Comparison of surveillance systems for monitoring COVID-19 in England: Lessons for disease surveillance

Summary

Background

During the COVID-19 pandemic, cases were tracked using multiple surveillance systems. Some systems were completely novel and others incorporated multiple data streams to estimate case incidence and/or prevalence. How well these different surveillance systems worked as epidemic indicators is unclear. This has implications for future disease surveillance and outbreak management.

Methods

Data from twelve surveillance systems used to monitor the COVID-19 in England were extracted (Jan20-Nov21). These were integrated as daily time-series and comparisons undertaken using Spearman correlation between candidate alternatives and the most timely (updated daily, clinical case register) and the least-biased (from comprehensive household sampling) COVID-19 epidemic indicators, with comparisons focused on the period Sep20-Nov21.

Findings

Spearman statistic correlations during the full focus period between least-biased indicator (from household surveys) and other epidemic indicator time series were 0.94 (clinical cases, the most timely indicator), 0.92 (self-report case status on a digital App), 0.67 (emergency department attendances), 0.64 (NHS111 website visits), 0.63 (wastewater concentrations), 0.60 (admissions to hospital with +COVID-19 status), 0.45 (NHS111 calls), 0.08 (Google search rank for 'covid'), -0.04 (consultations with general practitioners) and -0.37 (Google search rank for 'coronavirus'). Time lags (-14 to +14 days) did not markedly improve these rho statistics. Clinical cases (the most timely indicator) captured a more consistent proportion of cases than the self-report digital App did.

Interpretation

A suite of monitoring systems is useful. The household-survey system was a most comprehensive and least-biased epidemic monitor but not very timely. Data from laboratory testing, self-reporting digital App and attendances to emergency departments were comparatively useful, fairly accurate and timely epidemic trackers.

Keywords: surveillance ; epidemic ; COVID-19 ; symptoms ; health care seeking

Research in context

Evidence before this study

We searched PubMed on Jun 14, 2023, using search term "(("SARS-CoV-2" or "COVID-19") and "surveillance")[Title/Abstract] followed by forward and backward citation searches. 71 articles were relevant, 11 of these were UK-specific, most others came from high-income countries. Correlations with clinical case data from another source were determined for 57 studies about concentration of virus in wastewater samples, 10 studies about Internet search phrases, 2 studies with estimates from self-reporting (mobile phone) applications, and 2 studies with data generated by different population testing frameworks. Nearly all of these studies compared only two surveillance datasets with each other, maximum 4 different time series together, as epidemic indicators, typically with many adjustments for socio-demographic or other confounders, and sometimes for specific sampling sites (in case of wastewater samples). Most time series correlations were assessed using regression models or more sophisticated statistical analysis, often from monitoring over relatively short time periods, sometimes adjusting for individual wastewater catchments and allowing for many confounders.

Added value of this study

We consider epidemic monitoring suitable for an emerging data context for 12 surveillance systems simultaneously. We calculated relatively unadjusted correlations to monitor the Covid-19 epidemic in England in 2020-2021, comparing 9 alternatives to 3 nominated best (least biased or most timely) epidemic indicators. We wanted relatively unadjusted (for confounders) data, because this is how epidemic monitors tend to have to be used in daily decision making. The 9 datasets considered for potentially acceptably timely epidemic monitoring were cases suggested by a self-reporting application, syndromic surveillance datasets, wastewater monitoring, and Internet searches for relevant words. Integrating recent population survey estimates with data from a daily self-reporting symptom monitoring application was an effective approach to epidemic monitoring. Unadjusted data from wastewater sampling, Internet searches and health advice-seeking (rather than treatment seeking) activity did not provide good quality epidemic indicators.

Implications of all the available evidence

No ideal monitoring system (both very accurate and very timely) existed. Compared to the most timely or least biased estimates of prevalence and incidence, in England in 2020-2021, laboratory-confirmed case counts and emergency department attendances were the most timely and also independent indicators of concurrent epidemic status.

Introduction

Surveillance systems are an essential tool in the control of infectious diseases [1]. They were important during the COVID-19 pandemic [2]. COVID-19 surveillance was essential for monitoring trends in COVID-19 morbidity and mortality, identifying impacts on high-risk groups, informing modelling studies, targeting the delivery of health services and measuring the impact of vaccination and non-pharmaceutical interventions [3]. However, comparing different surveillance systems can be challenging due to their diverse characteristics. Even within a single country like the UK, several different surveillance systems were simultaneously in operation. It is unclear which systems gave the best balance of timely and reliable information about the COVID-19 pandemic progress. It is therefore prudent to evaluate the utility of these systems and how well they tracked COVID-19 incidence, particularly for decision-makers. Understanding the strengths and weaknesses of different surveillance systems is key to preparedness for future epidemics.

This is a retrospective observational study of COVID-19 surveillance systems (SS) in England, UK. Assessment of each data available via each surveillance system was undertaken with focus on potential sensitivity i.e., especially monitoring rises or falls in the number of cases over time. In addition to sensitivity, we discuss practical challenges in utilising specific SS for COVID-19. We also consider other aspects of how surveillance systems should be evaluated, especially with respect to timeliness and comprehensiveness [4]. Timely information is useful for having early understanding epidemic developments and to inform decision makers about when to make specific decisions with regard to epidemic control [5].

Our approach was to use two 'best standard' community-based COVID-19 surveillance datasets to compare to case counts suggested by other available SS that had apparent potential to indicate COVID-19 incidence and/or prevalence. We describe a systematic, replicable and consistent set of procedures for comparing case estimates that could be derived from these different SS, and implemented without sophisticated calibration or data adjustments. We used data in an unadjusted form (as published), because this unadjusted analysis is exactly what decision makers usually have to do with emerging data. We

examined correlations for both the full monitoring period and sub-periods. Finally, we discuss the utility of different systems for future outbreaks and factors influencing their effectiveness.

Methods

Best standard surveillance data

We judged that the best standard datasets in use for COVID-19 epidemic monitoring in 2020-2022 were from two sources. Comprehensive and *least-biased* case counts were from modelled estimates of new case counts (incidence) and concurrently infectious cases (prevalence), as generated by the Office for National Statistics (ONS) from data collected in their Coronavirus Infection Survey for England [6] (ONSCISE). The ONSCISE estimates were for swab dates. The estimates were typically updated weekly and retrospectively using post-stratification adjustments [7]. The ONSCISE was designed to detect COVID-19 infections through a nationally representative cohort sample of the entire population from May 2020 onwards, to generate estimates of both incidence (new case count) and prevalence (estimates of total concurrently infectious cases). The ONSCISE collected about 150,000 swab samples every two weeks [6]. By swabbing persons chosen at random, the ONSCISE had a high probability of detecting asymptomatic cases, and was otherwise designed to generate unbiased estimates of prevalence and incidence at daily or weekly periods. This system was also fairly representative as it was a random sample of the population aged 2 years and older and was adjusted for demographic variation between sampled and general population [8]. This system had high stability (data collection did not change and models only changed slightly) throughout the period of this analysis. However, given the time taken to collect, analyse and publish its estimates (10-24 day delays between swab collection and reporting date were typical), the ONS surveillance lacked timeliness.

The UK government also collated combined case counts from people testing positive for coronavirus under laboratory testing frameworks denoted as Pillar 1 (tests for those with occupational risk or clinical need) and Pillar 2 (tests for the general population). This was used as the second best standard dataset. Pillar 1 & 2 (P12) swab results [9,10] were mostly available within one day after sample collection [11], and thus were very timely and as well as the most comprehensive clinical case data available. We refer to P12 data as the *most timely* epidemic indicator, although biased by under-sampling of asymptomatic cases. By

"epidemic indicator", we mean a metric that indicates change in case counts. The dates we use for P12 case counts is for swab date not date that the information was published.

Comparator surveillance systems/datasets

Many other surveillance datasets were potential epidemic indicators, using methods other than counting known cases. Counts of persons newly admitted to NHS hospitals with COVID-19 were available publicly on a government website: coronavirus.data.gov.uk. Syndromic surveillance data with 'COVID-19-like' indicators (i.e. patients presenting to a NHS health service with symptoms likely to be caused by COVID-19) were accessed from the UK Health Security Agency (UKHSA) real-time syndromic surveillance service, which routinely monitors national syndromic SS to respond to a broad range of public health issues. Syndromic surveillance data comprised NHS 111 calls and online assessments, emergency department (EDSS) attendances and general practitioner (GP) consultations [12]. COVID-19 syndromic indicators were newly developed within existing syndromic surveillance systems early during the COVID-19 pandemic [13].

A further source of COVID-19 surveillance data was UKHSA wastewater sampling, specifically counts of SARs-CoV-2 viral genome copies in wastewater samples in most areas of England [14]. Wastewater samples had the merit that they should capture samples from asymptomatic cases. In addition, we made comparisons with three other potential surveillance datasets, all available in the public domain, as concurrent candidates for tracking the UK COVID-19 epidemic: ZoeApp incidence estimates [15] as they were originally published, and two search terms ("covid" and "coronavirus") from Google Trends [16]. ZoeApp was a nutritional health and wellness programme in development by early 2020 which was adapted to support symptom tracking during the pandemic. ZoeApp relied on users entering daily information about possible COVID-19 symptoms. ZoeApp published estimates of community and incidence prevalence that were derived from models that incorporated what percentage of their App users had reported new case status, with adjustments for demographic imbalances. The precise algorithm for how ZoeApp generated its incidence estimates does not appear to be published, but has been described by the App developers as based in part by combining ZoeApp data with historical ONS estimates [15 17]. That frequency of relevant search phrases on Internet search engines (such as Google)

might indicate underlying community prevalence or incidence of infectious diseases has been explored previously [18].

The 3 "best" datasets and the 9 other candidate COVID-19 SS and their derived datasets that are compared in this study are described in Table 1. All SS and datasets cover the geographical area of England only. Table S1 in the supplementary material describes the population coverage for each of the best standard and candidate alternative systems. By timeliness in Table 1, we mean the most typical delay after infection was detected until data publication. We note that some SS and datasets were not initially as comprehensive (at their start date) as they became by the start of the monitoring period (1 Sept 2020). For instance, Pillar 2 testing did not start until late May 2020, while wastewater sampling only started in July 2020 and over time coverage gradually increased. In contrast, UKHSA syndromic surveillance systems are in operation daily for routine all-hazard surveillance and were operational for COVID-19-like surveillance from March 2020. Data cleaning, additional processing and specific source details are described in Supplementary material.

Monitoring period

Narratively, we compare the datasets from earliest date of joint existence up to and including 30 November 2021. This end point was chosen because it precedes the emergence of the Omicron variant. The incubation period [19] , duration of shedding [20] and risk of reinfection with Omicron [21] all changed, compared to previous COVID-19 variants. Our quantitative analysis focuses on comparisons in the period September 1st 2020 to November 30th 2021. 1st September 2020 was deemed a plausible first date by when each of the SS selected for this study had likely established routine and standardised data collection and reporting procedures.

Table 1. Surveillance systems, derived dataset descriptions

	Short name	Description	Primary purpose of dataset	Types of cases most likely to be detected or indicated	Timeliness	Start date
Best standard datasets	P12 [9,10]	Pillar 1 & 2 case counts	Epidemic control and monitoring	Persons compliant with policy, having symptoms, medical need, &/or exposure	1-3 days	30 Jan 2020
	ONSincid [8]	ONSCISE incidence estimates	Estimate new cases, epidemic control	All population with new infections	17-24 days	11 May 2020
	ONSprev [8]	ONSCISE prevalence estimates	Estimate currently infectious cases, epidemic control	All population likely to be infectious and/or ill	10-17 days	26 Apr 2020
Comparator surveillance systems /datasets	HospAdm	New admissions to hospital	Epidemic monitoring and health care usage	More ill and/or vulnerable cases	1 day	20 Mar 2020
	EDSS [12]	Emergency Department attendances	Syndromic surveillance, health care usage	Patients with symptoms suggestive of COVID-19 infection who attended emergency department; possibly more severe cases	2 days	1 Mar 2020
	GPIH[12]	Consultations with GPs in usual hours	Syndromic surveillance, information seeking, care needs and care seeking	Patients with symptoms suggestive of COVID-19 infection who consulted a GP	1 day	4 Feb 2020
	111 calls[12]	NHS 111 telephone calls	Syndromic surveillance, health care information seeking and system usage	Patients with symptoms suggestive of COVID-19 infection who consulted NHS111 via phone	1 day	16 Mar 2020
	111 web[12]	NHS 111 online assessments	Syndromic surveillance, health care information seeking and system usage	Patients with symptoms suggestive of COVID-19 infection who consulted NHS111 via website	1 day	30 Mar 2020
	GTcov [18]	Google Trends search rank for “covid”	Monitor information seeking; user interests; commercial potential	Individuals who use Internet search engines	< 7 days	< 1 Jan 2020
	GTcor [18]	Google Trends search rank for “coronavirus”	Monitor information seeking; user interests; commercial potential	Individuals who use Internet search engines	< 7 days	< 1 Jan 2020

WW [14]	Wastewater	Passive epidemic monitoring	All cases in monitored areas	≥ 1 day	15 Jul 2020
ZoeApp [15] [17]	Zoe symptom tracker application	To monitor symptoms and disease progress associated with COVID-19 infection	Individuals who engaged with App	2 days	5 Jul 2020

Analysis: Quantitative comparisons

We (the authors) mostly treat ONSCISE incidence and prevalence as though they are interchangeable epidemic measures and that it is valid to compare them both with most of the other candidate datasets we describe. In actuality, two datasets (Zoe and P12) indicated incidence, one dataset (wastewater) indicated prevalence and all other datasets were imperfect indicators of incidence and prevalence. The other systems indicate prevalence or incidence because they can count the same patients in the same episode of illness multiple times (e.g. multiple calls for NHS health advice made by the same individual on one day). We have observed that many surveillance studies have not distinguished whether a system was measuring incidence and prevalence. For our study to have relevance to real world practice, we felt obliged to also blur the distinction between incidence and prevalence thus mimicking common practices applied in other surveillance system studies.

Otherwise, for comparisons, the ONSCISE estimates of incidence and prevalence, as well as the P12 case counts were used as the 'best standards'. Comparator datasets were compared to the likely best standards in daily time series. P12 was also compared to the ONSCISE estimates. We applied statistical comparisons between these time series that were simple and replicable. Spearman *rho* correlation statistics were calculated for the entire comparison period, as well as on a moving basis for 60 day periods: e.g., the value for 30 January 2021 describes time series correlation over the 60 day period from 1 January 2021 to 1 March 2021 inclusive. The 60-day period correlations were calculated to see if the time series correlations varied over a period of time that was neither very short nor the full period. We chose 60 day moving windows over other possible time divisions that might be meaningful for ascertainment, such as predominant COVID-19 variants, predominant symptom severity, public concern to seek treatment, likely vaccination status or data completeness because of the difficulty in consistently defining start and end dates for each of these conditions. Although we believe that the reporting systems were reasonably stable 1 September 2020 to 30 November 2021, there were still many reasons why ascertainment might vary over time, such as changes in criteria that made a person eligible for free testing [22], or if severity of disease declined after vaccination programmes which also may have reduced index of suspicion [23] and thus motivation to obtain tests. We report the median and interquartile range for the all-period and moving 60 day window correlations. A non-

parametric correlation test (Spearman rho) was appropriate because the data had strongly nonparametric distributions.

The (whole period) correlation analysis was replicated but with +/- 14 day lags (the values from best standard(s) were kept at original date, but the comparator data were moved by +/- 14 days). Ascertainment ratios were calculated as the candidate alternative case counts divided by the best estimates of actual cases (P12 or ONSCISE prevalence / ONS incidence). We only calculated ascertainment ratios for the hospital admissions, syndromic surveillance and ZoeApp datasets; the search ranks and wastewater viral count data were not suitable to compare to an estimate of case counts as denominator. We report the results with respect to a conventional significance indicator (p-value, significance threshold set at $p < 0.05$), but we note that correlation p-values must be carefully interpreted; they can be highly over-simplistic in describing closeness of variation between time series [24]. This is why we also report the coefficient of variation [25] (for the ascertainment ratios) as an alternative to Spearman rho with p-values. We present the ascertainment ratios as means (for all daily data) and coefficient of variation (CoV, standard deviation divided by the mean) in monitoring period. CoV can indicate consistency of a relationship.

The funder had no role in data collection, analysis, interpretation, writing of the manuscript or the decision to submit.

Results

Time series

Supplementary Figure S1 shows plots of the best standard times series (columns) against each candidate alternative (rows), with no time lags. All plots were designed to plot the candidate on left axis, and each best standard on right axis, and coincidence at bottom (value=0) and at their relative peaks (which varied). For these 30 illustrations (but not for actual comparisons and analysis), all time series were smoothed (7 day moving average on central date) except the ONSCISE incidence and prevalence which were already modelled data.

Full period correlations

Table 2 shows correlation (Spearman rho with p-value) between all candidate best standard and candidate alternative surveillance datasets over the entire monitoring period: 1 Sept 2020 to 30 Nov 2021 inclusive. Correlations are colour coded to make the table quick to interpret, with respect to correlation value: green 95% confidence interval ≥ 0.7, orange = 95% confidence interval below 0; grey = 95% confidence interval includes 0; black = significant (95% CI > 0) but < 0.7. Most correlations were significant at $p < 0.05$, meaning there was evidence of correlation; search rank for 'covid' is a noticeable exception with ONS incidence or prevalence. ZoeApp and COVID-19 attendances to ED had the highest correlation with the best standard data. Search ranks for "coronavirus" had a negative relationship, which possibly reflects changes in nomenclature preference (the term 'covid' came to be preferred) rather than evidencing lack of interest concurrent with high COVID-19 prevalence in community).

Moving 60 day windows

When moving 60 day windows are used to determine between time series correlation, Table 3 shows median values achieved for correlation (Spearman rho) and IQR of the central estimates of Spearman rho, comparing candidate and best standards. In these comparisons, the most consistently high correlations were observed for the ZoeApp with ONSCISE, with hospital admissions, Pillar 1 & 2 and ED attendances also highly correlated (lower boundary of 95% confidence interval for rho > 0.70) with the ONSCISE. Wastewater data were on average positively correlated overall with ONSCISE incidence, but also had great variability (IQR -.26 to 0.79).

P12	rho	**1**											
	ci	**(1, 1)**											
ONSincid	rho	0.943	**1**										
	ci	(0.92, 0.95)	**(1, 1)**										
ONSprev	rho	0.929	0.929	**1**									
	ci	(0.91, 0.95)	(0.91, 0.94)	**(1, 1)**									
HospAdm	rho	0.656	0.601	0.754	**1**								
	ci	(0.59, 0.72)	(0.52, 0.66)	(0.69, 0.80)	**(1, 1)**								
EDSS	rho	0.748	0.673	0.813	0.932	**1**							
	ci	(0.69, 0.80)	(0.60, 0.73)	(0.76, 0.86)	(0.91, 0.95)	**(1, 1)**							
GPIH	rho	0.141	-0.036	0.032	0.256	0.232	**1**						
	ci	(0.06, 0.23)	(-0.12, 0.05)	(-0.06, 0.12)	(0.16, 0.35)	(0.14, 0.32)	**(1, 1)**						
111 calls	rho	0.464	0.445	0.533	0.566	0.528	0.253	**1**					
	ci	(0.38, 0.54)	(0.36, 0.52)	(0.45, 0.61)	(0.49, 0.64)	(0.44, 0.60)	(0.16, 0.35)	**(1, 1)**					
111 web	rho	0.601	0.643	0.647	0.368	0.458	-0.120	0.163	**1**				
	ci	(0.56, 0.64)	(0.60, 0.68)	(0.61, 0.68)	(0.29, 0.43)	(0.40, 0.51)	(-0.20, -0.03)	(0.06, 0.25)	**(1, 1)**				
GTcor	rho	-0.296	-0.374	-0.235	0.219	0.125	0.435	0.452	-0.475	**1**			
	ci	(-0.39, -0.20)	(-0.46, -0.28)	(-0.33, -0.13)	(0.12, 0.31)	(0.03, 0.22)	(0.34, 0.52)	(0.35, 0.53)	(-0.55, -0.39)	**(1, 1)**			
GTCov	rho	0.117	0.079	0.084	0.133	0.188	0.149	0.179	-0.207	0.429	**1**		
	ci	(0.01, 0.21)	(-0.03, 0.18)	(-0.01, 0.18)	(0.03, 0.23)	(0.08, 0.28)	(0.05, 0.24)	(0.08, 0.27)	(-0.29, -0.12)	(0.34, 0.51)	**(1, 1)**		
WW	rho	0.658	0.627	0.654	0.674	0.750	0.167	0.517	0.335	0.110	0.149	**1**	
	ci	(0.59, 0.72)	(0.56, 0.69)	(0.59, 0.71)	(0.61, 0.73)	(0.70, 0.79)	(0.07, 0.26)	(0.43, 0.59)	(0.25, 0.40)	(-0.02, 0.19)	(0.05, 0.23)	**(1, 1)**	
ZoeApp	rho	0.907	0.921	0.920	0.560	0.644	-0.052	0.459	0.606	-0.398	-0.045	0.630	**1**
	ci	(0.88, 0.93)	(0.90, 0.94)	(0.90, 0.93)	(0.48, 0.63)	(0.56, 0.71)	(-0.14, 0.03)	(0.38, 0.54)	(0.55, 0.65)	(-0.49, -0.29)	(-0.14, 0.04)	(0.55, 0.69)	**(1, 1)**
		P12	**ONSincid**	**ONSprev**	**HospAdm**	**EDSS**	**GPIH**	**111 calls**	**111 web**	**GTcor**	**GTCov**	**WW**	**ZoeApp**

Table 2. Cross correlations between surveillance datasets. rho = Spearman rho estimate, possible range is -1 to +1, ci = 95% confidence interval for rho. Font colours with respect to correlation: Green: 95% confidence interval ≥ 0.7; black = 95% confidence interval is > 0 but lower bound is > 0.70; grey = 95% confidence interval crosses zero; orange 95% confidence interval < 0. See Table 1 for surveillance set descriptions.

Figure S2 in the Supplementary Material illustrates correlation between best standards and alternatives with respect to moving 60 day periods: the correlation plotted is for the middle most date in each period. Note that the vertical scale is consistent across rows (for each alternative dataset) but vertical scale is not consistent between columns. We varied vertical scales between systems to give the most information visually. No system had a constantly high correlation with ONSCISE. Pillar 1 &2 and ZoeApp had the highest average correlation, so were closest to reliably reflecting the ONSCISE estimates in these 60 day moving windows. GPIH and Google search terms were least likely to correlate with the ONSCISE in moving windows.

Ascertainment ratio

Table 4 gives ascertainment ratios: case counts indicated by other systems compared to the P12, ONSCISE incidence or prevalence were evidenced in the alternative daily case count data. Results are shown as mean ascertainment ratio per day and coefficient of variation (CoV: standard deviation divided by the mean) in monitoring period. CoV was lowest for P12 compared with ONSCISE data. The ZoeApp CoVs were next lowest, around 0.43 compared to P12 and either ONS dataset. The EDSS and hospital admissions datasets also provided relatively lower CoV values (respectively ≤ 0.63 and ≤ 0.77). In contrast, the GP in hours data had least consistent relationship (highest CoV values, ~1.60) with P12 or the ONSCISE datasets. Figure S4 in the Supplementary Material shows the variation in the ZoeApp ascertainment ratio against ONSCISE incidence, which ranged from 0.22 to 1.47, mean 0.65. The P12 ascertainment ratio was lower (0.49) than ZoeApp, with a range that was narrower (0.25 to 0.90), making P12 more consistent.

Leads or lags between datasets

Figure S3 in the Supplementary Material indicates the median (black line) full monitoring period correlation values achieved with +/- 14 day lead/lags for each candidate dataset, compared to each best standard, with shaded 95% confidence intervals around the medians. The x-axis shows the lead or lag (-14 to +14 days). Zigzag patterns appear where both datasets had day-of-week effects (when more cases reported on some days than others), especially evident for Pillar 1 & 2 with each of GP consultations, hospital admissions, NHS111 calls and wastewater. Again, the vertical scales are varied between

rows to give maximum visual information. Pillar 1 & 2 and ZoeApp estimates stayed closest to predicting ONSCISE allowing for possible lags (when the candidate might predict past positivity) or leads (when the candidate dataset might predict future positivity). Wastewater and the ZoeApp were particularly good as leads (anticipator) of total prevalence. Google search terms and GP in hours consultations were the poorest lead indicators of Pillar 1 & 2 or ONS prevalence/incidence.

Table 3. Correlation between candidates and best standard databases, in 60 day moving windows, median and IQR, centre-most dates are 1 Oct 2020 to 31 Oct 2021 inclusive

	Best standard datasets		
	P12	**ONS incid**	**ONS prev**
Comparator dataset			
P12	1.0 (1, 1)	0.69 (0.54, 0.84)	0.79 (0.64, 0.90)
HospAdm	0.73 (0.48, 0.67)	0.62 (0.16, 0.81)	0.82 (0.60, 0.75)
EDSS	0.75 (0.56, 0.90)	0.53 (0.19, 0.84)	0.89 (0.73, 0.96)
GPIH	0.54 (0.35, 0.64)	0.11 (-0.02, 0.22)	0.15 (-0.01, 0.23)
111 calls	0.30 (0.01, 0.74)	0.43 (0.22, 0.61)	0.39 (0.15, 0.75)
111 web	0.10 (-0.20, 0.41)	0.22 (-0.06, 0.44)	0.16 (-0.30, 0.85)
GTcor	0.18 (-0.11, 0.81)	0.24 (-0.36, 0.74)	0.47 (0.01, 0.84)
GTcov	0.37 (0.04, 0.69)	0.40 (0.03, 0.66)	0.41 (0.05, 0.70)
Wastewater	0.38 (-0.21, 0.81)	0.55 (-0.26, 0.79)	0.48 (0.07, 0.94)
ZoeApp	0.75 (0.49, 0.88)	0.74 (0.52, 0.89)	0.94 (0.81, 1.00)

Notes: median rho (Q1, Q3).

Table 4. Full monitoring period, daily ascertainment ratios, candidates / best standards

	Best standard datasets		
	P12 counts	**ONS incid**	**ONS prev**
Comparator dataset			
P12	1.0 (0)	0.49 (0.30)	0.04 (0.30)
HospAdm	0.05 (0.67)	0.03 (0.77)	0.00 (0.51)
EDSS	0.02 (0.55)	0.01 (0.63)	0.00 (0.40)
GPIH	3.22 (1.60)	1.74 (1.60)	0.12 (1.63)
111 calls	0.28 (1.28)	0.14 (1.19)	0.01 (1.19)
111 web	0.29 (1.22)	0.13 (1.17)	0.01 (1.28)
ZoeApp	1.38 (0.43)	0.65 (0.41)	0.05 (0.44)

Notes: Values expressed as mean (coefficient of variation).

Discussion

We focused analysis on 9 surveillance systems (SS) for COVID-19 monitoring in England from September 2020 to November 2021. With more sophisticated statistical analysis, perhaps adjusting for inherent sampling biases and confounders, any of the 9 SS might correlate very well with our nominated 3 best standards. However, we wanted to assess how well the data as published or with only the minimal adjustment (i.e., for wastewater) from each alternative SS correlated with the best standard datasets. All time series except ZoeApp and wastewater counts had poorer correspondence with the best standards after June 2021. This decline in correspondence possibly arises from the successful and rapid COVID-19 vaccination programme [26]. As noted elsewhere, the vaccination programme effectively 'broke the link' between positivity and health care needs [27]. The wastewater data show that prevalence continued to be high, however there was a deterioration in the sensitivity of the wastewater data to indicate incidence or prevalence from about September 2021 for unknown reasons. In contrast, the ZoeApp estimates persisted in corresponding closely with the ONSCISE data, especially for prevalence of COVID-19.

Although the absolute ascertainment ratio for ZoeApp estimates were high (mean 0.65) the ratio also varied greatly (0.22 to 1.47). This high variability is undesirable, because it suggests an inconsistent relationship. When a system captures the same percentage of cases consistently, there can be greater confidence that the system will accurately indicate whether the situation is improving or worsening. High consistency therefore can make ascertainment ratios more useful as epidemic predictors. Although the P12 ascertainment ratio was lower (0.49) than ZoeApp, its range was also much narrower (0.25 to 0.90) making it more consistent. P12 should therefore be interpreted as a more reliable indicator of epidemic progress than the ZoeApp data.

Infectious disease SS have three main functions [1] : (1) to describe the burden and epidemiology of disease, (2) to monitor trends, and (3) to identify outbreaks. In the early days of the pandemic, such information was essential for planning health care delivery, identifying those groups most at risk and targeting interventions aimed at reducing transmission. We have not explicitly analysed the data with regard to localised outbreaks

(function 3) but we do suggest that the P12 surveillance was best designed to do that. In terms of understanding the direction of travel of the UK epidemic (function 2, whether infections were increasing or decreasing), the P12 and ZoeApp achieved this most effectively and were timely as measured by their correlation with the ONS estimated incidence. Emergency Department attendance, hospital admissions and wastewater sampling also tracked incidence and prevalence relatively well. Many other SS (GP in hours, NHS 111 calls/web, Google Trends search terms) had little value for indicating trend. This may be because these services operated less for health care delivery and more in the role of providing information and reassurance. The ONSCISE was explicitly designed to be effective at detecting total burden of infection (function 1). Although overall, the ZoeApp estimates were highly correlated with the ONS incidence, the large variation in the ratio between ZoeApp and ONS incidence means that ZoeApp gave an inconsistent estimate of disease burden at any one time. P12 had slightly lower correlation but more consistency with ONSCISE estimates. Other systems gave less reliable estimates of total infections.

Other important considerations of surveillance systems are timeliness, availability and relative cost of each system. Most of the SS that we describe sampled broadly across the population (Table S1 in supplementary material) and were in the public domain and/or were made available to public health officials, so availability was relatively high. No system was perfect for timeliness but some systems reported within 24-36 hours. However, many systems detected cases only after symptoms developed which, in the case of COVID-19, is after infectiousness starts. COVID-19 specific SS (e.g., P12 and ONSCISE) were established in England rapidly to respond to the pandemic, but at significant cost. There was therefore merit in considering if pre-existing, routine SS (e.g. syndromic [13] and hospital admissions) could be adapted to monitor COVID-19 activity simply through the addition of new clinical codes. Indeed, UKHSA syndromic SS were utilised for all acute respiratory surveillance from January 2020 to provide reassurance of lack of changes in health care seeking behaviour for respiratory illness before any COVID-specific systems were in place. While more specific systems may provide more accurate estimates of COVID-19 activity and health system burden, if those estimates are quite delayed then having less sensitive, but more timely surveillance data is useful for planning the management and public health response to the pandemic. We do not argue that any one surveillance system (SS) can or should meet all

information needs. A complement of SS is preferable. That said, for decision-makers in resource-scarce settings, it is worthwhile to describe which SS are likely to be the best single epidemic monitors.

We demonstrate how changes in the national response to the pandemic affected the sensitivity of SS. For example, GP in-hours syndromic COVID-19-like consultation data became less sensitive to best standard systems due to increased access over time to rapid and free testing. Due to changes in dominant variants and fast vaccine rollout with high uptake (started December 2020), symptom presentation changed over the course of the COVID-19 pandemic and thus during our monitoring period [28]. Because symptoms varied over time, presentations by infected persons changed. These developments likely altered the most accurate ways to identify COVID-19 cases over time.

ZoeApp estimates were evidently generated by incorporating recent ONSCISE incidence/prevalence data, which means that ZoeApp *should* closely correspond with ONSCISE data (Table 3). As a result, the ZoeApp estimates are not independent of the ONS estimates. Our objective was not to only assess mutually independent data sources; the novel technology aspects of the ZoeApp and its unusual status as a participatory surveillance system made it worthwhile for consideration as an epidemic tracker. Integrating recently collected data with recent historical population-based survey and modelled incidence/prevalence (like ONSCISE) seems to have been a good strategy, producing a compromise in terms of timely and relatively accurate estimates of COVID-19 incidence and prevalence. This strategy could be adopted by other surveillance systems.

In contrast, the syndromic surveillance ED attendance data (EDSS) are independently recorded counts of attendances for persons attending an ED with a COVID-19-like diagnosis. For this reason, the EDSS data may be considered the better quality indicator if independent confirmation is desired. However, the good correspondence of many systems, such as the EDSS or hospital admissions data with ONSCISE, declined in summer 2021, as the full benefits of the vaccination programme became apparent and medical care demand due to COVID-19 infection subsided.

A limitation of our analysis is that we have not addressed possible completeness of each dataset. Full period response rates for ONSCISE and ZoeApp data are not available. The frequency of wastewater samples varied over time, and omissions in NHS service activity records are not uncommon in the authors' experience. The impacts of these data completeness issues on our results could be complex, although they also seem likely to have had been relatively minor. We did not analyse effects of variations in estimation methods over time (used by ONSCISE or ZoeApp, for instance), because this would be outside our aims (to use the candidate alternative data as they were published). To analyse how the estimation algorithms affected results would hugely complicate our analysis and neither the relevant raw data nor the precise estimation methods are publicly available. Our comparisons are a form of agreement analysis but they do not extend to assessing how differences between the SS estimates of case numbers might have changed estimates in key epidemic parameters such as the basic reproduction number.

There exist other UK surveillance datasets we could have also considered (eg., REACT study [29]), we cannot comment on their potential utility as epidemic monitors. For practicality, our analysis is very specific to a single region (England) in a single sovereign state (UK). The UK has been lauded for having especially complete, comprehensive and complementary COVID-19 surveillance systems [30]. Of relevance to settings with fewer resources, our analysis shows how fairly simple and relatively less expensive epidemic data (clinical case counts, self-reported case status and emergency department attendances) may serve as good proxies for less biased estimates of community incidence/prevalence.

In conclusion, none of the surveillance systems we tested here could meet all information needs. Different systems were most useful at different points in the pandemic. The ONSCISE estimates were probably least biased but were not timely. Syndrome surveillance was especially useful early in the pandemic because it was cheaply and quickly customised for case monitoring using existing surveillance systems. P12 and ZoeApp had reasonably good correlation with the ONSCISE; of these, the P12 data were most likely to suggest impacts on health care system. The ZoeApp practice of incorporating data that were only a few days old (from their App users) with the least biased source (ONSCISE) is a useful practice that other SS could emulate. However, a weakness of this integrated modelling

approach was sometimes delayed detection of rapid change. Methods used for collecting data and generating wider estimates should be well-documented and replicable. Overall we highlight the importance of range of different SS for epidemic tracking.

Author contributions

Analysis plan: IRL, RAM, AJE, PRH, JB
Calculating statistics: JB, PRH, NRJ
Conception: PRH, IRL
Data acquisition and extraction: JB, PRH, IRL, AJE
Data cleaning: PRH, NRJ, IRL, JB
Data curation: JB
Funding: IRL, PRH
Interpretation: PRH, IRL, AJE, RAM, JB
Research governance: AJE, IRL
Visualisations: JB, IRL
Writing first draft, assembling revisions: JB, IRL, PRH

Data sharing:

Applications for requests to access relevant anonymised data included in this study should be submitted to the UKHSA Office for Data Release at: https://www.gov.uk/government/publications/accessing-ukhsaprotected-data/accessing-ukhsa-protected-data)

Declaration of Interests

We declare no competing interests. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of our funders.

Ethical approval

The research protocol was approved by the UKHSA. The anonymised health data used in this study were either published (fully in the public domain) or collected and aggregated as part of the public health function of the UK Health Security Agency (UKHSA). UKHSA has access to a range of data sources under Regulation 3 (Health Protection) of The Health Service (Control of Patient Information) Regulations 2002. Informed consent from individuals was not required.

Acknowledgements

We thank syndromic data providers: NHS 111 and NHS England supporting NHS 111 surveillance; emergency department clinicians and NHS Trusts and NHS England supporting the Emergency Department Syndromic Surveillance System; and participating TPP (The Phoenix Partnership) SystmOne practices supporting GP in-hours surveillance. Thanks to staff at the UKHSA Environmental Monitoring for Health Protection project for providing wastewater data. The authors thank members of the UKHSA Real-time Syndromic Surveillance Team (Sally Harcourt, Sue Smith, Paul Loveridge, Christopher Bennett) for data

verification and technical support in accessing the syndromic data for this study. Gillian Smith (UKHSA) gave us comments to improve the manuscript.

Funding
This work was funded and authors were financially renumerated by the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Emergency Preparedness and Response at King's College London in partnership with the UK Health Security Agency (UKHSA) in collaboration with the University of East Anglia. AJE is also affiliated with the NIHR HPRU in Gastrointestinal Infections at the University of Liverpool. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, UEA, any HPRU, Department of Health or UKHSA. The funders had no role in writing the manuscript. Authors were not precluded from accessing data that supported this study and authors accept responsibility for this submission.

References

1. Murray J, Cohen AL. Infectious disease surveillance. *International Encyclopedia of Public Health* 2017: 222.
2. World Health Organization. Public health surveillance for COVID-19: interim guidance. Jul 22 2022. https://www.who.int/publications/i/item/WHO-2019-nCoV-SurveillanceGuidance-2022.2 (accessed Mar 9 2023).
3. Hunter PR, Colón-González FJ, Brainard J, Rushton S. Impact of non-pharmaceutical interventions against COVID-19 in Europe in 2020: a quasi-experimental non-equivalent group and time series design study. *Eurosurveillance* 2021; **26**(28): 2001401.
4. German RR, Horan JM, Lee LM, Milstein B, Pertowski CA. Updated guidelines for evaluating public health surveillance systems; recommendations from the Guidelines Working Group. 2001. https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm (accessed Mar 9 2023).
5. Wilkins K, Nsubuga P, Mendlein J, Mercer D, Pappaioanou M. The data for decision making project: assessment of surveillance systems in developing countries to improve access to public health information. *Public Health* 2008; **122**(9): 914-22.
6. Pouwels KB, House T, Pritchard E, et al. Community prevalence of SARS-CoV-2 in England from April to November, 2020: results from the ONS Coronavirus Infection Survey. *The Lancet Public Health* 2021; **6**(1): e30-e8.
7. Pouwels KB, Eyre DW, House T, et al. Improving the representativeness of UKs national COVID-19 Infection Survey through spatio-temporal regression and post-stratification. *medRxiv* 2023: 2023.02. 26.23286474.
8. Office for National Statistics. Protocol and information sheets. 2020-2023. https://www.ndm.ox.ac.uk/covid-19/covid-19-infection-survey/protocol-and-information-sheets (accessed Apr 18 2023).
9. Department of Health and Social Care. Coronavirus (COVID-19) scaling up our testing programmes. Apr 4 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/878121/coronavirus-covid-19-testing-strategy.pdf (accessed Apr 21 2023).
10. Elson R, Davies TM, Lake IR, et al. The spatio-temporal distribution of COVID-19 infection in England between January and June 2020. *Epidemiology & Infection* 2021; **149**: e73.
11. UK Health Security Agency. Weekly statistics for NHS Test and Trace (England): 18 to 24 November 2021. Dec 2 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1038052/test-and-trace_week-78.pdf (accessed Aug 17 2023).

12. Ferraro CF, Findlater L, Morbey R, et al. Describing the indirect impact of COVID-19 on healthcare utilisation using syndromic surveillance systems. *BMC Public Health* 2021; **21**(1): 2019.
13. Elliot AJ, Harcourt SE, Hughes HE, et al. The COVID-19 pandemic: a new challenge for syndromic surveillance. *Epidemiology & Infection* 2020; **148**.
14. Wade MJ, Jacomo AL, Armenise E, et al. Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: Lessons learned from the United Kingdom national COVID-19 surveillance programmes. *Journal of Hazardous Materials* 2022; **424**: 127456.
15. Zoe Health Study. COVID estimates revised after change to methodology. May 12 2021. https://health-study.joinzoe.com/post/covid-estimates-revised-after-change-to-methodology (accessed Mar 10 2023).
16. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases* 2009; **49**(10): 1557-64.
17. JoinZoe. COVID Estimates Updated. Mar 2 2023. https://health-study.joinzoe.com/blog/covid-estimates-updated (accessed Apr 18 2023).
18. Li S, Zhou X. Research of the correlation between the H1N1 morbidity data and Google Trends in Egypt. *arXiv preprint arXiv:151105300* 2015.
19. Tanaka H, Ogata T, Shibata T, et al. Shorter incubation period among COVID-19 cases with the BA. 1 Omicron variant. *International Journal of Environmental Research and Public Health* 2022; **19**(10): 6330.
20. Kang SW, Kim JY, Park H, et al. Comparison of secondary attack rate and viable virus shedding between patients with SARS-CoV-2 Delta and Omicron variants: A prospective cohort study. *Journal of Medical Virology* 2023; **95**(1): e28369.
21. Pulliam JR, van Schalkwyk C, Govender N, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science* 2022; **376**(6593): eabn4947.
22. Wise J. Covid-19: What's going wrong with testing in the UK? *British Medical Journal* 2020; **370**.
23. Grissinger M. Raising the Index of Suspicion: Red Flags That Represent Credible Threats to Patient Safety. *Pharmacy and Therapeutics* 2015; **40**(11): 713.
24. Aggarwal R, Ranganathan P. Common pitfalls in statistical analysis: The use of correlation techniques. *Perspectives in Clinical Research* 2016; **7**(4): 187.
25. Shechtman O. The coefficient of variation as an index of measurement reliability. *Methods of Clinical Epidemiology* 2013: 39-49.
26. Iacobucci G, Mahase E. Covid-19: Vaccine success drives England's lockdown exit. *British Medical Journal* 2021; **372**(528).
27. Public Health England. Vaccines highly effective against hospitalisation from Delta variant. June 14 2021. https://www.gov.uk/government/news/vaccines-highly-effective-against-hospitalisation-from-delta-variant (accessed Mar 9 2023).
28. Bernal JL, Andrews N, Gower C, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. *British Medical Journal* 2021; **373**.
29. Riley S, Atchison C, Ashby D, et al. REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: study protocol. *Wellcome Open Research* 2020; **5**.
30. Clarke J, Beaney T, Majeed A. UK scales back routine covid-19 surveillance. *British Medical Journal* 2022; **376**.