

1 **Forecasting the spatial spread of an Ebola epidemic in real-time:**  
2 **comparing predictions of mathematical models and experts**

3

4 James D Munday\*<sup>†1,2,3</sup>, Alicia Rosello<sup>†1,2</sup>, W. John Edmunds<sup>§1,2</sup>, Sebastian Funk<sup>§1,2</sup>

5

6 <sup>1</sup> Centre for mathematical modelling of infectious diseases, London School of Hygiene and  
7 Tropical Medicine, United Kingdom

8 <sup>2</sup> Department of infectious disease epidemiology, London School of Hygiene and Tropical  
9 Medicine, United Kingdom

10 <sup>3</sup> Department of Biosystems Science and Engineering, ETH Zürich, Switzerland

11

12 † These authors contributed equally

13 § These authors contributed equally

14

15 \*Corresponding author: [james.munday@bsse.ethz.ch](mailto:james.munday@bsse.ethz.ch)

16

17 **Keywords:** Ebola, Forecast evaluation, Expert elicitation, Mathematical modelling, Outbreak  
18 response

19 **Abstract**

20 Ebola virus disease outbreaks can often be controlled, but require rapid response efforts  
21 frequently with profound operational complexities. Mathematical models can be used to  
22 support response planning, but it is unclear if models improve the prior understanding of  
23 experts.

24  
25 We performed repeated surveys of Ebola response experts during an outbreak. From each  
26 expert we elicited the probability of cases exceeding four thresholds between two and 20 cases  
27 in a set of small geographical areas in the following calendar month. We compared the  
28 predictive performance of these forecasts to those of two mathematical models with different  
29 spatial interaction components.

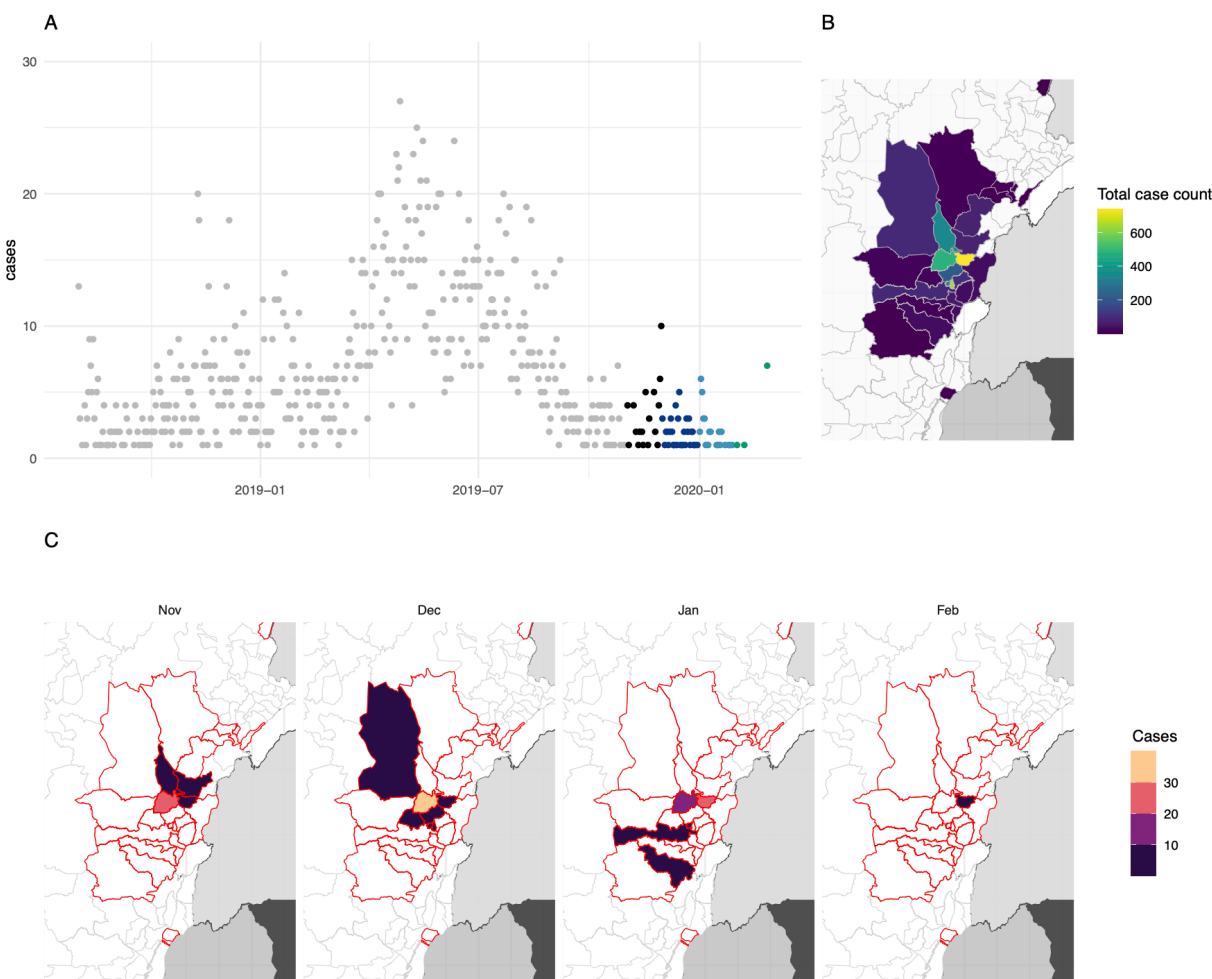
30  
31 An ensemble combining the forecasts of all experts performed similarly to the two models.  
32 Experts showed stronger bias than models forecasting two-case threshold exceedance. Experts  
33 and models both performed better when predicting exceedance of higher thresholds. The  
34 models also tended to be better at risk-ranking areas than experts.

35  
36 Our results support the use of models in outbreak contexts, offering a convenient and scalable  
37 route to a quantified situational awareness, which can provide confidence in or to call into  
38 question existing advice of experts. There could be value in combining expert opinion and  
39 modelled forecasts to support the response to future outbreaks.

40

41 **Background**

42 Following the initial emergence in 1976 in Zaire (now the Democratic Republic of the Congo,  
43 DRC)[1], epidemics of Ebola Virus Disease (EVD) have occurred, on average, every 12 - 24  
44 months[2]. EVD is a viral haemorrhagic fever first caused by the Ebola Zaire virus (EZV), with a  
45 case fatality rate of 25-90%[3]. A major outbreak in North-Eastern provinces of DRC between  
46 2018-2020 resulted in over 3300 reported cases and over 2100 deaths [4](Figure 1).  
47



48 **Figure 1. The extent of the 2018-2020 ebola outbreak in north-eastern DRC and areas included in our study.** A)  
49 Daily incidence in north-eastern DRC between August 2018 and March 2020. Grey points show days prior to the  
50 study period, coloured points show days within the study period (November 2019 - March 2020), hue indicates  
51 month. B) Shows the total number of cases of ebola recorded in each Health Zone. C) Number of cases in each  
52 month and health zone during the period covered by this study, health zones outlined in red show all health zones  
53 affected by the entire epidemic.  
54

55  
56 Transmission of EZV occurs mainly through direct contact during the symptomatic phase of  
57 infection; therefore, isolation of infected individuals with strict infection control, contact  
58 tracing, and safe burials have been key to controlling past EVD outbreaks[5], although setting

59 specific challenges can hamper containment efforts [6]. More recently, vaccination has also  
60 become a tool for outbreak control, with two vaccines now licensed for use[7][8].

61  
62 EVD outbreaks typically occur in resource-poor settings where limited communication and poor  
63 accessibility make logistics of surveillance and vaccination campaigns challenging.  
64 Understanding the spatial risk of future spread is therefore useful to allow response teams to  
65 focus efforts on high-risk areas. Mathematical and statistical models have been used  
66 extensively to forecast the spread of infectious diseases, including EVD[9]. Such models rely on  
67 a combination of statistical inference based on epidemiological data and information about the  
68 mechanisms underlying the dynamics of infection. However, the dynamics of EVD are  
69 frequently governed by changing contextual factors which are challenging to forecast  
70 quantitatively. For example, violent conflicts or flooding can seriously hinder, interrupt, or even  
71 reverse the impact of containment efforts [6,10]. Moreover, changes in healthcare capacity  
72 and health seeking behaviour of patients can strengthen or weaken efforts to reduce  
73 transmission [11]. The timing and impact of these factors is notoriously difficult to predict using  
74 mathematical models.

75  
76 Models are used by epidemic response experts to support decision making in the field. In  
77 addition to models, experts also make judgments as to the future spread of the virus based on  
78 their interpretation of the current status of the outbreak combined with their knowledge of  
79 other less tangible factors such as the geography, climate (eg. seasonal variation in accessibility  
80 of particular areas) and soft intelligence about the escalation of conflict in areas which may as a  
81 result, be harder to access by response teams. There are clear costs and benefits to human-  
82 made and modelled-based forecasts. Whereas models are objectively based on observations of  
83 the past outbreak dynamics and current case data, experts have additional knowledge of the  
84 complex factors surrounding the outbreak response. It is therefore difficult to assess the impact  
85 mathematical models have on decision making, how much modelled forecasts differ from those  
86 made by experts in the field and whether either modelled or human forecasts are  
87 systematically more accurate or useful. Moreover, the knowledge of experts in the field of EVD  
88 epidemiology, with a good understanding of the geographical area of study may provide an  
89 invaluable resource that is currently underused in forecasting.

90  
91 Previous studies have aimed to establish the relative performance of humans and models in  
92 predicting infectious disease spread in human populations, particularly in the context of acute  
93 respiratory infections such as Influenza and SARS-CoV-2. Three studies have evaluated the  
94 predictions of humans against models, explicitly. The first of these evaluated short-term  
95 forecasts and season-wide predictions of reports of influenza-like-illness (ILI) in the United  
96 States of America (USA) [12] and two studies [13,14] compared short-term forecasts of cases of  
97 and deaths from COVID-19, firstly in Germany and Poland and secondly in the United Kingdom  
98 (UK). All three studies found that humans tended to perform better than the mathematical and  
99 statistical models selected for comparison when predicting cases. However, the COVID-  
100 focussed studies found the human ensembles performed worse than the ensemble prediction  
101 of the models when predicting deaths - These results were maintained when only self-declared  
102 'experts' were included in the forecasts. A number of other studies recorded expert predictions

103 without comparison to mathematical models. A study conducted early in the COVID-19  
104 pandemic [15] evaluated the relative ability of laypeople and experts to predict the course of  
105 the UK epidemic over the first calendar year. The study found that both experts and laypeople  
106 typically under-predicted the impact overall, however experts' forecasts were more accurate  
107 and better calibrated than laypeople. A study of expert predictions in the United States of  
108 America [16] evaluated their weekly forecasts of case incidence and total deaths in the first  
109 year against a pooled ensemble of all predictions. The study found that the ensemble  
110 outperformed every expert individually over the period of the study. A similar study surveyed  
111 experts regarding the total number of cases and deaths from MPox in the USA during 2022 [17],  
112 however these predictions are yet to be evaluated. Overall, these studies provide evidence that  
113 human predictions can play a valuable role in epidemiological prediction, providing a  
114 comparator and complementary method to mathematical and statistical modelling.

115  
116 In this paper we extend the use of expert forecasters to predict spatial risk of transmission in  
117 the context of a local outbreak. We made monthly forecasts of the geographic spread of Ebola  
118 Virus Disease (EVD) from November 2019 to March 2020 during the declining phase of the  
119 2018-2020 outbreak in the Democratic Republic of the Congo (DRC) using both expert  
120 predictions collected through regular interviews and with two spatially explicit computational  
121 transmission models with different spatial interaction assumptions: a gravity model and an  
122 adjacency model (where transmission can only occur between contiguous regions), see the  
123 methods section for details. Alongside supporting situational awareness, these forecasts were  
124 motivated by an aim to inform site selection for a planned vaccine trial. The objective was to  
125 identify areas that had seen no cases yet and thus were not already being supported by  
126 vaccination and other interventions, but were at high risk of still becoming affected by the EVD  
127 outbreak, thus allowing estimation of efficacy [18]. Here we evaluate the performance of the  
128 forecasts and select ensembles of the methods in predicting continued transmission and flare-  
129 ups of EVD in health zones (HZs) close to the affected area. We further study variation in  
130 forecast quality against a selection of factors related to local demography, case history and  
131 forecast implementation.

132 **Methods**

133

134 *Expert elicitation*

135 Experts in EVD epidemiology with knowledge of the local geography speaking English or French  
136 were identified originally by convenience sampling. The pool of experts was then expanded  
137 through recommendation from the identified experts (snowball sampling). This approach was  
138 best suited to capture the expertise of individuals who were most often temporarily based in  
139 the field.

140

141 A pilot study was carried out in November 2019. Subsequently, monthly interviews were held  
142 over WhatsApp in December 2019, January 2020, February 2020, and March 2020. All  
143 interviews were scripted. The main biases of this type of study (availability bias,  
144 representativeness bias, overconfidence, motivational bias, anchoring on past estimates) were  
145 briefly discussed during the first interview.

146

147 Experts also were provided with an interactive map of the outbreak area and surrounding  
148 health zones (HZ) showing the number of total cases during the outbreak and during the two  
149 preceding weeks for reference (supplementary figure S1). HZ were numbered to facilitate  
150 communication with the experts.

151

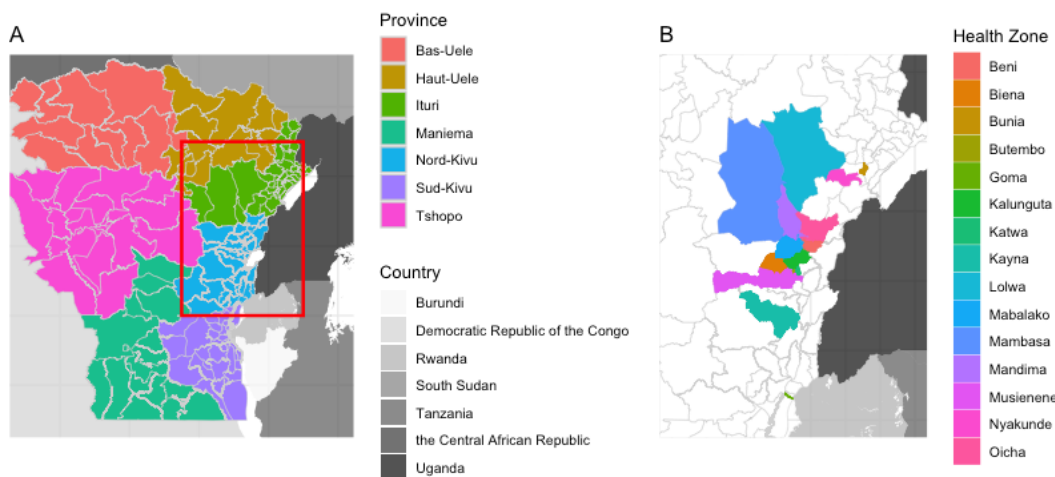
152 Experts were asked to estimate the number of reported probable and confirmed cases they  
153 would expect per HZ during the following month using the online MATCH Uncertainty  
154 Elicitation Tool [19] (supplementary figure S1). Through this platform, the experts and the  
155 researcher (AR) interacted in real time. The “roulette” (chips and bins) method was used.  
156 Experts were instructed to place a total of 20 chips over the available bins (0-1 cases, 2-4 cases,  
157 ..., 48-50 cases). Therefore, each chip represented for the expert a 5% probability that the  
158 number of cases was in the bin where the chip was placed. This process aimed to capture the  
159 uncertainty surrounding the expert’s estimates.

160

161 The experts were asked to estimate the number of reported cases they would expect in the HZ  
162 where there had been 1 or more cases in the 2 preceding weeks, as well as Goma (Figure 2).  
163 The experts were then asked to identify any additional HZ where they would predict 1 or more  
164 cases during the following month with >5% probability, and to also estimate the number of  
165 reported cases they would expect in these HZ. In the pilot study, carried out in October 2019,  
166 experts were asked to forecast the number of cases they expected during November 2019 in 10  
167 HZs: Beni, Goma, Kalunguta, Katwa, Lolwa, Mabalako, Mambasa, Mandima, Nyankunde, and  
168 Oicha.

169

170



171  
172 **Figure 2. Health zones included in the model and Expert Elicitation survey.** A shows the provinces around the  
173 affected area, and included in the transmission model, the red box shows the area detailed in panel B. B shows the  
174 health zones included at least once in the Expert Elicitation survey we conducted.

175  
176 *Ethics*

177  
178 LSHTM ethics approval was obtained for this study (reference: 17633). Signed informed consent  
179 was taken from experts willing to participate and their verbal consent was requested again at  
180 the beginning of each elicitation.

181  
182 *Modelling framework*

183  
184 In parallel with the expert elicitation programme, we developed a modelling framework to  
185 forecast spatial risk of infection. In the framework, incidence of cases is forecast in each Health  
186 Zone based on historical case reports. The model was formed of two components, the  
187 autoregressive component, and the spatial component.

$$\lambda_{it} = \sum_{s=t-(D+L)}^{t-D} (\gamma N_{i,s} + \alpha \sum_j w_{ij} N_{j,s}) \quad (1)$$

189  
190  
191 The auto-regressive component modelled the rate of infections in a particular health zone i, on  
192 day t, to be proportional to the number of cases in the same health zone (i) between dates t-  
193 (D+L) and t-D, where L is the estimated latent period and D is the estimated infectious period.  
194 The spatial component accounts for transmission between health zones, where rate of  
195 infection was proportional to the cases in each other Health Zone (i.e.  $\forall j \neq i$ ) and moderated by  
196 a pairwise specific factor defined by a spatial kernel  $w_{ij}$  We used two spatial kernels, both of

197 which use proximity of health zones to each other and their respective population size,  $P_i$  and  
198  $P_j$ . Firstly, the gravity model which treats interaction in an analogous way to Newtonian gravity  
199 with population size in place of mass, such that interaction reduces distance,  $d$ , raised to a  
200 power,  $k$ .

$$201 \quad w_{ij} = \frac{P_i P_j}{d_{ij}^k} \quad (2)$$

202  
203  
204 Secondly, we applied a model with adjacency-based interaction. In this model only adjacent HZs  
205 can interact. The strength of interaction between HZs is proportional to the product of their  
206 population sizes.

$$207 \quad w_{ij} = \delta_{ij} P_i P_j \quad (3)$$

$$208 \quad \delta_{ij} = \begin{cases} 1, & \text{if adjacent} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

209  
210  
211  
212 Cases were modelled as Poisson distributed such that:

$$213 \quad N_{i,t} = \text{Poisson}(\lambda_{i,t}) \quad (5)$$

214  
215  
216 To forecast cases, we fitted the spatiotemporal model to historical data from the 60 days prior  
217 to the date the forecast was made, accounting for cases in health zones in seven regions (169  
218 HZs) centred on the location of the epidemic; Nord-Kivu, Ituri, Tshopo, Maniema, Sud-Kivu,  
219 Haut-Uele, Bas-Uele. We fit the model using the No U-Turn Sampling (NUTS) method for  
220 Hamiltonian Monte Carlo with Stan [20], a probabilistic programming framework. We  
221 estimated  $\alpha$  and  $\gamma$ , which vary the contribution of within-health-zone and between-health-  
222 zone transmission. We also estimated  $k$ , which determines how rapidly transmission rate  
223 decays with distance in the spatial component of the model. We sampled parameters from the  
224 resultant joint posterior distribution to simulate daily incidence in all HZs in the seven regions,  
225 up to and including the last day of the following month. We performed 1000 iterations for each  
226 forecast date. We then extracted the full distribution of the number of cases incident within the  
227 calendar month of interest. Forecasts were made using data up to the last day of the month  
228 prior to the forecast period.

229  
230  
231 *Quantification of risk and forecast evaluation*

232 To compare the model and the expert forecasts and score them according to the eventual true  
233 number of cases we calculated the probability attributed to cases over four thresholds,  $\geq 2$ ,  
234  $\geq 6$ ,  $\geq 10$  and  $\geq 20$  cases.

235  
236 We evaluated the forecasts using the Brier Score, a proper scoring rule which quantifies how  
237 accurate a forecast or a group of forecasts are when compared to true data after the event. The



238 Brier score, BS, is defined as the square of the difference between the probability of observing  
239 an event and the observation  $o_i$  status, which takes a value 1 or 0 for cases observed and none  
240 observed respectively. We calculated this for multiple (N) forecasts by taking the mean of the  
241 individual forecast scores.  
242

$$BS = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \quad (6)$$

243  
244 We also quantified the general bias and calibration of the forecasts by considering the hazard  
245 rate predicted by each forecast, which we calculated as the sum of probabilities attributed to  
246 exceeding each threshold. This gives the number of HZs the forecast 'expected' to cross the  
247 threshold in each month. To quantify the bias of each set of predictions, we took the difference  
248 between the hazard rate and the actual number of HZs that exceeded each threshold in each  
249 month. We refer to this as the *hazard gap* (HG).  
250

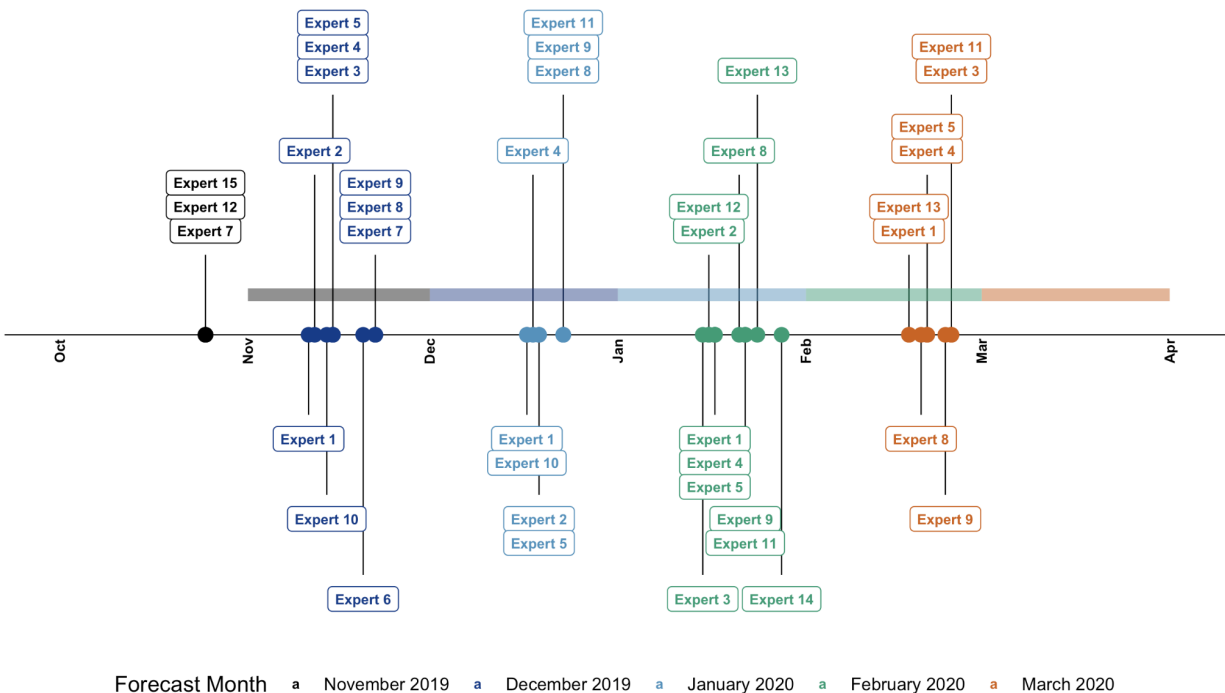
$$HG = \sum_{\forall HZ} p(c > c_{thresh}) - \sum_{\forall HZ} o(c > c_{thresh}) \quad (7)$$

251  
252  
253  
254  
255  
256

257 **Results**

258 *Expert panel and health zones included in survey*

259 Over the study period, we conducted a total of 40 interviews with 15 experts, three of which  
 260 took place during the pilot phase (November 2019). Figure 3 shows the timeline of the expert  
 261 elicitations.  
 262



263  
 264 **Figure 3. Timeline of the expert elicitation.** Each point shows the date of the interview of the expert labelled to  
 265 obtain forecasts for the following month. Colour indicates the month for which the forecast was made, the  
 266 forecast windows are highlighted with a shaded band of the same colour.  
 267

268 Eight experts worked at the World Health Organization, four at the London School of Hygiene &  
 269 Tropical Medicine, two for Médecins Sans Frontières, and one at the DRC Ministry of Health.  
 270 Most experts (10/15) had more than five years of experience working in infectious disease  
 271 epidemiology. About half of the interviews (21 of 40) were conducted with experts that were in  
 272 the outbreak area (defined HZs affected by EVD or Goma, the site of the international response  
 273 base) or had been there within 2 weeks of the interview. Four experts had never been in the  
 274 outbreak area.

275  
 276  
 277  
 278  
 279  
 280  
 281  
 282

**Table 1: Experts and health zones included in each round of the survey**

Month forecasted	Experts interviewed (Highlighted)			Health zones in interview (HZs)	No. experts	HZ nominated
<b>December</b>	1	2	3	Beni	7	Oicha
	4	5	6	Goma	5	Komanda
	7	8	9	Kalunguta	2	Butembo
	10	11	12	Mabalako	2	Katwa
	13	14		Mambasa	1	Lolwa
				Mandima	1	Makiso - Kisangani
					1	Nyankunde
<b>January</b>	1	2	3	Beni	6	Butembo
	4	5	6	Biena	3	Katwa
	7	8	9	Goma	1	Kalunguta
	10	11	12	Mabalako	1	Manguerdjipa
	13	14		Mandima		
				Oicha		
<b>February</b>	1	2	3	Beni	6	Oicha
	4	5	6	Bunia	4	Biena
	7	8	9	Butembo	2	Vuhovi
	10	11	12	Goma	1	Lolwa
	13	14		Kalunguta		
				Katwa		
				Kayna		
				Mabalako		
				Mambasa		
				Mandima		
			Musienene			
<b>March</b>	1	2	3	Beni	3	Mabalako
	4	5	6	Butembo		
	7	8	9	Goma		
	10	11	12	Mandima		
	13	14				

283  
284

285 Eight to ten experts were interviewed each month between December and March where they  
286 were asked to forecast cases in between four and eleven health zones (supplementary figures  
287 S2 - S5).

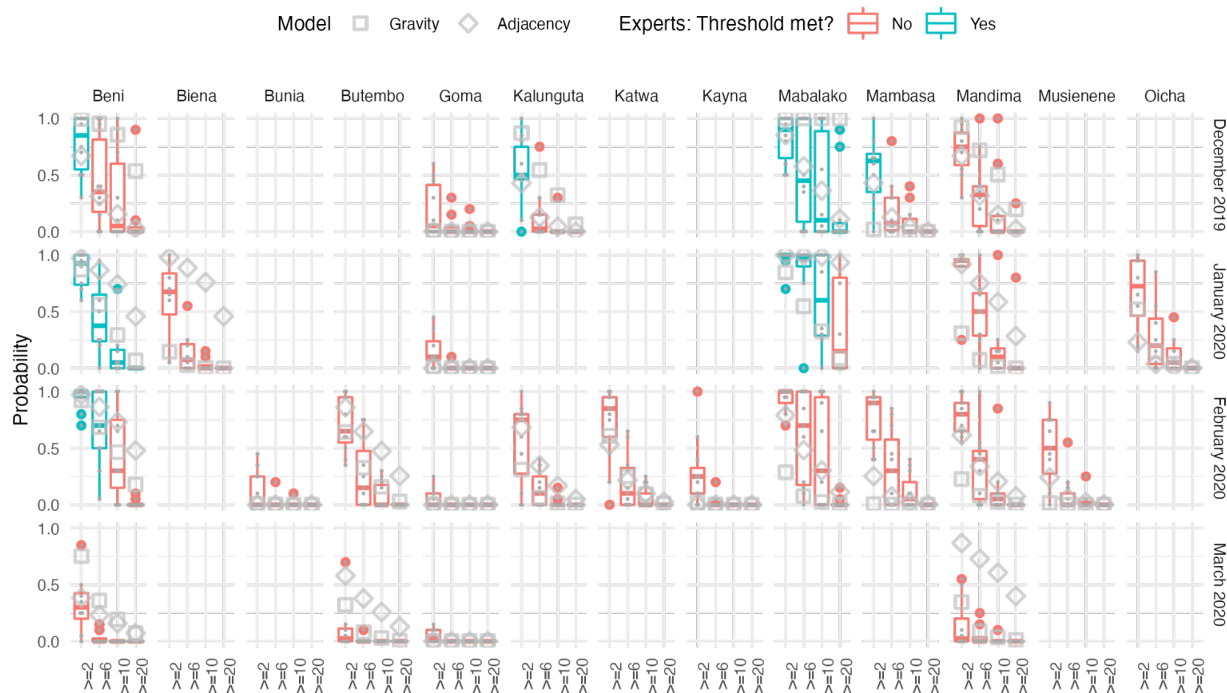
288  
289 In December there were four health zones that reported 2 or more cases, Beni (3), Kalunguta  
290 (5) and Mambasa (4) did not reach the 6 case threshold. Mabalako reported 38 cases, therefore  
291 exceeding all of the thresholds (Figure 4). All experts attributed a greater than 50% chance of  
292 Mabalako exceeding 2 cases in December with a mean of 82%. The experts were collectively

293 less confident of it exceeding the higher thresholds, with a number of experts attributing no  
294 probability of crossing these thresholds at all. The experts ranked Beni, which reported 3 cases,  
295 as the second most likely to pass all of the thresholds. They also attributed Mambasa and  
296 Kananguta, which reported 4 and 5 cases respectively, with lower probability (53% and 54%)  
297 than Mandima (72%), where no cases were confirmed in December.

298  
299 The experts correctly identified Beni and Mabalako as high risk HZs in January, where there  
300 were 22 and 11 cases reported respectively. However, they generally expected more cases in  
301 Mabalako than Beni, and in fact attributed 0% probability of exceeding 20 cases in Beni (where  
302 the threshold was exceeded) but 11% chance in Mabalako, where the threshold had been  
303 exceeded the previous month - but was not exceeded this month. Similarly to December the  
304 experts expected to see high numbers of cases in Mandima, where no cases were reported in  
305 January. There was near unanimity amongst the experts that cases would arise here, with all  
306 but one expert attributing a probability of at least 90% that 2 or more cases would be reported.  
307 The experts also collectively predicted a probability of at least two cases in Oicha in January but  
308 were more cautious with a mean probability of 63% for crossing the 2 case threshold. Similarly,  
309 for Biena experts gave a mean probability of 60% for exceeding the 2 case threshold. In both  
310 HZs no cases were confirmed in January.

311  
312 In February, of the eleven HZs nominated, only Beni reported confirmed cases. Here, 9 cases  
313 were confirmed in total, meaning the 6 case threshold was passed. Experts collectively assigned  
314 a probability of 70% for this. Seven and five of the ten experts were certain that the 2 case and  
315 6 case thresholds would be crossed in Beni respectively. Collectively, the experts expected  
316 similar case numbers in Mabalako attributing 60% probability of exceeding the 6 case  
317 threshold, however no cases were reported here in February. Notably, cases were reported  
318 here in December and January. Experts also assigned a probability of over 50% for crossing the  
319 2 case threshold for Butembo, Kalunguta, Katwa, Mombasa, Mandima and Musienene. None of  
320 which reported confirmed cases in February. Experts, however, considered Bunia, Kayna and  
321 Goma to be at low risk of having 2 or more cases, with the exception of one expert—who  
322 attributed a probability of 100% to the 2 case threshold in Kayna.

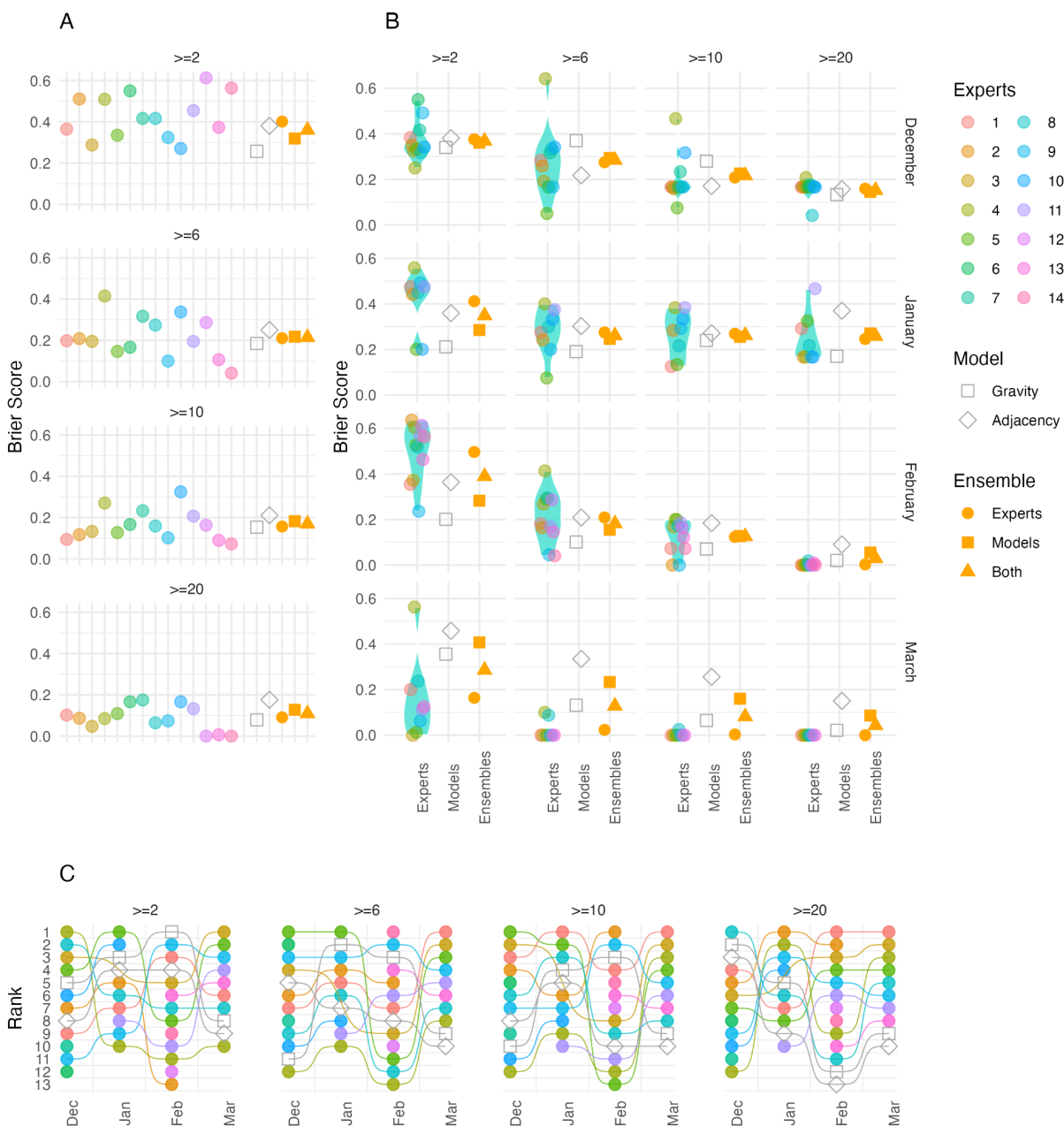
323  
324 In March no cases were reported in any HZ. Experts broadly predicted this well with only one  
325 expert assigning a probability of greater than 50% for exceedance of the 2 case threshold. The  
326 HZ with the highest average assigned probability was Beni with a mean probability of 33% of  
327 exceeding the 2 case threshold.



328  
329 **Figure 4. Expert elicitation results and accuracy of predictions.** Only the HZs that were rated by all experts are  
330 included here. Results are shown as probabilities (vertical axes) that a given health zone (horizontal panels)  
331 exceeds a given threshold (horizontal axes) according to the experts (box plots) or models (square / diamond for  
332 gravity and adjacency models, respectively) across different months (vertical panels). Health zone / month  
333 combinations where the given thresholds were exceeded are marked in cyan, and ones where they weren't in red.  
334

### 335 *Performance evaluation*

336 We evaluated the forecasts using the Brier score. The overall scores of individual experts varied  
337 between 0 and 0.6 across the four thresholds. Collectively, the experts scored best at the  
338 highest threshold (20 cases) and worst for forecasts of the lowest threshold (2 cases). The  
339 models also performed better at higher thresholds than low thresholds, but the difference was  
340 less pronounced. Overall, the gravity model ranked best amongst all forecasts at the 2 case  
341 threshold. It also ranked best for this threshold in the month of February and consistently in the  
342 top half of forecasts in December and January, however performed comparatively poorly in  
343 March, ranking higher than only one of the experts. The adjacency model also performed better  
344 than the experts overall for the 2 case threshold. Related to this, including the models  
345 improved the ensemble forecast. Although the gravity model performed better than the  
346 adjacency model for higher thresholds, together the models performed similarly to the expert  
347 ensemble forecast overall. In January and February the gravity model performed well compared  
348 to the adjacency model and the expert ensemble, however in March both models performed  
349 particularly badly compared to the experts for all thresholds. None of the experts performed  
350 consistently well relative to the others, experts 3 and 10 performed best for the 2 case  
351 threshold, whereas experts 13 and 14 did best for higher thresholds.  
352



353  
354  
355  
356  
357  
358  
359  
360  
361  
362

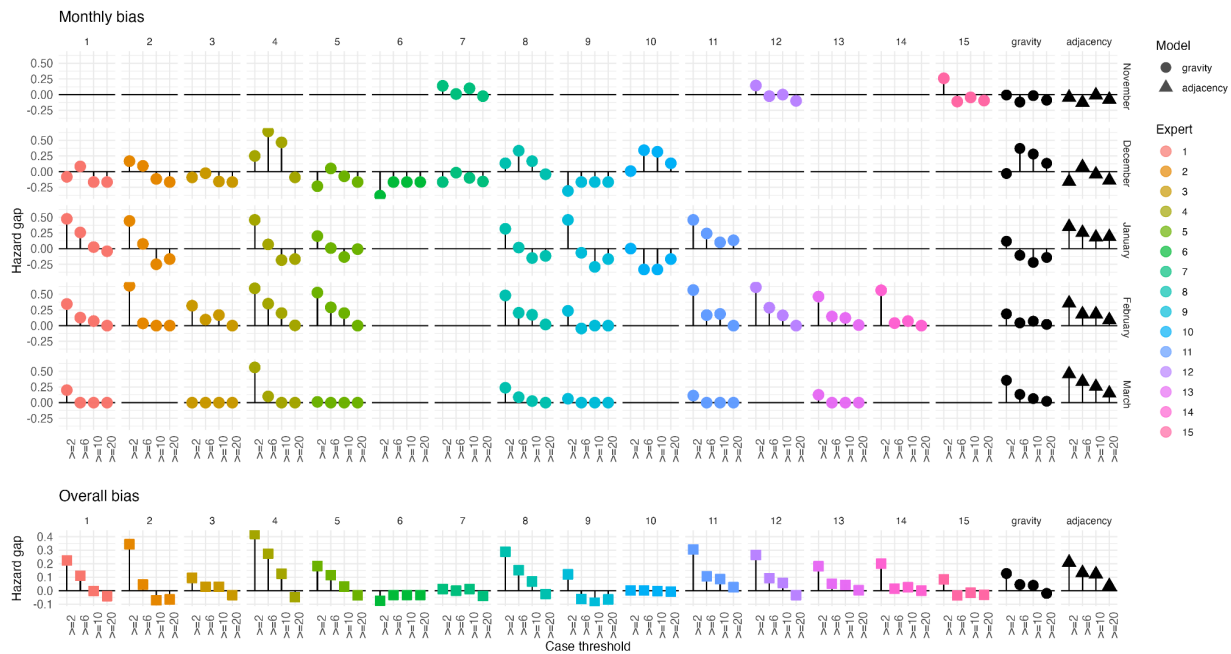
**Figure 5. Evaluation of forecasts made by the experts, models, and ensembles.** A shows the overall Brier Score for each expert, model and ensemble (calculated over all forecasts included in the study). In B each panel shows the Brier score across all health zones for each month (vertical) at each case threshold (Horizontal). Coloured points show each expert score, the violin plot shows their distribution. The grey hollow points show the model scores, the yellow points show the ensemble scores (circles show experts alone, squares show models alone and triangles show experts and models with 50% weight given to each). C Shows the ranking of each expert and model in terms of forecast performance

363 To evaluate how the different forecasts may impact decision making we ranked the health  
364 zones for each month, based on the probability of exceeding each threshold of cases forecast  
365 by each ensemble and by the model alone (supplementary figure S6). In general, the model and

366 the ensembles all ranked health zones that did reach the threshold highly. In some cases the  
 367 model performed better, ranking health zones that did meet the threshold higher than the  
 368 experts, specifically ranking Beni Higher than Mandima in higher thresholds ( $\geq 6$  and 10 cases)  
 369 for the forecast of January, where Beni ultimately had cases and Mandima did not, in that  
 370 month. Considering the models separately, the gravity model performed better than the  
 371 adjacency model in general, with the adjacency model occasionally performing worse than the  
 372 experts when ranking the HZs. This was clearest in the forecasts of November and January.

373  
 374 *Bias and calibration in forecasts*

375 We evaluated the bias in each forecast type by considering the hazard gap between forecasts  
 376 and actual cases. We found that experts systematically forecasted higher risk of the lowest  
 377 threshold ( $\geq 2$  cases) than was warranted, but tended to forecast lower risk of exceeding the  
 378 highest threshold ( $\geq 20$  cases) than was borne out across all HZs (Figure 6). When calculated  
 379 across all months, this bias was present in 12 of the 15 experts. The models did not show clear  
 380 consistent bias in either direction.  
 381



382  
 383 **Figure 6 Bias and calibration of forecasts.** Panels show the Hazards gap difference between the Hazard rate  
 384 (expected number of exceedances across all health zones) for each threshold attributed by the forecast and the  
 385 actual number of Health Zones that exceeded the associated threshold. Each panel shows one forecast (expert or  
 386 model) in each month. The bottom row shows the same for each forecast calculated over the entire study period.

387  
 388 *Forecasting flare-ups*

389 In addition to the health zones presented to all experts, each expert was able to nominate  
 390 health zones, which they deemed at risk. Experts nominated seven further HZs to forecast in  
 391 December, four in January, four in February and one in March (Table 1).

392



393 Of the seven HZs nominated in December, all except Lolwa had at least one expert attributing  
394 some probability of more than 2 cases. Oicha and Komanda had the most nominations with  
395 seven and five each, and six and four of the ten experts interviewed allotted greater than 5%  
396 chance of 2 or more cases in December. For Oicha, attributed probabilities of observing at least  
397 2 cases ranged between 5% and 95%, for Komanda probabilities of between 30% and 80% were  
398 attributed to crossing the threshold. Neither HZ passed the threshold in November. Experts 4  
399 and 8 both nominated Katwa (both gave 50% chance of 2 or more cases) and Butembo  
400 (attributed probabilities of 45-50%) - the only health zones not included in the survey for  
401 December to have greater than 2 cases. Makiso-Kisangani and Nyankunde) were nominated by  
402 one expert each (expert 4 with 50% and 10 with 35% respectively). Lolwa was nominated by  
403 expert 3, but attributed all probability to less than 2 cases.

404  
405 In January all nominations of health zones were accompanied by an attribution at least 5%  
406 probability of exceeding the 2 case threshold. Six of the eight experts interviewed nominated  
407 Butembo (1,2,4,5,9, and 11 with probabilities of 5% to 85% of crossing the 2 case threshold)  
408 three of them also nominated Katwa (2 and 9 giving 85% and 5 giving 5%) and Kalunguta and  
409 Manguredjipa were also nominated by one expert each, 5 with 10% and 11 with 50%  
410 respectively. None of the nominated health zones crossed the threshold in January.

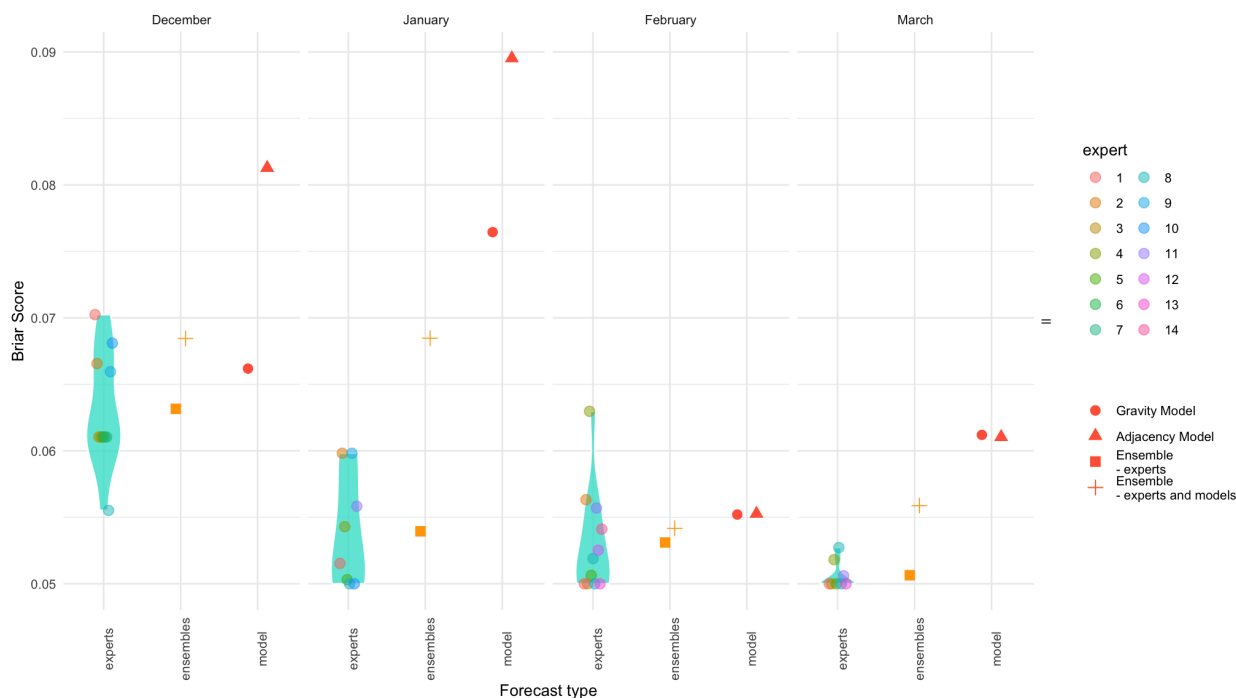
411  
412 In February six of the ten experts interviewed (2, 4, 5, 11, 12 and 14) nominated Oicha with  
413 probabilities between 10% and 95% of crossing the 2 case threshold. Four (2, 4, 5 and 11)  
414 nominated Biena with probabilities between 10% and 95% of exceedance. Experts 4 and 8 also  
415 nominated Vuhovi with attributing 55% and 20% probability of threshold exceedance  
416 respectively. Expert 3 nominated Lolwa alone but gave no probability of exceeding 1 case. No  
417 HZs not included in the interview as default crossed the 2 case threshold in February.

418  
419 In March three (4, 8 and 11) of the eight experts interviewed gave probabilities of 35%, 50%  
420 and 15% of exceeding the 2-case threshold respectively. No HZs not included in the interview as  
421 default crossed the 2 case threshold in March.

422  
423 To compare the model with the experts we included all HZs modelled and attributed all HZs not  
424 nominated by experts an exceedance probability of 0%. To allow comparison, we also set all  
425 HZs given a probability of lower than 5% to 0% for both the gravity and adjacency models.  
426 When considering the Brier Score (Figure 7), we found that the gravity model performed  
427 comparably to some experts when forecasting for December, and February. The adjacency  
428 model performed worse than all the experts in every month except February. In every month  
429 the ensemble of experts did better than the models and including the models in the ensembles  
430 reduced performance.

431  
432  
433  
434





435  
436 **Figure 7 Evaluation of forecasts made in health zones not included in the main survey.** Each panel (right to left)  
437 shows the Brier score across all health zones for each month. Coloured points show each expert score, density plot  
438 shows their overall distribution. The red points show the model scores, the yellow points show the ensemble  
439 scores (squares show experts alone, crosses show experts and models with 50% weight given to each).  
440

## 441 Discussion

442 We compared forecasts of the geographic spread of Ebola made by experts, with those made  
443 using a modelling framework. Since the outbreak dynamics of Ebola are highly sensitive to the  
444 changeable context in which they take place, mathematical models and expert opinions are  
445 expected to have different strengths and weaknesses, with models benefiting from objective  
446 inference from previous observations and experts able to utilise detailed knowledge about the  
447 outbreak and the changing surrounding context to make informed projections of risk. By  
448 interviewing experts and asking them to forecast risk in a structured way, we were able to  
449 compare the performance of their forecasts against those made with well-established  
450 modelling approaches in a quantifiable and robust way.  
451

452 Overall, the forecasts made by the group of experts as a whole performed similarly to those of  
453 the model, with a few consistent exceptions. The model performed better than the experts  
454 when considering the lowest threshold in four of the five months covered by the survey, but  
455 performance was more comparable for the higher thresholds. The model also performed  
456 marginally better when ranking health zones by risk of ongoing transmission, indicating that use  
457 of a model may improve prioritisation of health zones when attributing resources.  
458

459 We found that both methods performed better when considering higher case thresholds. This is  
460 likely to be due to a combination of bias in both forecast types towards under prediction of  
461 cases and the fact that there were few instances where the higher thresholds were reached.

462  
463 Although individual experts frequently out-performed ensembles in individual instances, no  
464 individual expert outperformed the ensembles overall. This supports the practice of considering  
465 predictions from a range of experts over a smaller number of more specialist or experienced  
466 experts. The models tended to perform similarly to the ensemble representing more consistent  
467 performance across all forecasts.

468  
469 Experts tended to be more biased than models, especially at low case thresholds with a  
470 tendency to over-predict cases to a greater degree than models. This bias reduced rapidly as  
471 the case threshold increased. This may be interpreted as over cautiousness from the experts  
472 regarding potential for geographic spread of the virus but confidence that transmission could  
473 be contained quickly. This trend reflects a pattern amongst previous introductions into new  
474 health zones earlier in the pandemic, where a small number of cases were reported, but the  
475 local outbreak was quickly stopped (Figure 1).

476  
477 To our knowledge, our study is the first to record experts' assessment of geographical risk at a  
478 local level during an epidemic and the first comparison of outbreak response experts'  
479 predictions to those of models in real-time. Although direct comparison is not possible, our  
480 results lead to conclusions that are broadly similar to those from previous studies [12–14],  
481 however each of these studies found that ensemble expert forecasts performed better than the  
482 comparison models, whereas our study found no clear performance difference. This may  
483 suggest that experts are better at predicting simple time series than geographic distribution of  
484 cases. However, we cannot view these findings independently from the different survey designs  
485 or study contexts.

486  
487 There are a number of important limitations to consider when interpreting our results. The  
488 context within which we conducted the study has important implications for interpretation.  
489 Due to the timing and logistics of setting up the questionnaire, the study only began in the  
490 closing phase of the epidemic, whereas the relative performance of experts and models may  
491 differ during different phases of the epidemic. For example, in the early phase where dynamics  
492 are driven more by infectious transmission than established response practices, or during the  
493 peak where changes in intervention strategy may be more influential. The stage of the epidemic  
494 also meant that there was a substantial trend towards 'negative' results (i.e. no threshold  
495 exceedance), which is likely to favour some forecasting methods over others.

496  
497 Additionally, experts were not all interviewed on the same day and interviews occurred several  
498 days before the beginning of the month they were forecasting. In some cases experts were  
499 interviewed up to 2 weeks prior to the beginning of the month. This means that the information  
500 available differed both between experts and with the model, which was run considering data  
501 up to the first day of each month. This reduces our ability to compare models to experts  
502 directly, however, it could also be argued that this is a 'built-in' factor which represents the  
503 inherent challenge of eliciting predictions from experts. In addition, the interview process for  
504 experts was quite taxing and required a phone call - which can cause scheduling challenges

505 during an epidemic response. It may be that other methods, with less arduous and more  
506 flexible data entry would improve responses.

507  
508 Our analysis represents the comparison of expert forecasts to only two specific forecasting  
509 models. There are a great range of models that could have been applied in this context which  
510 may have differed in performance to those we used. We chose these models for convenience  
511 since we were applying them to the outbreak at the time of the interviews. It is also possible  
512 that some of the experts involved in the study had ingested results from our model, which were  
513 available through our online dashboard, or other models being used at the time.

514  
515 Since our findings, like those of similar studies, suggest that models and experts perform  
516 comparably in this context, there is an argument that models have no value in informing expert  
517 decision making. It can be argued, however, that models remain useful in outbreak response.  
518 Firstly, while the models performed similarly to the ensemble forecasts of the experts, there  
519 was no individual expert that performed consistently better than the models. Secondly, models  
520 are much more easily scaled and generalised making them simple to deploy in new contexts  
521 and to adapt as epidemics grow. Expert interviews are time-consuming and often inconvenient,  
522 especially in the context of outbreak response activities, which are characteristically fast paced.  
523 Models therefore offer a more convenient route to a quantified insight, which from our results,  
524 performs comparably to the way groups of experts may think. Finally, there are ways to  
525 combine both methods. For example, in the event that expert forecasts can be garnered, joint  
526 ensembles can capture information from both the expert and modelled forecasts. Further, we  
527 suggest that models can offer a role in aiding decision making by providing confidence in or  
528 calling into question expert advice that is being considered.

## 529 530 **Conclusions**

531 Our analysis evaluated performance of experts and models when forecasting the spatial spread  
532 of Ebola, representing the first such study incorporating local geographic distribution and the  
533 first to focus on an epidemic in a resource poor setting. We found that forecasts made by  
534 experts and models performed comparably overall, but experts tended to be slightly more  
535 biased towards predicting that a small number of cases would persist. The results support the  
536 use of models in outbreak response and provide insight into how models and expert opinion  
537 could be combined when tackling future epidemics.

538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574

### **Acknowledgements**

We would like to thank Xavier de Radiguès, Neale Batra, Nabil Tabbal, Mathias Mossoko, Chris Jarvis, Thibaut Jombart, Denis Ardiét, Michel Van Herp, Silimane Ngoma, Olivier le Polain, Esther van Kleef, Noé Guinko, and Amy Gimma as well as 2 other experts who preferred to remain anonymous, for their participation as experts in this study. We would also like to thank David Smith, Thibaut Jombart, Chris Jarvis, Flavio Finger, and Anton Camacho for their helpful advice in conducting this survey.

### **Ethics**

LSHTM ethics approval was obtained for this study (reference: 17633). Signed informed consent was taken from experts willing to participate and their verbal consent was requested again at the beginning of each elicitation.

### **Code and data availability**

All data and code used to process the expert interview responses can be found here: <https://github.com/epiforecasts/Ebola-Expert-Interviews>. The forecasts were performed using the EpiCastR package <https://github.com/epiforecasts/EpiCastR>. The code used for the analysis and scoring of the forecasts can be found here: <https://github.com/epiforecasts/Ebola-Expert-Elicitation>

### **Author contributions**

AR and WJE conceived and designed the interviews. JDM and SF conceived and designed the modelling framework and the evaluation of forecasts. AR conducted the expert interviews and prepared the interview data for comparison. JDM implemented the model and performed the formal forecast evaluations. JDM, AR, WJE and SF interpreted the results. JDM and AR wrote the manuscript. JDM, AR, WJE and SF edited the manuscript.

### **Funding statement**

This study was partly funded by the Department of Health and Social Care using UK Aid funding and is managed by the National Institute for Health and Care Research (VEEPED: PR-OD-1017-20002; AR and WJE). This study was partly funded by the Wellcome Trust (210758/Z/18/Z : JDM and SF). The views expressed in this publication are those of the authors and not necessarily those of the funders.

### **Conflicts of interest**

The authors have no conflicts of interest to declare

575  
576  
577

## References

- 578 1. Report of an International Commission. 1978 Ebola haemorrhagic fever in Zaire, 1976. *Bull. World*  
579 *Health Organ.* **56**, 271–293.
- 580 2. Rosello A *et al.* 2015 Ebola virus disease in the Democratic Republic of the Congo, 1976-2014. *Elife*  
581 **4**. (doi:10.7554/eLife.09015)
- 582 3. In press. Ebola virus disease. See [https://www.who.int/news-room/fact-sheets/detail/ebola-virus-](https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease)  
583 [disease](https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease) (accessed on 19 December 2023).
- 584 4. WHO Regional Office for Africa. 2020 Outbreaks and Emergencies Bulletin, Week 5: 27 January -  
585 02 February 2020.
- 586 5. Kucharski AJ, Camacho A, Flasche S, Glover RE, Edmunds WJ, Funk S. 2015 Measuring the  
587 impact of Ebola control measures in Sierra Leone. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14366–14371.
- 588 6. Adongo PB, Tabong PT-N, Asampong E, Ansong J, Robalo M, Adanu RM. 2016 Preparing towards  
589 Preventing and Containing an Ebola Virus Disease Outbreak: What Socio-cultural Practices May  
590 Affect Containment Efforts in Ghana? *PLoS Negl. Trop. Dis.* **10**, e0004852.
- 591 7. In press. WHO prequalifies Ebola vaccine, paving the way for its use in high-risk countries. See  
592 [https://www.who.int/news/item/12-11-2019-who-prequalifies-ebola-vaccine-paving-the-way-for-its-](https://www.who.int/news/item/12-11-2019-who-prequalifies-ebola-vaccine-paving-the-way-for-its-use-in-high-risk-countries)  
593 [use-in-high-risk-countries](https://www.who.int/news/item/12-11-2019-who-prequalifies-ebola-vaccine-paving-the-way-for-its-use-in-high-risk-countries) (accessed on 20 December 2023).
- 594 8. Woolsey C, Geisbert TW. 2021 Current state of Ebola virus vaccines: A snapshot. *PLoS Pathog.* **17**,  
595 e1010078.
- 596 9. Chowell G, Viboud C, Simonsen L, Merler S, Vespignani A. 2017 Perspectives on model forecasts  
597 of the 2014-2015 Ebola epidemic in West Africa: lessons and the way forward. *BMC Med.* **15**, 42.
- 598 10. Wannier SR *et al.* 2019 Estimating the impact of violent events on transmission in Ebola virus  
599 disease outbreak, Democratic Republic of the Congo, 2018-2019. *Epidemics* **28**, 100353.
- 600 11. Funk S *et al.* 2017 The impact of control strategies and behavioural changes on the elimination of  
601 Ebola from Lofa County, Liberia. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**.  
602 (doi:10.1098/rstb.2016.0302)
- 603 12. Farrow DC, Brooks LC, Hyun S, Tibshirani RJ, Burke DS, Rosenfeld R. 2017 A human judgment  
604 approach to epidemiological forecasting. *PLoS Comput. Biol.* **13**, e1005248.
- 605 13. Bosse NI *et al.* 2022 Comparing human and model-based forecasts of COVID-19 in Germany and  
606 Poland. *PLoS Comput. Biol.* **18**, e1010405.
- 607 14. Bosse NI, Abbott S, Bracher J, van Leeuwen E, Cori A, Funk S. 2023 Human judgement forecasting  
608 of COVID-19 in the UK. *Wellcome Open Res.* **8**, 416.
- 609 15. Recchia G, Freeman ALJ, Spiegelhalter D. 2021 How well did experts and laypeople forecast the  
610 size of the COVID-19 pandemic? *PLoS One* **16**, e0250935.

- 611 16. McAndrew T, Reich NG. 2022 An expert judgment model to predict early stages of the COVID-19  
612 pandemic in the United States. *PLoS Comput. Biol.* **18**, e1010485.
- 613 17. McAndrew T *et al.* 2022 Early human judgment forecasts of human monkeypox, May 2022. *Lancet*  
614 *Digit Health* **4**, e569–e571.
- 615 18. Watson-Jones D *et al.* 2022 Protocol for a phase 3 trial to evaluate the effectiveness and safety of a  
616 heterologous, two-dose vaccine for Ebola virus disease in the Democratic Republic of the Congo.  
617 *BMJ Open* **12**, e055596.
- 618 19. 2014 A web-based tool for eliciting probability distributions from experts. *Environmental Modelling*  
619 *& Software* **52**, 1–4.
- 620 20. Stan Team. 2012 *Stan Modeling Language User's Guide and Reference Manual*.
- 621  
622  
623  
624