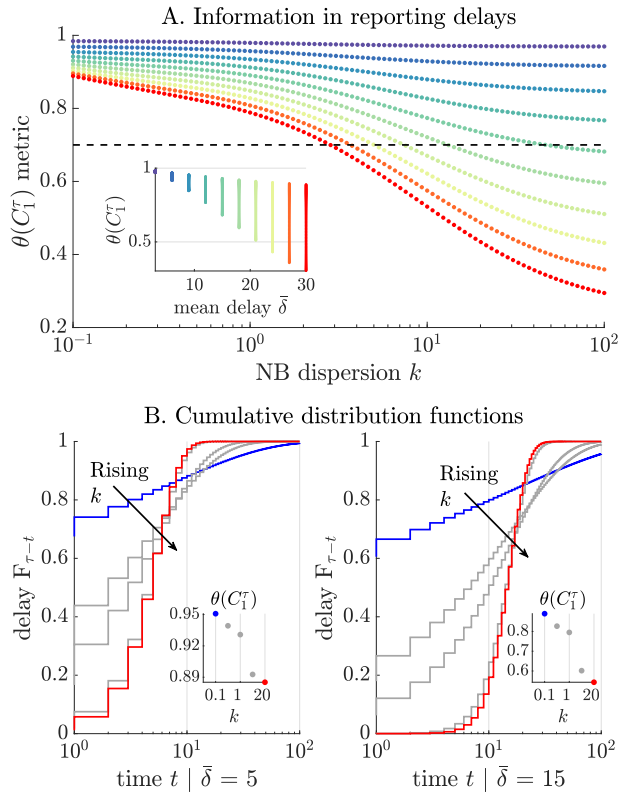

Fig. 3: **The information loss from delays.** We investigate the effective information metric ($\theta(C_1^\tau)$) for various delay distributions, which are negative binomial (NB) with dispersion k. Panel A examines how mean delay ($\bar{\delta}$) and k influence the information loss, showing that various combinations can result in the same loss (intersections of the coloured curves, each representing a different $\bar{\delta}$, with the dashed black line). Further, the inset illustrates the variations in our metric at a given mean due to the shape of the delay distribution. Panel B confirms this relationship and indicates that the most dispersed distributions (smallest k, blue, with largest start to cumulative delay distribution $F_{\tau-t}$) preserve the most information as compared to more deterministic delays (red, largest k). The insets additionally verify this point.

Comparing different epidemic data streams

Our metric (Eq. (15)) not only allows the comparison of different under-reporting schemes and reporting delay protocols (see above section) but also provides a common score for assessing the reliability or informativeness

of diverse data streams for inferring R_1^τ. The best stream, from this information theoretic viewpoint, maximises the product of the geometric means $\mathbb{G}(.)$ of the cumulative delay probabilities $F_{\tau-t}$ and reporting fractions ρ_t. Many common surveillance data types used for inferring pathogen transmissibility have been modelled within the framework of Eq. (5) and therefore admit related $\theta(.)$ metrics. Examples include time series of deaths, hospitalisations, the prevalence of infections and incidence proxies generated from viral surveys of wastewater.

We detail death count data in the following section but note that its model, provided in Eq. (6), is a simple extension of Eq. (5). Hospitalisations may be described similarly with the ifr term replaced by the proportion of infections hospitalised and the intrinsic delay distribution now defining the lag from infection to hospital admission [1]. The infection prevalence conforms to Eq. (5) because it can be represented as a convolution of incidence with a duration of infectiousness distribution, which essentially contributes a reporting delay [48]. Viral surveys also fit Eq. (5). They offer a downsampled proxy of incidence, which is delayed by a shedding load distribution defining the lag before a case is detected in wastewater [49]. Consequently, our metrics are widely applicable.

While in this study we focus on developing methodology for estimating and contrasting the information from the above surveillance data we find that our metric is also important for defining the complexity of a noisy renewal epidemic model. Specifically, we re-derive Eq. (15) as a key term of its *description length* (L). Description length theory evaluates the complexity of a model from how succinctly it describes its data (e.g., in bits) [35, 50]. This measure accounts for model structure and data quality and admits the approximation $L_C \approx -\ell(\hat{R}_1^\tau) + \frac{p}{2}\log\frac{m}{2\pi} + \log\int \det\left[\frac{1}{m}\mathbb{F}_C(R_1^\tau)\right] \mathrm{d}R_1^\tau$. Here the first term indicates model fit by assessing the log-likelihood at our MLEs $\hat{R}_1^\tau$. The second term includes data quality through the number of parameters (p) and data size (m). The final term defines how model structure shapes complexity with the integral across the parameter space of R_1^τ.

This formulation was adapted for renewal model selection problems in [31] assuming perfect reporting. We extend this and show that our proposed total information $\mathbb{T}(C_1^\tau)$ plays a central role. Given some epidemic curve C_1^τ we can rewrite the previous integral as $-\frac{p}{2}\log m + \log\prod_{t=1}^{\tau}\int\sqrt{\mathbb{F}_C(R_t)}\,\mathrm{d}R_t$ and observe that $m = p = \tau$. It is known that under a robust transform such as $\mathcal{R}_t = 2\sqrt{R_t}$ this integral is conserved [35, 36]. Consequently, $\int\sqrt{\mathbb{F}_C(R_t)}\,\mathrm{d}R_t = \sqrt{\mathbb{F}_C(\mathcal{R}_t)}\int_0^{2\sqrt{R_{\max}}} 1\,\mathrm{d}\mathcal{R}_t$ with $R_{\max}$ as some maximum value that every R_t can take. Combining these expressions we obtain Eq. (16), highlighting the importance of our total information metric.

$$L_C \approx -\ell(\hat{R}_1^\tau) + \frac{\tau}{2}\log\frac{2R_{\max}}{\pi} + \log\mathbb{T}(C_1^\tau). \qquad (16)$$

If we have two potential data sources for inferring R_1^τ then we should select the one with the smaller L_C value. Since the middle term in Eq. (16) remains unchanged in this comparison, the key points when comparing model complexity relate to the level of fit to the data and the total Fisher information of the model given that data [50]. Using Δ to indicate differences this comparison may be formulated as $\Delta L_C \approx -\Delta\ell(\hat{R}_1^\tau) + \Delta\log\mathbb{T}(C_1^\tau)$. The second term can be rewritten as $\Delta\log\eta(C_1^\tau)$ (see Eq. (14)). This signifies that our metrics play a central part when comparing different data streams.

Are COVID-19 deaths or cases more informative?

In the above sections we developed a framework for comparing the information within diverse but noisy data streams. We now apply these results to better understand the relative reliabilities of two popular sources of information about transmissibility R_1^τ; the time series of new cases C_1^τ and of new death counts D_1^τ. Both data streams have been extensively used across the ongoing COVID-19 pandemic to better characterise pathogen spread [1]. Known issues stemming from fluctuations in the ascertainment of COVID-19 cases [18, 21] have motivated some studies to assert D_1^τ as the more informative and hence trustworthy data for estimating R_1^τ [2, 11].

These works have reasonably assumed that deaths are more likely to be reliably ascertained. Case reporting can be substantially biased by testing policy inconsistencies and behavioural changes (e.g., symptom based healthcare seeking). In contrast, given their severity, deaths should be less likely to be under-ascertained [1]. However, no analysis, as far as we are aware, has explicitly tested this assumption. Here we make some progress towards better understanding the relative merits of both data streams. We start by computing ratios of our metric in Eq. (15) for both C_1^τ and D_1^τ via Eq. (5) and Eq. (6).

This gives Eq. (17) where $\succcurlyeq$ indicates greater than or equals in terms of total information with the infection to death cumulative distribution $H_{\tau-t} := \sum_{x=0}^{\tau-t}\gamma_x$.

$$C_1^\tau \succcurlyeq D_1^\tau : \mathbb{G}\left(\frac{\rho_t}{\mathrm{ifr}_t\sigma_t}\right) \geq \mathbb{G}\left(\frac{H_{\tau-t}}{F_{\tau-t}}\right). \qquad (17)$$

Eq. (17) states that case data is more informative if the geometric mean of the case to death reporting fractions is at least as large as that of the death and case delays. Studies preferring death data effectively claim that the variation in case reporting probabilities ρ_t (which we proved in a previous section always decreases the

geometric mean of sampled incidence at a given mean probability) is sufficiently strong to mask the influences of the infection fatality ratio (ifr_t), the death reporting probability (σ_t) and any variations in those quantities.

Proponents of using death data to infer R_1^τ recognise that the infection to death delay (cumulative distribution $H_{\tau-t}$) is appreciably larger in mean than that of corresponding reporting lags from infection ($F_{\tau-t}$) and therefore unsuitable for real time estimation (where this extra lag denatures recent information as we showed in above sections). We allow for all of these adjustments. We assume that the infection fatality ratio is constant at ifr (maximising $\mathbb{G}(\text{ifr}_t)$) and that death ascertainment is perfect ($\sigma_t = 1$). Even for purely retrospective estimation with correction for delays we expect $\mathbb{G}\left(\frac{H_{\tau-t}}{F_{\tau-t}}\right) \leq 1$. We allow this to be 1 maximising the informativeness of D_1^τ.

Combining these assumptions we reduce Eq. (17) into Eq. (18). This presents a sufficient condition for case data to be more reliable than the death time series.

$$C_1^\tau \succcurlyeq D_1^\tau : \mathbb{G}(\rho_t)_{\bar{\rho}\in[0.07,0.38]} \geq \text{ifr} \approx 0.01. \qquad (18)$$

Here we choose an ifr for COVID-19 of 1% [2]. Estimates of the mean sample fraction of cases range from about 7% to 38% [18], forming a constraint on $\bar{\rho}$, the average ρ_t. Using these figures we examine possible ρ_t sampling distributions under the $\text{Beta}(a, b)$ formulation from earlier sections. Our main results are in Fig. 4. We take 10^4 samples of ρ_t from each of 2000 distributions. These are parametrised over $10^{-1} \leq b \leq 10^2$ with a set to satisfy our mean $\bar{\rho}$ reporting constraints.

Panel A plots our metric against those constraints (a different colour for each $\bar{\rho}$) and the ifr threshold (black dashed). Whenever $\theta(C_1^\tau) \geq \text{ifr}$ we find that case data is more reliable. This appears to occur for many possible combinations of ρ_t. The inset charts the proportion of Beta distributions that cross that threshold. This varies from about 45% at $\bar{\rho} = 0.07$ to 90% at $\bar{\rho} = 0.38$. While these figures will differ depending on how likely a given level of variability is, they offer robust evidence that death counts are not necessarily more reliable. Even when deaths are perfectly ascertained ($\sigma_t = 1$) the small ifr term in D_1^τ means that 99% of the original incidence data is lost, contributing appreciable uncertainty.

These points are reinforced by the numerous assumptions we have made, which inflate the information in the death time series (e.g., it is likely $\sigma_t < 1$, $\text{ifr} < 0.01$ and neither is constant). Panel B displays the distributions of our sampling fractions with red (blue) indicating which shapes provide more (less) information than death data (see Eq. (18)). Note that these results hold for both real time and retrospective analyses as we ignored the noise

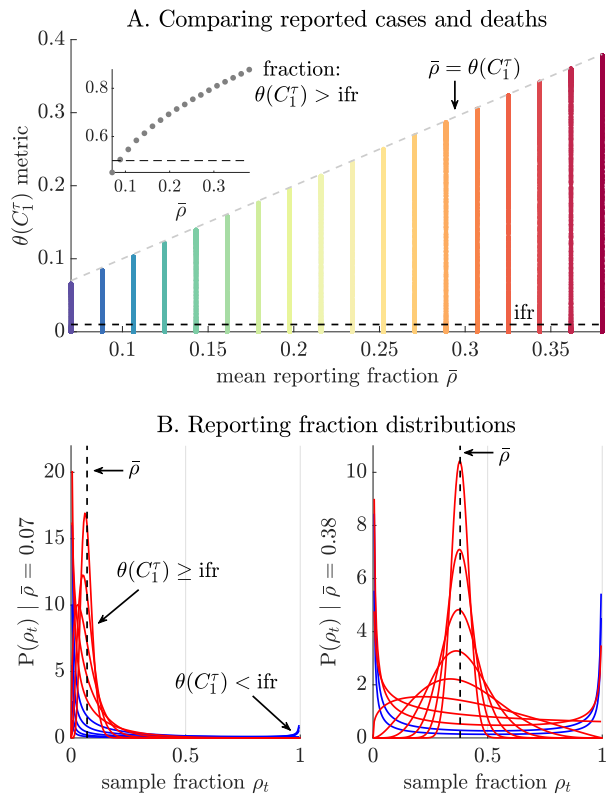

Fig. 4: **Epidemic case records may often be more informative than death counts.** Using metrics ($\theta(C_1^\tau)$) we compare the effective information in epidemic case curves C_1^τ and death counts D_1^τ under assumptions that lead to Eq. (18). We examine various reporting strategies parametrised as Beta distributions with means $\bar{\rho}$ ranging from 0.07 to 0.38 [18] and compare the resulting $\theta(C_1^\tau)$ against the equivalent from deaths (which reduces to just the infection fatality ratio, ifr). Panel A indicates that many such distributions for sampled cases still contain more information than available from deaths (proportion of vertical lines above the black dashed threshold, plotted inset). Panel B plots those distributions at the ends of the $\bar{\rho}$ range with red indicating when C_1^τ is more reliable. Substantial fluctuations in C_1^τ reporting can still preserve more information than might be found in D_1^τ.

induced by the additional delays that death data contain (relative to case reports) when we maximised $\mathbb{G}\left(\frac{H_{\tau-t}}{F_{\tau-t}}\right)$.

Consequently, death data cannot be assumed, without rigorous and context-specific examination, to be generally more epidemiologically meaningful. For example, while D_1^τ is unlikely to be more reliable in well-mixed populations, it may be in high-risk settings where the local ifr is notably larger (e.g., care homes). Vaccines, which substantially reduce ifr values in most contexts,

will make death time series less informative about R_1^τ. However, pathogens such as Ebola virus, which induce large ifr parameters, will likely lead to death data that is more reliable than their case counts.

DISCUSSION

Public health policymaking is becoming progressively data-driven. Key infectious disease parameters [6] such as effective reproduction numbers and growth rates, fitted to heterogeneous outbreak data sources (e.g., case, death and hospitalisation curves), are increasingly contributing to the evidence base for understanding pathogen spread, projecting epidemic burden and designing effective interventions [4, 6, 51]. However, the validity and value of such parameters depends substantially on the quality of surveillance data [1, 7]. Although many studies have made important advances in underscoring and correcting errors in these data [13, 42] no work (to our knowledge) has yet attempted to directly quantify their quality.

Here we have made some progress towards this aim. We applied Fisher information and experimental design principles to derive a novel framework for quantifying the information within common outbreak data when inferring pathogen transmissibility. Our approach involved finding the total information, $\mathbb{T}(.)$, available from epidemic curves corrupted by reporting delays and under-reporting. These are predominant noise sources that limit surveillance quality. By maximising $\mathbb{T}(.)$ we minimise the overall uncertainty of our transmissibility estimates, hence measuring the reliability of that data.

This approach yielded a new non-dimensional metric, $\theta(.)$, that allows analytic and generalisable insights into how noisy surveillance data degrades estimate precision. Using this metric we characterised the impact of different types of delay and under-reporting schemes. We demonstrated that under mean surveillance constraints, constant under-reporting of cases minimises loss of information. However, constant delays in reporting maximise this loss. The first result bolsters conventional thinking [9], while the second highlights the need for timely data [47].

Importantly, our metric provided insight into the nuances of noise, elucidating how the mean and variability of schemes both matter. For example, fluctuating reporting protocols with larger mean may outperform stable ones at lower mean. Exploiting this and the flexibility of our framework, which can describe the noise in cases, death counts, hospitalisations, infection prevalence and wastewater virus surveys, we demonstrated how diverse data sources could be ranked. Specifically, we critiqued a common assertion about death and case data.

Because the reporting of cases can vary significantly when tracking acute diseases such as COVID-19, various studies have assumed death data to be more reliable [1]. Using our metric, we presented one of the first qualifications of this claim. We found that the infection fatality ratio acts as an under-reporting factor with very small mean. Only the most severely varying case reporting protocols result in larger information loss, suggesting that in many instances this assertion may not hold. Note that this analysis does not even consider the additional advantages that case data bring in terms of timeliness.

However, there may be other important reasons for preferring death data (e.g., when little is known about the reporting protocol). As hospitalisation counts effectively interpolate among the types of noise in cases and deaths, this might serve as the best *a priori* choice of data for inferring transmissibility. Some studies also propose to circumvent these ranking issues by concurrently analysing multiple data sources [43, 51]. This then opens questions about how each data stream should be weighed in the resulting estimates. Our framework may also help by quantifying the most informative parts of contributing streams. A common way of deriving consensus weighs individual estimates by their inverse variance [52]. As the Fisher information defines the best possible inverse variance of estimates, our metrics naturally apply.

While our framework can enhance understanding and quantification of surveillance noise, it has several limitations. First, it depends on renewal model descriptions of epidemics [26]. These models assume homogeneous mixing and that the generation time distribution of the disease is known. While the inclusion of more realistic network-based mixing may not improve transmissibility estimates [53] (and this extra complexity may occlude insights), the generation time assumption is a true constraint that may only be ameliorated through the provision of updated, high quality line-list data [54].

Further, our analysis is contingent on having knowledge of the delays, under-ascertainment rates and other noise sources within data. These may be unavailable or themselves highly unreliable. To include this additional uncertainty is an important next step for this work, which will likely involve recomputing our metrics using posterior Fisher information terms [34, 55] that allow prior distributions on the noise parameters. We also assumed that the time scale chosen ensures that R_t parameters are independent. This may be invalid. In such instances we can append non-diagonal terms to Fisher information matrices or use our metric as an upper bound.

Last, we defined the reliability or informativeness of a data stream in terms of minimising the joint uncertainty of the entire sequence of reproduction numbers R_1^τ. This is known as a D-optimal design [38]. However, we may instead want to minimise the worst uncertainty among

the R_1^τ (which may better compensate known asymmetries in inferring transmissibility [56]). Our framework can be reconfigured to tackle such problems by appealing to other design laws. We can solve this specific problem by deriving an *E-optimal* design, which maximises the smallest eigenvalue of our Fisher information matrix.

Code Availability

All data and code (Matlab v2021a) necessary for reproducing the analyses and figures in this manuscript, as well as for applying the methods we have developed are freely available at https://github.com/kpzoo/information-in-epidemic-curves.

Acknowledgments

Thanks to Matthew Hickman for providing useful and interesting comments on the manuscript.

Funding

KVP acknowledges support from the NIHR Health Protection Research Unit in Behavioural Science and Evaluation at University of Bristol. CAD thanks the UK National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections for funding (grant no. HPRU200907).

Author Contributions

Conceptualization, investigation, methodology, formal analysis, funding acquisition and writing (original draft preparation): KVP. Validation: KVP and AEZ. Writing (review and editing): all authors.

References

[1] R. Anderson, C. Donnelly, D. Hollingsworth, *et al.*, "Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation," tech. rep., The Royal Society, 2020.

[2] S. Flaxman, S. Mishra, A. Gandy, *et al.*, "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe," *Nature*, vol. 584, pp. 257–261, 2020.

[3] Y. Li, H. Campbell, D. Kulkarni, *et al.*, "The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries," *Lancet Infect. Dis*, vol. 21, no. 2, pp. 193–202, 2020.

[4] S. Cauchemez, N. Hoze, A. Cousien, *et al.*, "How modelling can enhance the analysis of imperfect epidemic data," *Trends. Parasitol*, vol. 35, no. 5, pp. 369–79, 2019.

[5] S. Funk, A. Camacho, A. Kucharski, *et al.*, "Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the western area region of Sierra Leone, 2014-15," *PLoS Comput. Biol*, vol. 15, no. 2, p. e1006785, 2019.

[6] GOV.UK, "The R value and growth rate." https://www.gov.uk/guidance/the-r-value-and-growth-rate, 2021.

[7] S. Riley, K. Ainslie, O. Eales, *et al.*, "Resurgence of SARS-CoV-2: Detection by community viral surveillance," *Science*, vol. 372, no. 6545, pp. 990–5, 2021.

[8] K. Gostic, L. McGough, E. Baskerville, *et al.*, "Practical considerations for measuring the effective reproductive number, Rt," *PLoS Comput. Biol*, vol. 16, no. 12, p. e1008409, 2020.

[9] L. White and M. Pagano, "Reporting errors in infectious disease outbreaks, with an application to pandemic influenza A/H1N1," *Epidemiol. Perspec. Innov*, vol. 7, no. 12, 2010.

[10] E. Goldstein, J. Dushoff, J. Ma, *et al.*, "Reconstructing influenza incidence by deconvolution of daily mortality time series," *PNAS*, vol. 106, no. 51, pp. 21825–9, 2009.

[11] P. Nouvellet, S. Bhatia, A. Cori, *et al.*, "Reduction in mobility and COVID-19 transmission," *Nat. Comms*, vol. 12, p. 1090, 2021.

[12] J. Wallinga and P. Teunis, "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures," *Am. J. Epidemiol*, vol. 160, no. 6, pp. 509–16, 2004.

[13] P. Yang and G. Chowell, *Quantitative Methods for Investigating Infectious Disease Outbreaks*, vol. 70 of *Texts in Applied Mathematics*. Cham, Switzerland: Springer, 2019.

[14] J. Lawless, "Adjustments for reporting delays and the prediction of occurred but not reported events," *Can. J. Statist*, vol. 22, pp. 15–31, 1994.

[15] M. Salmon, D. Schumacher, K. Stark, *et al.*, "Bayesian outbreak detection in the presence of reporting delays," *Biometr. J*, vol. 57, no. 6, pp. 1051–67, 2015.

[16] B. Dalziel, M. Lau, M. Tiffany, *et al.*, "Unreported cases in the 2014-2016 Ebola epidemic: Spatiotemporal variation, and implications for estimating transmission," *PLOS Negl. Trop. Dis*, vol. 12, no. 1, p. e0006161, 2018.

[17] S. Funk, S. Bansal, C. Bauch, *et al.*, "Nine challenges in incorporating the dynamics of behaviour in infectious diseases models," *Epidemics*, vol. 10, pp. 21–5, 2015.

[18] G. Pullano, L. Di Domenico, C. Sabbatini, *et al.*, "Underdetection of cases of COVID-19 in France threatens epidemic control," *Nature*, vol. 590, pp. 134–9, 2021.

[19] A. Cori, N. Ferguson, C. Fraser, *et al.*, "A new framework and software to estimate time-varying reproduction numbers during epidemics," *Am. J. Epidemiol*, vol. 178, no. 9, pp. 1505–12, 2013.

[20] K. Parag, C. Donnelly, R. Jha, *et al.*, "An exact method for quantifying the reliability of end-of-epidemic declarations in real time," *PLoS Comput. Biol*, vol. 16, no. 11, p. e1008478, 2020.

[21] V. Pitzer, M. Chitwood, J. Havumaki, *et al.*, "The Impact of Changes in Diagnostic Testing Practices on Estimates of COVID-19 Transmission in the United States," *Am. J. Epidemiol*, vol. 190, no. 9, pp. 1908–17, 2021.

[22] C. Fraser, C. Donnelly, S. Cauchemez, *et al.*, "Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings," *Science*, vol. 324, no. 5934, pp. 1557–61, 2009.

[23] M. Hohle and M. der Heiden, "Bayesian Nowcasting during the STEC O104:H4 Outbreak in Germany, 2011," *Biometrics*, vol. 70, pp. 993–1002, 2014.

[24] S. Ali, A. Kadi, and N. Ferguson, "Transmission dynamics of the 2009 influenza (H1N1) pandemic in India: The impact of holiday-related school closure," *Epidemics*, vol. 5, pp. 157–63, 2013.

[25] R. Li, S. Pei, B. Chen, *et al.*, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)," *Science*, vol. 368, pp. 489–93, 2020.

[26] C. Fraser, "Estimating individual and household reproduction numbers in an emerging epidemic," *PLoS One*, vol. 8, p. e758, 2007.

[27] J. Wallinga and M. Lipsitch, "How generation intervals shape

the relationship between growth rates and reproductive numbers," *Proc. R. Soc. B*, vol. 274, pp. 599–604, 2007.
[28] C. Fraser, D. Cummings, D. Klinkenberg, *et al.*, "Influenza transmission in households during the 1918 pandemic," *Am. J. Epidemiol*, vol. 174, no. 5, pp. 505–14, 2011.
[29] WHO Ebola Response Team, "Ebola virus disease in West Africa – the first 9 months of the epidemic and forward projections," *N. Engl. J. Med*, vol. 371, no. 16, pp. 1481–95, 2014.
[30] H. Bourhy, E. Nakoune, M. Hall, *et al.*, "Revealing the Microscale Signature of Endemic Zoonotic Disease Transmission in an African Urban Setting," *PLoS Pathog*, vol. 12, no. 4, p. e1005525, 2016.
[31] K. Parag and C. Donnelly, "Adaptive estimation for epidemic renewal and phylogenetic skyline models," *Syst. Biol*, vol. 69, no. 6, pp. 1163–79, 2020.
[32] E. Lehmann and G. Casella, *Theory of Point Estimation*. Springer-Verlag, second ed., 1998.
[33] N. Brunel and J. Nadal, "Mutual information, Fisher information, and population coding," *Neural Comp*, vol. 10, no. 7, pp. 1731–57, 1998.
[34] K. Parag, O. Pybus, and C. Wu, "Are skyline plot-based demographic estimates overly dependent on smoothing prior assumptions?," *Syst. Biol*, vol. 71, no. 1, pp. 121–38, 2022.
[35] P. Grunwald, *The Minimum Description Length Principle*. The MIT Press, 2007.
[36] K. Parag and O. Pybus, "Robust design for coalescent model inference," *Syst. Biol*, vol. 68, no. 5, pp. 730–43, 2019.
[37] M. Bartlett, "The use of transformations," *Biometrics*, vol. 3, pp. 39–52, 1947.
[38] A. Atkinson and A. Donev, *Optimal experimental designs*. London, UK: Oxford University Press, 1992.
[39] I. Myung, V. Balasubramanian, and M. Pitt, "Counting probability distributions: Differential geometry and model selection," *PNAS*, vol. 97, no. 21, pp. 11170–5, 2000.
[40] T. Churcher, J. Cohen, N. Ntshalintshali, *et al.*, "Measuring the path toward malaria elimination," *Science*, vol. 344, no. 6189, pp. 1230–32, 2014.
[41] A. Azmon, C. Faes, and N. Hens, "On the estimation of the reproduction number based on misreported epidemic data," *Stats. Med*, vol. 33, pp. 1176–92, 2014.
[42] J. Polonsky, A. Baidjoe, Z. Kamvar, *et al.*, "Outbreak analytics: a developing data science for informing the response to emerging pathogens," *Phil. Trans. R. Soc B*, vol. 374, p. 20180276, 2019.
[43] J. Brauner, S. Mindermann, M. Sharma, *et al.*, "Inferring the effectiveness of government interventions against COVID-19," *Science*, vol. 371, no. 6531, p. eabd9338, 2021.
[44] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," *IEEE Trans. Info. Theo*, vol. 44, no. 3, pp. 1246–50, 1998.
[45] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, second ed., 2006.
[46] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and its Applications*. Springer Science + Business Media, second ed., 2011.
[47] F. Casella, "Can the COVID-19 Epidemic Be Controlled on the Basis of Daily Test Reports?," *IEEE Control Syst. Lett*, vol. 5, no. 3, pp. 1079–84, 2021.
[48] F. Vanni, D. Lambert, L. Palatella, *et al.*, "On the use of aggregated human mobility data to estimate the reproduction number," *Sci. Rep*, vol. 11, p. 23286, 2021.
[49] J. Huisman, J. Scire, L. Caduff, *et al.*, "Wastewater-based estimation of the effective reproductive number of SARS-CoV-2," *medRxiv*, vol. 2021.04.29.21255961, 2021.
[50] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Info. Theo*, vol. 42, no. 1, pp. 40–7, 1996.
[51] D. De Angelis, A. Presanis, P. Birrell, *et al.*, "Four key challenges in infectious disease modelling using data from multiple sources," *Epidemics*, vol. 10, pp. 83–7, 2015.
[52] J. Hartung, G. Knapp, and B. Sinha, *Statistical meta-analysis with applications*. Wiley Series in Probability and Statistics, New Jersey, USA: John Wiley and Sons, 2008.
[53] Q. Liu, M. Ajelli, A. Aleta, *et al.*, "Measurability of the epidemic reproduction number in data-driven contact networks," *PNAS*, vol. 115, no. 50, pp. 12680–85, 2018.
[54] S. Ali, L. Wang, E. Lau, *et al.*, "Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions," *Science*, vol. 369, no. 6507, pp. 1106–9, 2020.
[55] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior Cramer-Rao Bounds for Discrete-Time Nonlinear Filtering," *IEEE Trans. Signal Process*, vol. 46, no. 5, pp. 1386–95, 1998.
[56] K. Parag and C. Donnelly, "Fundamental limits on inferring epidemic resurgence in real time using effective reproduction numbers," *PLoS Comput. Biol*, vol. 18, no. 4, p. e1010004, 2022.
[57] J. Lloyd-Smith, S. Schreiber, P. Kopp, *et al.*, "Superspreading and the effect of individual variation on disease emergence," *Nature*, vol. 438, no. 17, pp. 355–9, 2005.

APPENDIX

Heterogeneous transmission noise models

For some infectious outbreaks the case (Eq. (5)) and death (Eq. (6)) count time series might be overdispersed [11] i.e., the mean-variance equality inherently assumed by the Poisson renewal model may not be valid. This can result when transmission is strongly influenced by heterogeneities in contacts and infectiousness or when incidence data are corrupted by additional intrinsic noise. These effects are often modelled by generalising Eq. (1) to a negative binomial (NB) form [2, 57], with dispersion k and success probability p. This leads to the expression below, with mean $\Lambda_t R_t$ and second argument as p.

$$I_t \sim \mathrm{NB}\left(k,\ \Lambda_t R_t (k + \Lambda_t R_t)^{-1}\right)$$

As we focus on the impact of the heterogeneity here, we assume perfect reporting (reporting noise as in the main text only affects the mean of this model). Taking derivatives of the log-likelihood corresponding to the NB model, $\ell(R_t)$, we get $\frac{\partial \ell(R_t)}{\partial R_t} = I_t R_t^{-1} - (I_t + k)\Lambda_t (k + \Lambda_t R_t)^{-1}$. This gives the MLE $\hat{R}_t = I_t \Lambda_t^{-1}$, which is the same as for Eq. (1). Computing $\mathbb{E}[-\frac{\partial^2 \ell(R_t)}{\partial R_t^2}]$ yields the Fisher information (FI) that I_t contains about R_t below.

$$\mathbb{F}_I(R_t) = \Lambda_t R_t^{-1} - \Lambda_t (R_t + k\Lambda_t^{-1})^{-1}$$

The first term above is the FI from Eq. (3). Heterogeneity therefore subtracts from its FI with a dispersion controlled term. As $k \to \infty$ this disappears and $\mathrm{NB} \to \mathrm{Pois}$.

While we do not explicitly include heterogeneity in the analyses of the main text, we do show, importantly, that our results do not qualitatively change if overdispersion

is included. There we rank epidemic and death curves, which have been corrupted by under-reporting and delays, according to their Poisson FI. Curves with larger FI are deemed more reliable sources of information about R_t. If the delay and under-reporting noise do not alter k (i.e., the level of transmission heterogeneity is stable), then this Poisson FI ordering remains valid. We confirm this in Fig. 5, demonstrating a monotonic relationship between the FI from both models at fixed k. Our main results are therefore moderately robust to heterogeneities.

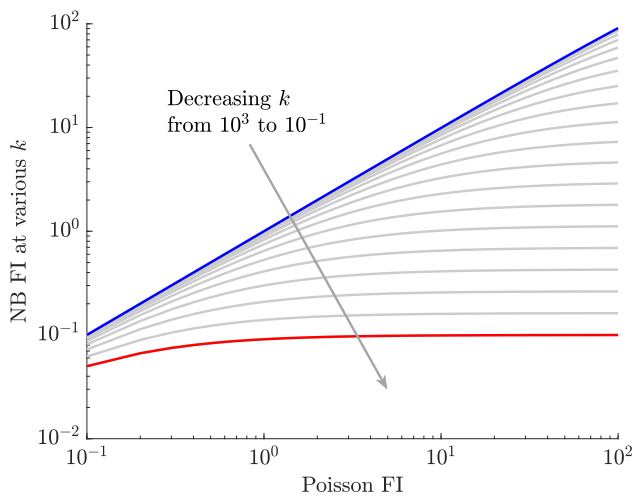

Fig. 5: **Overdispersion maintains Fisher information based rankings.** We plot the FI under a NB observation model that models overdispersion or heterogeneities in transmission, against the FI from the corresponding Poisson renewal model (with the same mean) at $R_t = 1$. For various dispersion parameters, k (smaller k means more heterogeneity), we find a monotonic ordering between these two FI values. This suggests that the rankings of data sources derived from their Poisson FI will likely also remain preserved under more complex NB models, provided those data sources have similar k values.

Information if case source times are never reported

Eq. (9) allows us to compute the FI of under-reported and delayed surveillance data. However, it assumes that we eventually uncover the true time of infection of the delayed cases. This is only valid for OBNR data [13], for which we can separate C_t from Eq. (5) into components $\sum_{x=1}^{t} \text{Pois}(\delta_{t-x}\rho_x\Lambda_x R_x)$ for every t. Here we compute the FI for the more complex case of NEVR delays, in which the source data of delayed cases are never reported and hence this sum cannot be decomposed. Specifically, we solve Eq. (10) for a three-dimensional example i.e., $\tau = 3$ and prove that we obtain a triangular FI matrix that has diagonal terms as in Eq. (11).

We then compute the total information, $\mathbb{T}(C_1^\tau)$ for never reported delays (the NEVR analogue to Eq. (13)) showing that it involves a product of those diagonal terms. This allows us to assert Eq. (13) as an upper bound and to prove convergence of these when the effective reproduction numbers are constant i.e., $R_t = R$ for all t. The algorithms we apply serve for any τ by inductively repeating the procedure here (for code see https://github.com/kpzoo/information-in-epidemic-curves).

We start from Eq. (10) with $\alpha_{(t-x)x} := \delta_{t-x}\rho_x\Lambda_x$ to define the complete log-likelihood below.

$$\ell(R_1^3) = \sum_{t=1}^{3} C_t \log \sum_{x=1}^{t} \alpha_{(t-x)x} R_x - \sum_{x=1}^{t} \alpha_{(t-x)x} R_x.$$

Taking derivatives for R_1 gives $\frac{\partial \ell(R_1^3)}{\partial R_1} = \frac{C_1}{R_1+b_1} - \delta_0 R_1 + \frac{C_2}{R_1+b_2} - \delta_1 R_1 + \frac{C_3}{R_1+b_3} - \delta_2 R_1$ with $b_1 = 0$, $b_2 = \frac{\alpha_{02}}{\alpha_{11}} R_2$ and $b_3 = \frac{\alpha_{12}}{\alpha_{21}} R_2 + \frac{\alpha_{03}}{\alpha_{21}} R_3$. This is repeated for all other R_t and conforms to the description in the main text. When solved and rearranged this gives the MLEs in Eq. (12).

We then calculate $\mathbb{E}[-\frac{\partial^2 \ell(R_1^\tau)}{\partial R_t R_x}]$ for every combination of t and x with the expectation of any C_t following as the mean in the Poisson model Eq. (5). This generates the complete FI matrix, $\mathbb{F}_C$, below with $\beta_1 = \alpha_{01} R_1$, $\beta_2 = \alpha_{02} R_2 + \alpha_{11} R_1$ and $\beta_3 = \alpha_{03} R_3 + \alpha_{12} R_2 + \alpha_{21} R_1$.

$$\mathbb{F}_C = \begin{bmatrix} \frac{\alpha_{01}^2}{\beta_1} + \frac{\alpha_{11}^2}{\beta_2} + \frac{\alpha_{21}^2}{\beta_3} & \frac{\alpha_{11}\alpha_{02}}{\beta_2} + \frac{\alpha_{21}\alpha_{12}}{\beta_3} & \frac{\alpha_{21}\alpha_{03}}{\beta_3} \\ \frac{\alpha_{11}\alpha_{02}}{\beta_2} + \frac{\alpha_{21}\alpha_{12}}{\beta_3} & \frac{\alpha_{02}^2}{\beta_2} + \frac{\alpha_{12}^2}{\beta_3} & \frac{\alpha_{12}\alpha_{03}}{\beta_3} \\ \frac{\alpha_{21}\alpha_{03}}{\beta_3} & \frac{\alpha_{12}\alpha_{03}}{\beta_3} & \frac{\alpha_{03}^2}{\beta_3} \end{bmatrix}.$$

The total information that C_1^3 contains about R_1^3 is already computable from the determinant of this matrix. However, we can obtain a more illuminating form by applying elementary column operations.

Such operations do not change the determinant. Using $\text{col}(x)$ for column x we successively apply $\text{col}(2) \to \text{col}(2) - \frac{\alpha_{12}}{\alpha_{03}}\text{col}(3)$, $\text{col}(1) \to \text{col}(1) - \frac{\alpha_{21}}{\alpha_{03}}\text{col}(3)$ and $\text{col}(1) \to \text{col}(1) - \frac{\alpha_{11}}{\alpha_{02}}\text{col}(2)$, revealing the triangular matrix below. This procedure of subtracting multiples of the later columns from earlier ones can be repeated for any τ to also extract analogous triangular forms.

$$\mathbb{F}_C = \begin{bmatrix} \frac{\alpha_{01}^2}{\beta_1} & \frac{\alpha_{11}\alpha_{02}}{\beta_2} & \frac{\alpha_{21}\alpha_{03}}{\beta_3} \\ 0 & \frac{\alpha_{02}^2}{\beta_2} & \frac{\alpha_{12}\alpha_{03}}{\beta_3} \\ 0 & 0 & \frac{\alpha_{03}^2}{\beta_3} \end{bmatrix}, \quad \mathbb{T}(C_1^3) = \prod_{t=1}^{3} \frac{\alpha_{0t}}{\sqrt{\beta_t}}.$$

The total information $\mathbb{T}(C_1^3)$ therefore depends on the diagonals of this triangular matrix as shown above.

When written out this corresponds to a product of the terms given in Eq. (11) for all t. Each term is smaller than or equal to the corresponding term from an OBNR

delay, given in Eq. (9). We can therefore use the total information in Eq. (13) to upper bound that available from a time series with NEVR delays. Intriguingly, this bound is sharp in some important instances. Specifically, if transmissibility is constant so $R_t = R$ for all t then $\frac{\partial \ell(R)}{\partial R} = \sum_{t=1}^{3} C_t R^{-1} - F_{3-t}\rho_t\Lambda_t$ and consequently $\mathbb{F}_C(R) = \sum_{t=1}^{3} F_{3-t}\rho_t\Lambda_t R^{-1}$. This is precisely the FI obtained under an OBNR delay (by summing Eq. (9)). The MLEs for both delay types are also equal.