

1 **MSM with HIV: Improving prevalence and risk estimates by a Bayesian small area**
2 **estimation modelling approach for public health service areas in the Netherlands**

3 Authors: Haoyi Wang¹, Chantal den Daas², Eline Op de Coul³, Kai J Jonas^{1*}

4 Affiliation:

5 1 Department of Work and Social Psychology, Maastricht University, Maastricht, the Netherlands

6 2 Health Psychology, University of Aberdeen, Aberdeen, United Kingdom

7 3 Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM),

8 Bilthoven, The Netherlands

9

10 *Correspondence to: Kai J Jonas

11 Postal address: Department of Work and Social Psychology, Maastricht University, Maastricht, 6200ER, the

12 Netherlands

13 Phone: +31 43 38 84068

14 Email: kai.jonas@maastrichtuniversity.nl

15

16 ORCIDs of authors:

17 HW: [0000-0002-1844-7156](https://orcid.org/0000-0002-1844-7156)

18 CDD: [0000-0003-0955-3691](https://orcid.org/0000-0003-0955-3691)

19 EODC: [0000-0002-8263-9187](https://orcid.org/0000-0002-8263-9187)

20 KJJ: [0000-0001-6607-1993](https://orcid.org/0000-0001-6607-1993)

21

22

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

23 Abstract

24 In many countries, HIV infections among MSM (MSMHIV) are closely monitored, and updated epidemiological
25 reports are made available annually, yet the true prevalence of MSMHIV can be masked for areas with small
26 population density or lack of data. Therefore, this study aimed to investigate the feasibility of small area estimation
27 with a Bayesian approach to improve HIV surveillance. Data from the European MSM Internet Survey 2017
28 (EMIS-2017, Dutch subsample, n=3,459) and the Dutch survey 'Men & Sexuality-2018' (SMS-2018, n=5,653)
29 were utilized in this study. We first applied a frequentist calculation to compare the observed relative risk of
30 MSMHIV per Public Health Services (GGD) region in the Netherlands. We then applied a Bayesian spatial
31 analysis and ecological regression to account for variance due to space and determinants associated with HIV
32 among MSM to obtain more robust estimates. Results of the prevalence and risk estimations from EMIS-2017 and
33 SMS-2018 converged with minor differences. Both estimations confirmed that the risk of MSMHIV is
34 heterogeneous across the Netherlands with some GGD regions, such as GGD Amsterdam [RR=1.21 (95% credible
35 interval 1.05-1.38) by EMIS-2017; RR=1.39 (1.14-1.68) by SMS-2018], having a higher-than-average risk.
36 Results from our ecological regression modelling revealed significant regional determinants which can impact on
37 the risk for MSMHIV. In sum, our Bayesian approach to assess the risk of HIV among MSM was able to close
38 data gaps and provide more robust prevalence and risk estimations. It is feasible and directly applicable for future
39 HIV surveillance as a statistical adjustment tool.

40

41 Word counts (247/250)

42

43 Keywords: HIV surveillance, small area estimation, Bayesian spatial analysis, MSM

44

45 Introduction

46 Background

47 For epidemiology in HIV, data are often characterised by a spatial or a spatio-temporal structure [1], however most
48 studies in the field of HIV often ignore these spatial characteristics during data analysis. Examining the data
49 provided by spatial analysis allows to identify men who have sex with men (MSM) with HIV (MSMHIV) clusters
50 and to explore how these clusters originate [1, 2]. Therefore, spatial information and influence are likely to produce
51 better surveillance models [1, 3, 4] in the context of a declining HIV epidemic in certain Global North regions,
52 and reaching the 90-90-90 goals set by UNAIDS [5, 6].

53 For instance, in the Netherlands, despite annual epidemiological reports of MSMHIV having been
54 provided by Stichting HIV Monitoring (SHM, the Dutch HIV monitoring foundation) [7], more accurate estimates
55 on the spatial distribution are still desirable to close data gaps and to identify contexts in need of targeted
56 interventions. Precise spatial distribution models of MSMHIV are crucial to assess areas of increased intervention
57 need, to better programme services and to eventually end the HIV epidemic [8]. In line with the suggestion from
58 Khan et al., we agree that HIV monitoring should go beyond urban/rural distinction to better inform policymakers
59 [8], and spatial analysis on a smaller geographical scale is necessary. However, in the case of incomplete data for
60 areas with small populations, the true risk of HIV can be hidden. Hence, more advanced techniques and
61 methodologies are required to obtain robust estimates.

62 To respond, multiple studies proposed and used different small area estimation (SAE) techniques: From
63 generalized additive models [9], over basic area-level models [10], to Poisson regression models [8]. In this study,
64 taking the Netherlands as an example, we propose a Bayesian solution, which has been shown to be particularly
65 effective and has been applied in several other epidemiological fields [3, 11-13], to estimate the posterior
66 distribution (the revised or updated probability of an event occurring after considering new information and other
67 uncertainties in the Bayesian inference [14]) of the prevalence and the risk of MSMHIV.

68 Bayesian approach to estimate better HIV clusters

69 Comparing aforementioned approaches which usually fail to pick up the random effects due to fixed geographical
70 foci, Bayesian modelling allows to account for similarities based on the neighbouring regions or by proximity, and
71 to present data on a spatial hierarchical structure that borrows strength from the overall geo-spatial entity [3]. This
72 hierarchical structure can thus be considered as a multi-level component which makes it possible to smoothen
73 estimates based on the structural relationship during the structured random effects estimation instead of only

74 treating the spatial information as a factor. Consequently, the smoothed HIV prevalence per small area by
75 accounting for the geo-spatial structure can be more robust and closer to the true prevalence [3]. It would thus be
76 possible to estimate a relative risk (RR) of HIV per area compared to the whole of the Netherlands, and to identify
77 regions with higher-than-average risk of MSMHIV [3].

78 In addition, it is important to understand socio-ecological facilitators and hurdles of MSMHIV jointly
79 with spatial distribution to unravel co-variations. A previous study has provided insights into the determinants of
80 HIV transmission in the Netherlands on individual level, such as HIV testing, older age, and other diagnosed
81 sexually transmitted infections (STIs) [15]. To evaluate these determinants of HIV transmission in the Netherlands
82 from an ecological perspective to better understand the national and local MSMHIV epidemic, we also included
83 these risk factors in a Bayesian spatial ecological regression model to explore how these risk factors impact on the
84 HIV spatial distribution in the Netherlands (for more details see Methodology section and Online supplementary
85 material).

86 This study thus sought to use data more effectively in order to generate opportunities to better understand
87 the HIV epidemic by identifying the regions with higher-than-average risk of MSMHIV, using data from the
88 Netherlands. We also aimed to fill the knowledge gap of how areal characteristics may impact MSMHIV by
89 applying Bayesian spatial modelling methodology to provide a more accurate epidemiologic spatial pattern of
90 MSMHIV using two independent survey-based datasets on HIV among MSM in the Netherlands. In addition, we
91 compared the results by two survey-based datasets to explore the stability and robustness of Bayesian spatial
92 analysis. Moreover, we compared the results with/without Bayesian inference to support future HIV surveillance
93 and to support local HIV prevention efforts.

94 Methodology

95 Study population and data sources

96 EMIS-2017

97 All MSM included in this dataset were recruited between 19 October 2017 and 30 January 2018 via the European
98 MSM Internet Survey (EMIS-2017, www.emis2017.eu) and were drawn from the Dutch subsample. EMIS-2017
99 was an anonymous, self-administered, and cross-sectional online survey conducted across 50 countries to inform
100 interventions for MSM which are highly affected by infections with HIV and other STIs [16]. EMIS-2017 recruited
101 3,851 MSM in the Netherlands. We excluded 392 (10.2%) men that failed to provide information on their place

102 of residence (final dataset n=3,459). Ethical approval for this survey was obtained from the Observational Research
103 Ethics Committee at the London School of Hygiene & Tropical Medicine (review reference 14421 /RR/8805).

104 SMS-2018

105 The cross-sectional Survey Men & Sexuality (SMS-2018), led by Soa Aids Netherlands and Utrecht University,
106 aimed to investigate the health, well-being, and sexuality of MSM in the Netherlands. It enrolled MSM between
107 February and June 2018 through social media, gay media and dating apps. The survey was distributed in six
108 languages to include a culturally diverse sample of MSM. All participants had provided informed consent prior to
109 accessing the survey [17]. SMS-2018 recruited 6,206 MSM in the Netherlands. We excluded 552 (8.9%) men
110 failed to provide their postal code (final dataset n=5,654). Ethical approval for this survey was obtained from the
111 Ethics Committee of the Faculty of Social and Behavioural Sciences, Utrecht University (FETC17-131).

112 Study area

113 There are 25 Public Health Services (GGD) regions in the Netherlands. The GGDs are organized regionally and
114 independently responsible to provide healthcare services and prevention, and to monitor the general health of the
115 population. Therefore, by summarizing results on the GGD level, we can better support both local and national
116 HIV monitoring and prevention. The geographical structure of the Netherlands with boundary geoinformation of
117 GGD regions were retrieved from Statistics Netherlands (CBS) [18] and linked with postal code datasets [19].

118 All MSM data retrieved were deidentified and were aggregated on GGD regional level to represent HIV
119 cases among MSM in the Netherlands separately from both datasets. Out of both samples, we included all MSM
120 who provided information on 2-digit postal code (4-digit postal code for SMS-2018), which were then linked to
121 the GGD regions. All data-aggregation processes were done separately for the two datasets without combination,
122 hence there are no overlapping observations concerns in this study, given that participants could participate in both
123 surveys.

124

125 **Small area estimation analysis**

126 Frequentist analysis

127 We first calculated the observed HIV prevalence per GGD region with its 95% confidence interval (95%CI) by
128 dividing the HIV counts by the numbers of MSM inhabitants participated in the surveys per region. We then
129 calculated the observed standardized prevalence ratio (SPR) per GGD region, which is defined as the ratio of the
130 observed counts to the expected counts using an indirect standardization approach, based on the overall risk of
131 HIV in the Netherlands and the total MSM population in each GGD region [20]. As a spatial epidemiological

132 measure, SPR could be applied to present the risk of HIV per GGD region compared to the overall risk of HIV in
133 the Netherlands on the regional/population level. In GGD regions where $SPR > 1$, this denotes a higher risk of
134 occurrence of HIV compared to the overall risk across the country, and GGD regions where $SPR < 1$, this denotes
135 a lower risk. However, in turn of regions with smaller populations sizes, SPR can be insufficiently reliable and
136 stable due to pure chance calculation [21].

137 Bayesian modelling (null model)

138 To estimate a more accurate prevalence by Bayesian approach, we first conducted model accounting only
139 the hierarchical spatial structure of GGD regions. Fig 1 presented the spatial structure on region connectivity of
140 the Netherlands on GGD regional level based on the common sharing borders or on proximity. To conduct the
141 modelling analysis, we used the Integrated Nested Laplace Approximation (INLA), which is designed for latent
142 Gaussian models, for the Bayesian computation. For the model parameters, we employed the Besag-York-Mollie
143 (BYM) model, which specifies the spatially structured residual using an intrinsic conditional autoregressive (iCAR)
144 distribution [22]. In addition, to specify the prior distribution for the Bayesian modelling, due to lack of information
145 on GGD regions in the Netherlands, we assigned a weak, understandable, conservative and useful Penalized
146 Complexity (PC) prior for the precision of the exchangeable random effects. For more detailed model assumptions
147 and prior distribution, please see the online resource file S1.

148



149

150 **Fig 1. Region connectivity matrix of the Netherlands at the Public Health Services level**

151 Note: for names and more details on the Public Health Services regions in the Netherlands, please see: <https://www.ggd.nl/>.

152

153 Bayesian spatial ecological regression modelling (final model)

154 Additional to the null model, we hypothesised that MSMHIV across the Netherlands can be influenced by the
155 established determinants of MSMHIV reported by den Daas et al. [15] using EMIS-2010 datasets. These
156 determinants included: prevalence of HIV testing (% ever tested); age (% ≥ 35 years old); median number of sex
157 partners; proportion of injecting drug users [(IDU), % IDU in EMIS-2017]/proportion of injecting drug use during
158 sex [(SLAM), % SLAM in SMS-2018]; proportion of never using condom with last partner (% never); prevalence
159 of syphilis (% yes), prevalence of chlamydia (% yes) and prevalence of gonorrhoea (% yes) in EMIS-2017 and

160 prevalence of syphilis, chlamydia and gonorrhoea in the past 6 months in SMS-2018. Detailed definition of
161 variables can be found in the published methodology paper for EMIS-2017 [16], and SMS-2018 [17].

162 Therefore, we applied a spatial ecological regression modelling technique [3] which takes into account
163 these selected determinants of MSMHIV summarized by aggregating the selected datasets. We first conducted
164 univariable models which only include one of the selected regional determinants and the spatial connectivity. We
165 then conducted multivariable models with the significant determinants indicated by the univariable models to
166 evaluate the impact on HIV prevalence in the Netherlands. We selected the final model using the backward
167 approach by comparing models' Deviance Information Criterion (DIC). The smaller DIC indicates the better
168 goodness of fit in model. In other words, estimates by model with a smaller DIC are more robust, and can be
169 consider closer to the true prevalence. For more detailed model assumptions and parameters, please see the online
170 resource file S2. Finally, we quantified an Intra-class correlation (ICC) to evaluate the proportion of variance
171 explained by the structured spatial component [3], and quantified the spatial random effects per GGD region based
172 on the spatial structure of the Netherlands to estimate the influence from the spatial structure of the Netherlands
173 on the GGD regional level on HIV prevalence for both the null spatial model and the final spatial ecological
174 regression model.

175 Computational analysis

176 All analyses were conducted in R (version 4.0.4). For all Bayesian modelling analyses with INLA, we used R-
177 INLA package (version 21.05.02) to empower our computational process [3].

178

179 Results

180 In this section, we first present the GGD regional characteristics and the results of the frequentist analysis to show
181 the spatial distribution of HIV in the Netherlands by GGD regions, using a classical frequentist approach for both
182 datasets. We then present the posterior prevalence and RR of HIV per GGD region estimated by Bayesian
183 modelling with only taking the spatial structure into account (null model). By comparing the results from the null
184 model and the naïve analysis, we can identify how the spatial structure itself impacts on the distribution of HIV in
185 the Netherlands beyond the chance levels. Next, we include the procedure of model selection by presenting the
186 univariable and multivariable regression modelling results to unravel which known individual-level determinants
187 of HIV are significantly related to HIV status on the areal level. Finally, the posterior prevalence and RR of HIV
188 per GGD region, estimated by the multivariable regression modelling (final model), are presented to be compared

189 with the results from the null model. The comparison can, therefore, inform how the HIV spatial distribution could
190 change after conditioning other determinants of HIV.

191

192 Study population characteristics

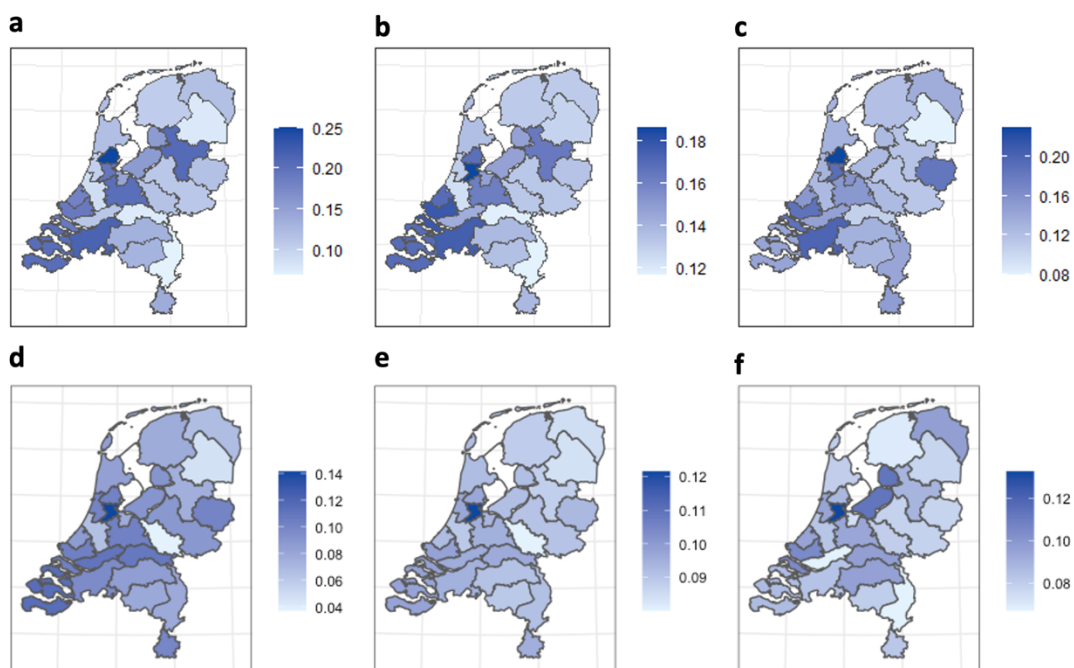
193 Regional characteristics relevant to MSMHIV across the Netherlands were heterogenous for both datasets (see
194 Online resource S1 table). For these established individual level determinants, HIV testing proportions ranged
195 from 64.1% in GGD Drenthe to 92.3% in GGD Amsterdam in EMIS-2017 (and ranged from 62.7% in GGD
196 Limburg-Noord to 92.7% in GGD Amsterdam in SMS-2018). Major differences between ever-/recent-diagnosed
197 STIs proportion among Dutch MSM were observed from the two datasets. For ever diagnosed STIs proportion in
198 EMIS-2017, ever diagnosed syphilis ranged from 10.3% in GGD Drenthe to 34.4% in GGD Zaanstreek/Waterland;
199 and ever diagnosed chlamydia ranged from 19.7% in GGD Gelderland-Zuid to 45.2% in GGD
200 Zaanstreek/Waterland. For recent diagnosed STIs (within six months) in SMS-2018, recently diagnosed syphilis
201 ranged from 18.0% in GGD Drenthe to 44.4% in GGD Amsterdam. More detailed information for other regional
202 characteristics (older than 35 years proportion, IDU proportion, and other STIs) per GGD region from both datasets
203 can be found in Online resource S1 table.

204

205 Frequentist observed HIV prevalence and risk among MSM

206 In terms of the prevalence of MSMHIV in the Netherlands, the observed overall prevalence of HIV among MSM
207 in 2017 was 14.2% in EMIS-2017 and 9.5% in SMS-2018. In EMIS-2017, the observed prevalence of HIV varied
208 by GGD regions in the Netherlands, with a range of 6.8% (95%CI 3.16-14.09) in GGD Limburg Noord to 25.0%
209 (95%CI 13.25-42.11) in GGD Zaanstreek/Waterland. In SMS-2018, the observed prevalence varied from 3.7%
210 (95%CI 1.59-8.38) in Veiligheids- en Gezondheidsregio Gelderland-Midden (VGGM) to 14.15% (95%CI 11.67-
211 17.06) in GGD Amsterdam (Fig 2a&2d, Online resource S2 table). The crude SPR in Fig 3-a and 3-d shows that
212 regions with higher-than-average risk of HIV exist in the Netherlands, with a range of 0.43 (GGD Limburg Noord)
213 to 1.59 (GGD Zaanstreek/Waterland) in EMIS-2017; and 0.39 (VGGM) to 1.49 (GGD Amsterdam) in SMS-2018.
214 The SPR trends corresponded with the patterns of the observed HIV prevalence. More detailed information for the
215 frequentist observed prevalence and SPR of MSMHIV can be found in Online resource S2 table.

216



217

218 **Fig 2. Choropleth map of the estimates of HIV prevalence by GGD regions in the Netherlands.**

219 A: Observed HIV prevalence by EMIS-2017. B: Posterior mean of HIV prevalence estimated by Bayesian spatial modelling (null model) by

220 EMIS-2017. C: Posterior mean of HIV prevalence estimated by Bayesian spatial ecological regression modelling (final model) by EMIS-2017.

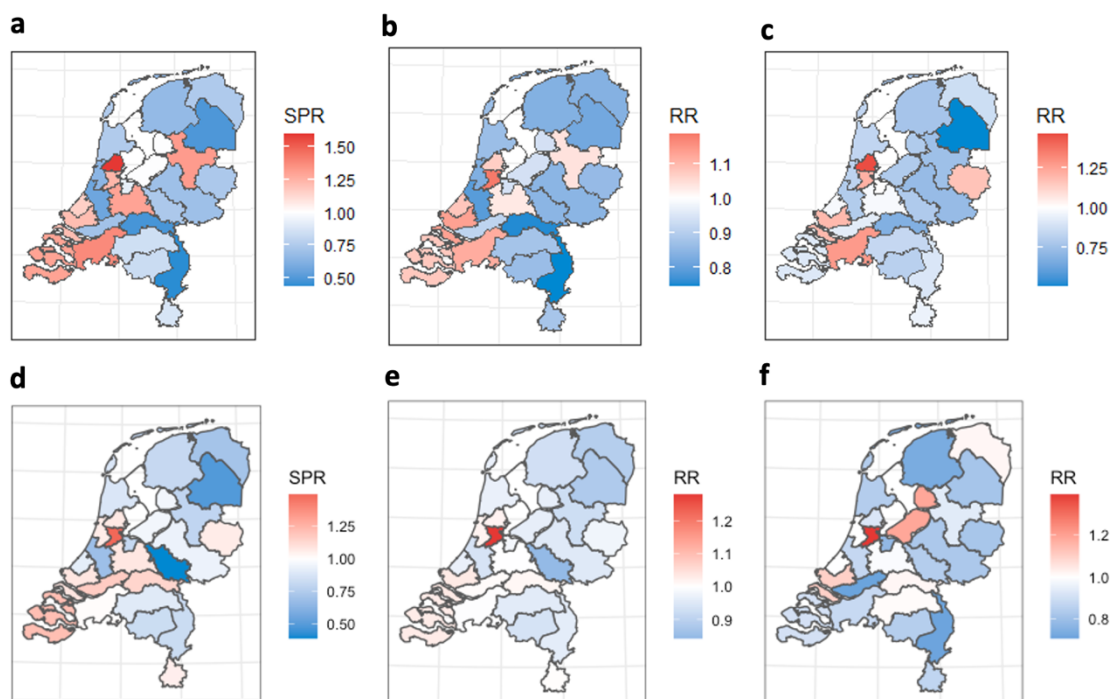
221 D: Observed HIV prevalence by SMS-2018. E: Posterior mean of HIV prevalence estimated by Bayesian spatial modelling (null model) by

222 SMS-2018. F: Posterior mean of HIV prevalence estimated by Bayesian spatial ecological regression modelling (final model) by SMS-2018.

223 See Online resource table S2 for the 95%CI or 95%CrI and other details.

224 Notes: the darker a GGD region, the higher the prevalence estimation of HIV among MSM.

225



226

227 **Fig 3. Choropleth map of the estimates of HIV risks by GGD regions in the Netherlands.**

228 A: Observed HIV standardised prevalence ratio by EMIS-2017. B: Posterior mean of HIV relative risk estimated by the Bayesian spatial
229 modelling (null model) by EMIS-2017. C: Posterior mean of HIV relative risk estimated by the Bayesian spatial ecological regression
230 modelling (final model) by EMIS-2017. D: Observed HIV standardised prevalence ratio by SMS-2018. E: Posterior mean of HIV relative risk
231 estimated by the Bayesian spatial modelling (null model) by SMS-2018. F: Posterior mean of HIV relative risk estimated by the Bayesian
232 spatial ecological regression modelling (final model) by SMS-2018. See Online resource table S2 for the 95%CI or 95%CrI and other details.
233 Notes: RR (or SPR) higher than 1 indicates a higher-than-average (average risk in the Netherlands) risk of HIV among MSM in that region
234 (red); RR (or SPR) lower than 1 indicates a lower-than-average risk of HIV among MSM in that region (blue).

235

236 HIV prevalence and risk among MSM after Bayesian spatial adjustment

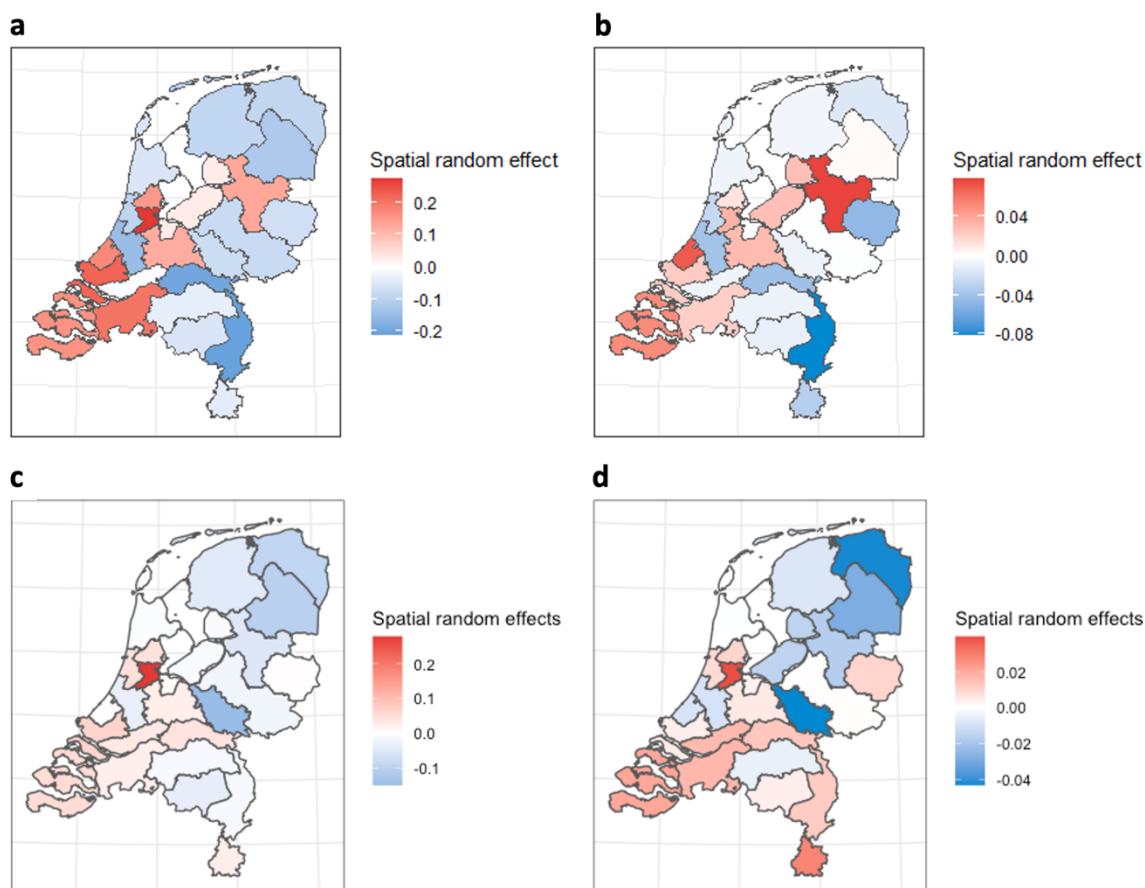
237 After accounting for the spatial effects based on the spatial structure of the Netherlands on the GGD region level
238 presented in Fig 1 without other regional determinants of HIV transmission, the EMIS-2017 ICC of the spatial
239 structure was estimated at 0.24, which indicates that around 24% of the observed variance of HIV among MSM in
240 the Netherlands can be explained by the spatial structure of the Netherlands on the GGD regional level, and the
241 SMS-2018 ICC was 0.27 (Table 1).

242 As indicated by the observed HIV prevalence, we observed heterogeneity of the posterior HIV prevalence
243 in the Netherlands estimated by the spatial null models in both datasets. In EMIS-2017, the highest posterior HIV

244 prevalence was found in the GGD Amsterdam of 18.6% (95%CrI 15.87-21.58) and the lowest in GGD Limburg-
245 Noord of 11.7% (95%CrI 7.2-16.4). GGD Amsterdam was estimated as the only region with statistically significant
246 higher-than-average risk of HIV among MSM in the Netherlands with a RR of 1.18 (95%CrI 1.01-1.37). Full
247 details of all GGD regions according to the spatial null model can be found in Online resource table S2 and Fig 3-
248 b. In SMS-2018, the posterior prevalence and RR was the similar as the observed prevalence, the highest posterior
249 HIV prevalence was found in Amsterdam, too, of 12.16% (95%CrI 9.58-15.14) with the only significant higher-
250 than-average risk of HIV of 1.28 (95%CrI 1.01-1.59), and the lowest in VGGM of 8% (95%CrI 5.09-10.69) with
251 RR of 0.84 (95%CrI 0.54-1.13), see Fig 3-e and Online resource S2 table.

252 Posterior spatial random effects on the HIV prevalence estimated by the null model can be found in Fig
253 4-a and 4-c, which confirms the spatial heterogeneity and indicates how the spatial structure impacts on the
254 estimated posterior RR per GGD region, with a range of [EMIS-2017: -0.21 (GGD Limburg-Noord) to 0.27 (GGD
255 Amsterdam)] and [SMS-2018: -0.15 (VGGM) to 0.28 (GGD Amsterdam)]. In other words, regions with a positive
256 (or negative) value of the spatial random effects indicate having an elevated (or lower) relative risk of HIV than
257 the overall risk in the Netherlands.

258



259

260 **Fig 4. Posterior spatial random effects on the GGD regional level in the Netherlands.**

261 A: posterior spatial random effects estimated by Bayesian spatial modelling (null model) by EMIS-2017. B: Posterior spatial random effects
262 estimated by the Bayesian spatial ecological regression modelling (final model) by EMIS-2017. C: posterior spatial random effects estimated
263 by Bayesian spatial modelling (null model) by SMS-2018. D: Posterior spatial random effects estimated by the Bayesian spatial ecological
264 regression modelling (final model) by SMS-2018. See Online resource table S2 for the 95%CrI and other details.

265 Notes: regions with a positive (negative) value of the spatial random effects indicate having an elevated (lower) relative risk of HIV among
266 MSM than the overall risk in the Netherlands.

267

268 HIV prevalence and risk among MSM after Bayesian spatial ecological adjustment

269 Univariable models

270 In EMIS-2017, after adjusting on the observed HIV testing prevalence as the regional determinant, a coefficient
271 of 2.72 (95%CrI 0.61-4.73, DIC=144.42, ICC=0.27) was modelled. This means that each increase of one percent
272 in HIV testing prevalence in a region is associated with an increase of around 2.8% ($=\exp(2.724 \times 0.01)$) in HIV
273 risk in that region. The coefficient for the observed syphilis prevalence was estimated at 3.55 (95%CrI 1.26-5.70,
274 DIC=142.41, ICC=0.28), which indicated that for every one percent increase of the regional prevalence of syphilis,

275 the regional risk of HIV increases by 3.5%. Similar for gonorrhoea, the increased risk was 2.3%. The other areal
276 determinants of HIV (age, number of partners, condom use, IDU prevalence and chlamydia prevalence were not
277 significantly associated with the posterior mean of RR of HIV in the Netherlands (Table 1).

278 In SMS-2018, we observed a significant positive association between HIV test and HIV with a coefficient
279 of 2.67 (95%CrI 0.74-4.50, ICC=0.31, DIC=118.97) in the univariable model. In addition, instead of STIs' spatial
280 prevalence, we found that proportion of higher age, coefficient=2.45 (95%CrI 0.55-4.23, ICC=0.28, DIC=119.84)
281 and proportion of never using a condom with non-steady partners, coefficient=3.12 (95%CrI 0.36-5.62, ICC=0.29,
282 DIC=121.04) were positively associated with the HIV prevalence on the GGD regional level. Other areal
283 characteristics of GGD regions were not estimated significant within this dataset (Table 1).

284

285 Multivariable models (final model)

286 After conditioning significant areal determinants of HIV, and selecting by the smallest DIC, in EMIS-2017, the
287 final model included HIV testing prevalence with a coefficient of 1.60 (95%CrI -0.60-3.74) and syphilis prevalence
288 with a coefficient of 2.67 (95%CrI 0.19-5.10), a DIC of 141.74, and an ICC of 0.28. The coefficients' estimations
289 indicate that both univariable models of HIV testing, and ever-diagnosed syphilis prevalence overestimated the
290 effects from these two regional determinants of HIV. Even though HIV testing was not statistically significant in
291 the final model, the DIC of the final model was smaller than the DIC of the model with only syphilis prevalence
292 (DIC=142.41). Therefore, we kept HIV testing in the final model (Table 1). In other words, with a smaller DIC,
293 the posterior prevalence and RR of MSMHIV estimated should be closer to the true prevalence and RR. In the
294 multivariable model based on SMS-2018, we included HIV test prevalence with coefficient=1.80 (95%CrI -0.29-
295 3.94) and proportion of higher age with coefficient=1.515 (95%CrI -0.55-3.57) with the smallest DIC of 118.58
296 and an ICC of 0.30 in the final model (Table 1).

297 The posterior prevalence of HIV was again heterogenous with both datasets in the Netherlands. In EMIS-
298 2017, the highest posterior prevalence of HIV was observed from GGD Zaanstreek/Waterland of 22.9% (95%CrI
299 16.25-30.8), and the lowest posterior prevalence of HIV was observed from GGD Drenthe of 7.89% (95%CrI
300 4.94-12.02). In addition, the final model succeeded to pick up the regions with higher-than-average risk of
301 MSMHIV in the Netherlands other than GGD Amsterdam (RR=1.21, 95%CrI 1.05-1.38): GGD Rotterdam-
302 Rijnmond (RR=1.19, 95%CrI 1.00-1.41) and GGD Zaanstreek/Waterland (RR=1.46, 95%CrI 1.04-1.96). Also,
303 the risk of HIV of GGD Noord- en Oost-Gelderland (RR=0.72, 95%CrI 0.54-0.94), GGD Fryslân (RR=0.76,
304 95%CrI 0.57-0.98), GGD Drenthe (RR=0.5, 95%CrI 0.31-0.77) and GGD Hollands-Midden (RR=0.79, 95%CrI

305 0.58-0.99) were found significantly lower than the risk on the average level in the Netherlands. In SMS-2018, the
306 lowest posterior prevalence was estimated in Dienst Gezondheid & Jeugd ZHZ of 6.71% (95%CrI 4.69-9.47) with
307 a RR of 0.71 (95%CrI 0.49-1.00) and GGD Limburg-Noord of 6.71% (95%CrI 3.46-11.42) with a RR of 0.71
308 (95%CrI 0.36-1.20). The highest posterior prevalence was again observed in GGD Amsterdam: 13.25% (95%CrI
309 10.84-15.93), and it was the only one GGD region with a significant higher-than-average risk of HIV in the
310 Netherlands (RR=1.39, 95%CrI 1.14-1.68). For full details of posterior HIV prevalence and RR per GGD region,
311 see Online resource table S2 and Fig 3-c and 3-f.

312 The spatial random effects indicated again the heterogeneity on the RR of MSMHIV across the
313 Netherlands with a range of [EMIS-2017: -0.08 (GGD Limburg Noord) to 0.08 (GGD Ijsselland); SMS-2018: -
314 0.04 (VGGM) to 0.04 (GGD Amsterdam)]. The posterior spatial random effect per GGD region were summarised
315 in Online resource table S2 and Fig 4-b and 4-d.

316

Table 1. Model comparison and selection for EMIS-2017 and SMS-2018

Models	HIV diagnosis									
	EMIS-2017					SMS-2018				
	Covariates	Coefficient	95%CrI	DIC	ICC	Covariates	Coefficient	95%CrI	DIC	ICC
Spatial Null model	Intercept	-0.108	(-0.257 – 0.025)	145.71	0.24	Intercept	-0.041	(-0.196 – 0.098)	123.99	0.27
	Intercept	-2.359	(-4.052 - -0.597)	144.42	0.27	Intercept	-2.206	(-3.729 - -2.217)	118.97	0.31
Spatial Univariable models	HIV test (%yes)*	2.724	(0.611 – 4.725)	145.89	0.24	HIV test (%yes)*	2.674	(0.738 – 4.501)	119.84	0.28
	Intercept	-0.591	(-1.376 – 0.172)	147.26	0.26	Intercept	-1.166	(-2.027 - -0.271)	125.66	0.27
	Age (%>= 35 y.o.)	1.089	(-0.613 – 2.786)	144.51	0.25	Age (%>= 35 y.o.)*	2.454	(0.551 – 4.229)	121.04	0.29
	Intercept	-0.063	(-0.232 – 0.083)	145.51	0.32	Intercept	0.073	(-0.498 – 0.645)	125.47	0.27
	Partner	-0.037	(-0.104 – 0.032)	142.41	0.28	Partner	-0.049	(-0.297 – 0.186)	125.62	0.27
	Intercept	-5.305	(-10.373 – 0.715)	146.6	0.24	Intercept	-0.971	(-1.763 - -0.130)	125.42	0.29
	Condom (%never)	41.315	(-6.485 – 81.421)	145.8	0.26	Condom (%never)*	3.121	(0.359 – 5.622)	125.64	0.27
	Intercept	-0.398	(-0.752 - -0.056)	141.74	0.28	Intercept	0.005	(-0.235 – 0.227)	118.58	0.30
	IDU (%yes)	4.41	(-0.369 – 9.051)	145.8	0.26	SLAM (%yes)	-0.561	(-2.812 – 1.601)		
	Intercept	-0.87	(-1.384 - -0.352)			Intercept	-0.049	(-0.327 – 0.220)		
	Syphilis (%yes)*	3.546	(1.263 – 5.703)			Syphilis (%yes)#	0.251	(-8.189 – 8.139)		
	Intercept	-0.662	(-1.281 - -0.029)			Intercept	-0.197	(-0.651 – 0.248)		
	Chlamydia (%yes)	1.724	(-0.200 – 3.514)			Chlamydia (%yes)#	1.992	(-3.513 – 7.301)		
	Intercept	-0.856	(-1.552 - -0.132)			Intercept	-0.125	(-0.533 – 0.277)		
Gonorrhoea (%yes)*	2.255	(0.116 – 4.221)			Gonorrhoea (%yes)#	1.195	(-4.376 – 6.498)			
Spatial Multivariable final model	Intercept	-2.021	(-3.624 - -0.358)			Intercept	-2.225	(-3.701 - -0.739)		
	HIV test (%yes)	1.607	(-0.6 – 3.737)			HIV test (%yes)	1.801	(-0.286 – 3.941)		
	Syphilis (%yes)	2.674	(0.192 – 5.099)			Age (%>= 35 y.o.)	1.515	(-0.549 – 3.571)		

Note: * = significant areal determinants of the univariable models. Partner = median number of partners. Condom = % never used condom with non-steady partners. # Indicates six-month prevalence instead of life-time prevalence. CrI = credible interval. DIC=Deviance Information Criterion, ICC= Intra-class correlation

320 Discussion

321 To illustrate the usefulness of SAE modelling with a Bayesian approach, we investigated the spatial distribution
322 of MSMHIV in conjunction with determinants of MSMHIV using data from the Netherlands at the level of the
323 Public Health Services regions (GGD). We applied this methodology on two independent survey-based datasets
324 to explore the applicability and the estimates of accuracy for the HIV surveillance at the same time.

325 Based on both datasets, we observed a heterogenous spatial distribution of MSMHIV: There are GGD
326 regions which showed higher-than-average risk. In particular, the GGD Amsterdam region as expected, and GGD
327 Zeeland, had a significantly higher-than-average risk of MSMHIV in the Netherlands. Jointly with the spatial
328 patterns, we identified regional determinants to be significantly associated with MSMHIV prevalence in the
329 Netherlands. Methodologically, we found that the observed prevalence estimated by the frequentist analysis was
330 less stable than the posterior prevalence estimated by the Bayesian spatial modelling in terms of the estimations
331 range and their uncertainty range (the 95% confidence interval and 95% credible interval), especially for regions
332 with smaller sample sizes. Despite a largely overlapping spatial distribution and heterogeneity of HIV in the
333 Netherlands (both Frequentist observed, and Bayesian smoothed) between the two datasets, minor differences in
334 the prevalence and spatial random effects were obtained.

335

336 Spatial distribution of HIV among MSM in the Netherlands

337 Overall, based on the overlapping results from both datasets, we observed a higher prevalence of HIV in the West
338 of the Netherlands where also the main urban areas (in Dutch: Randstad) are located, and in the GGD region of
339 Zeeland, which belongs to the area that has the highest concentration of conservative orthodox Calvinist
340 Protestants in the country [23].

341 It was within our expectation that the prevalence of MSMHIV was higher in the GGD regions in the
342 Randstad, such as GGD Amsterdam. This prevalence is also in line with findings from previous studies using
343 surveillance data by geographic information system and survey-based data [15, 24]. In addition, our analysis based
344 on both datasets suggested a significant higher-than-average risk of HIV among MSM in this GGD region
345 compared to other regions in the Netherlands. Few reasons may explain our findings. Firstly, Amsterdam which
346 is known as the ‘Gay Capital of Europe’ is the target of “gay tourism”, with more sexual encounters occurring
347 subsequently. [25]. Likewise, more Dutch MSM choose to relocate to these main urban areas [17], and the HIV
348 cases would, therefore, be concentrated there as well. Another reason that contributed to a higher HIV prevalence

349 is the high HIV testing rate among MSM in GGD Amsterdam region (Online resource S1 table). Our ecological
350 modelling analysis also confirmed this argument that with a higher HIV testing prevalence, the risk of HIV in that
351 region would be higher as well (Table 1).

352 It was, however, not expected that the GGD Zeeland also had a higher spatial risk (random effect, Fig 4)
353 of HIV among MSM compared to other regions. One reason for this higher risk found for this region may be due
354 to the religion/local culture. As one of the most conservative regions in the country with associated negative views
355 on same-sex sexual activities and relations, an overall negative attitude towards homosexuality may be greater
356 than in other regions [26]. In turn, some sexual behaviours may be stigmatized and MSM may experience more
357 barriers to HIV testing, which may influence the risk of MSMHIV at that region: according to both datasets GGD
358 Zeeland has one of the lowest HIV test prevalence among MSM (Online resource S1 table). Second, a longer
359 distance to the STI clinics could play a role as a barrier to HIV testing in GGD Zeeland [27], which could also
360 leave an influence on the spatial distribution of MSMHIV in the Netherlands. Therefore, future studies should
361 also investigate the distance to the STI clinics as a regional determinant for a more comprehensive model.

362 Differences between estimations by EMIS-2017 and SMS-2018

363 Despite the large overlap of the spatial pattern of MSMHIV by our analysis based on EMIS-2017 and SMS-2018
364 datasets, we observed some minor differences in terms of both the observed and estimated posterior prevalence,
365 which is generally lower in SMS-2018 data compared to EMIS-2017 data. One reason that may explain this
366 finding is that data collection variations existed between EMIS-2017 and SMS-2018. The collection methods and
367 process were different between the SMS-2018 and EMIS-2017. It would thus lead to collection variations and
368 resulted in different estimates. However, in terms of the posterior relative risk of HIV on the GGD regional level,
369 the results between these two datasets converged and can reflect how the regional risks differed between different
370 GGD regions on the national scale. This converged posterior RR estimations by both datasets indicate the strong
371 stability of the Bayesian spatial analysis to identify regions with higher risk for prevention efforts allocation
372 strategies.

373 We also observed a different impact of the areal determinants on our HIV prevalence and risk modelling
374 between these two datasets. Despite the discussed sampling variations, different definitions of the determinants
375 when collecting data through the surveys could also explain why our univariable models and final models are
376 different. For example, in EMIS-2017, men were asked if they were ever diagnosed with any type of STI instead
377 of STI diagnosis within the past six months, as in SMS-2018. The HIV-risk profile and sexual behaviour profile
378 of a MSM would thus be different and result in different impact on the ecological modelling analysis. Therefore,

379 based on our findings in the univariable models, we could also conclude that the impact of the lifetime STI
380 diagnoses should be greater than the recent STI diagnoses.

381

382 [Application of Bayesian spatial modelling analysis](#)

383 The application of Bayesian spatial modelling analysis in two survey-based datasets from the Netherlands proved
384 that modelling the HIV distribution with a Bayesian approach is feasible, and robust when comparing results
385 between two datasets. Compared to calculating the observed prevalence and SPR, results of the posterior
386 prevalence and risks estimations were smoother and more stable due to the narrower credible intervals estimated
387 by INLA, which has been proven helpful to estimate the more accurate prevalence and risk as an approximation
388 approach [3]. It thus delivers more certainty when interpreting the results and tailoring prevention programming
389 for HIV.

390 In addition, our spatial ecological modelling allowed us to investigate the variations of HIV based on the
391 spatial connectivity together with other regional determinants of HIV. We found several regional determinants
392 (Table 1) based on our survey data useful to estimate the posterior prevalence and risk of HIV in the Netherlands.
393 Consequently, both two final models for EMIS-2017 and SMS-2018 data improved the goodness of fit after
394 adding the regional determinants related to MSMHIV, and we believe the estimations from the final model should
395 be closer to the true prevalence compared to the observed frequentist calculation and models without covariates.
396 Therefore, the established spatial determinants from this study should be considered valuable for policymakers
397 and HIV surveillance authorities. Attention should be also given to these regional HIV determinants instead of
398 focussing only the numerical prevalence only.

399 We thus recommend promoting this novel methodology as a statistical adjustment for future HIV
400 national/local surveillance, especially when there are gaps due to missing data, or regional prevalence estimates
401 are needed. Even though we acknowledge that the complex statistical computation, unfamiliarity and limited
402 knowledge on Bayes' Theorem may limit the application of this methodology for non-Bayesian stakeholders, the
403 already available techniques and the various forms of open source statistical software [3, 28, 29] should help to
404 ease the computation process and help interpreting results. We believe applying SAE with a Bayesian approach
405 can help to robustly tailor HIV prevention programmes, especially local HIV prevention resources and services
406 navigation.

407

408 **Strength and limitations**

409 We acknowledge the following strengths and limitations of our study. One major strength of this study is the
410 introduction and the application of Bayesian spatial modelling analysis as a SAE method. We considered our
411 results, especially the posterior risks of MSMHIV, as robust and valuable for HIV related public health policies
412 and prevention strategies. The methodology in our study can be directly applied in other countries in the future
413 for small area estimations using surveillance data on HIV. Another strength is the convergence of the models
414 based on data from two survey-based datasets. Data from these two surveys made MSM individual level covariates
415 directly available for the posterior modelling analysis instead of using secondary area-level covariate data based
416 on the Dutch general population. Moreover, presenting data on the GGD regional level also helped to prevent
417 information bias due to the municipal location of HIV testing. Since the regional public health service runs the
418 majority of HIV tests in the Netherlands and since the sexual health clinics are located in the larger municipalities
419 in a GGD region, data may thus concentrate in these bigger cities if HIV among MSM would be assessed on the
420 municipality level.

421 In addition to the aforementioned limitations, one limitation can be the lack of data from the neighbouring
422 regions from other countries. Our Bayesian spatial analysis with a hierarchical structure revealed how regions
423 may influence each other to smoothen the risk estimates based on neighbouring information or on proximity.
424 However, given the smoothing by neighbouring regions, our analysis may be influenced by other regions outside
425 the Netherlands. It should be stressed that for some GGD regions which are located in the border regions of the
426 Netherlands, the estimated prevalence and risks of HIV of these regions would thus be less stable compared to
427 the rest due to the lower predictability as only one other node is available and thus part of spatial information is
428 missing. Regions that share a boarder with Germany and Belgium, especially for GGD Zuid-Limburg which is
429 only geographically connected with GGD Limburg-Noord and without other neighbouring regions in the
430 Netherlands (Fig 1), require additional cross-border data input. Therefore, a study including those neighbouring
431 regions in Belgium and Germany may be warranted in the future to compensate for the problem of lack of national
432 spatial connectivity for those boarder regions. To achieve this aim, comparable cross-border data needs to be
433 accessible, too. Moreover, our spatial analysis of MSMHIV across the Netherlands was based on survey data from
434 2017-2018 when the pre-exposure prophylaxis (PrEP) has yet to be formally introduced in the Netherlands (2019).
435 Our spatial model, therefore, did not include PrEP use among MSM per GGD region as a regional characteristic.
436 Consequently, the influence from PrEP use was not measured in our models. Given the established impacted on
437 the HIV prevention among MSM from using PrEP [30-34], future studies should therefore include PrEP use into

438 the spatial models for a more robust estimation. Another limitation can be the lack of an informative prior
439 distribution when conducting Bayesian spatial analysis. Previous studies which applied Bayesian statistic in other
440 epidemiologic field has suggested that to acquire the true prevalence and RR, an informative prior is preferred
441 and required in practice [35, 36]. Our application of the PC prior, as a weakly informative prior, may thus limit
442 the robustness of our posterior estimation and make them conservative [36]. However, we believe our estimations
443 were still robust and close to the true risk of HIV among MSM based on the previous sensitivity analysis of PC
444 in a prior experiment [37]. In addition, even though our Bayesian approach made our estimations more robust,
445 more comprehensive datasets, such as routine surveillance data are still warranted. Another major limitation in
446 our study may be the lack of temporal dimension in our models. The scope of our study to offer a time-dynamic
447 epidemiologic picture on how MSMHIV spatially distribute over the time is limited. Future studies thus should
448 include a wide temporal period to support a more comprehensive spatio-temporal analysis. Finally, ecological
449 fallacy is possible due to our ecological study design. We lose information on the individual-level due to
450 aggregating information spatially. Our results on the roles of the regional characteristics thus cannot be directly
451 applied to investigate/predict the MSM's HIV risk profile on the individual level.

452

453 Conclusion

454 In conclusion, our study proposed a Bayesian approach to more accurately assess the risk of HIV among MSM
455 using data from the Netherlands on the public health service regional level with more robust prevalence and risk
456 estimations over the use of crude proportions. Our findings based on two independent surveys can be considered
457 valuable for policymakers and HIV surveillance authorities for resources allocation decision by prioritizing
458 resources to the regions which require more efforts to reduce the burden of HIV among MSM accordingly. Based
459 on the Dutch data, our method has shown to be feasible and can be directly applied to achieve a more
460 comprehensive and robust surveillance of HIV in any geographic context.

461

462 Acknowledgement

463 We thank all participants of the surveys and the individuals involved in preparation, execution and analysis of the
464 surveys. We thank Wim Zuilhof, Bouko Bakker, Aryanti Radyowijati, Koenraad Vermey, Arjan van Bijnen and
465 the EMIS board for their invaluable help for the EMIS-2017 data. In addition, we thank John de Wit, Philippe
466 Adam, and Wim Zuilhof, for their role in the development and collection of SMS-2018 data.

467 Reference

- 468 1. Shrestha S, Bauer CXC, Hendricks B, Stopka TJ. Spatial epidemiology: An empirical
469 framework for syndemics research. *Social Science & Medicine*. 2020;113352.
- 470 2. Meyers DJ, Hood ME, Stopka TJ. HIV and Hepatitis C Mortality in Massachusetts, 2002–
471 2011: Spatial Cluster and Trend Analysis of HIV and HCV Using Multiple Cause of Death. *PLOS*
472 *ONE*. 2014;9(12):e114822.
- 473 3. Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with
474 R-INLA. *Spatial and Spatio-temporal Epidemiology*. 2013;4:33-49.
- 475 4. Omez-Rubio V, Best N, Richardson S, Li G, Clarke P. Bayesian Statistics Small Area
476 Estimation. 2010.
- 477 5. UNAIDS. 90–90–90 - An ambitious treatment target to help end the AIDS epidemic
478 2017 [Available from: [https://www.unaids.org/sites/default/files/media_asset/90-90-](https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf)
479 [90_en.pdf](https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf).
- 480 6. Marsh K, Eaton JW, Mahy M, Sabin K, Autenrieth CS, Wanyeki I, et al. Global, regional
481 and country-level 90-90-90 estimates for 2018: assessing progress towards the 2020 target.
482 *AIDS (London, England)*. 2019;33 Suppl 3(Suppl 3):S213-S26.
- 483 7. van Sighem A.I. WF, Boyd A., Smit C., Matser A., Reiss P. Monitoring Report 2020.
484 Human Immunodeficiency Virus (HIV) Infection in the Netherlands. Amsterdam: Stichting HIV
485 Monitoring, 2020. 2020 [Available from: [https://www.hiv-](https://www.hiv-monitoring.nl/en/resources/monitoring-reports)
486 [monitoring.nl/en/resources/monitoring-reports](https://www.hiv-monitoring.nl/en/resources/monitoring-reports).
- 487 8. Khan SS, McLain AC, Olatosi BA, Torres ME, Eberth JM. Small area estimation of
488 county-level U.S. HIV-prevalent cases. *Annals of Epidemiology*. 2020;48:30-5.e9.
- 489 9. Wand H, Morris N, Reddy T. Temporal and spatial monitoring of HIV prevalence and
490 incidence rates using geospatial models: Results from South African women. *Spatial and*
491 *Spatio-temporal Epidemiology*. 2021;37:100413.
- 492 10. Gutreuter S, Igumbor E, Wabiri N, Desai M, Durand L. Improving estimates of district
493 HIV prevalence and burden in South Africa using small area estimation techniques. *PLoS One*.
494 2019;14(2):e0212445.
- 495 11. Konstantinoudis G, Schuhmacher D, Ammann RA, Diesch T, Kuehni CE, Spycher BD.
496 Bayesian spatial modelling of childhood cancer incidence in Switzerland using exact point data:
497 a nationwide study during 1985-2015. *Int J Health Geogr*. 2020;19(1):15.
- 498 12. Tsiko RG. Bayesian spatial analysis of childhood diseases in Zimbabwe. *BMC Public*
499 *Health*. 2015;15:842.
- 500 13. Li M, Baffour B, Richardson A. Bayesian spatial modelling of early childhood
501 development in Australian regions. *International Journal of Health Geographics*.
502 2020;19(1):43.
- 503 14. Kelter R. Analysis of Bayesian posterior significance and effect size indices for the two-
504 sample t-test to support reproducible medical research. *BMC Medical Research Methodology*.
505 2020;20(1):88.
- 506 15. den Daas C, Goenee M, Bakker BHW, de Graaf H, Op de Coul ELM. Comparing
507 databases: determinants of sexually transmitted infections, HIV diagnoses, and lack of HIV
508 testing among men who have sex with men. *BMC Public Health*. 2015;15:1114-.
- 509 16. Weatherburn P, Hickson F, Reid DS, Marcus U, Schmidt AJ. European Men-Who-Have-
510 Sex-With-Men Internet Survey (EMIS-2017): Design and Methods. *Sexuality Research and*
511 *Social Policy*. 2020;17(4):543-57.

- 512 17. den Daas C, Zuilhof W, van Bijnen A, Vermey K, Dörfler T, de Wit J, et al. Rapport Survey
513 Mannen & Seksualiteit 2018 2018 [Available from: <https://www.soa aids.nl/files/2019-06/Rapport-SMS-Mannen-en-seksualiteit-MSM-Soa-Aids-Nederland-2018.pdf>.
514
- 515 18. CBS. CBS gebiedsindelingen 2020 [Available from:
516 [http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/effe1a](http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/effe1a b0-073d-437c-af13-df5c5e07d6cd?tab=relations)
517 [b0-073d-437c-af13-df5c5e07d6cd?tab=relations](http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/metadata/effe1a b0-073d-437c-af13-df5c5e07d6cd?tab=relations).
518
- 519 19. CBS. Dataset: CBS Postcode statistieken 2017 [Available from:
520 <https://www.pdok.nl/introductie/-/article/cbs-postcode-statistieken>.
521
- 522 20. Becher H, Winkler V. Estimating the standardized incidence ratio (SIR) with incomplete
523 follow-up data. BMC Medical Research Methodology. 2017;17(1):55.
- 524 21. Moraga P. Geospatial Health Data: Modeling and Visualization with R-INLA and
525 Shiny2019.
- 526 22. Morris M, Wheeler-Martin K, Simpson D, Mooney SJ, Gelman A, DiMaggio C. Bayesian
527 hierarchical spatial models: Implementing the Besag York Mollié model in stan. Spatial and
528 Spatio-temporal Epidemiology. 2019;31:100301.
- 529 23. CBS. Fewer churchgoers, especially among Catholics 2014 [Available from:
530 <https://www.cbs.nl/en-gb/news/2014/40/fewer-churchgoers-especially-among-catholics>.
531
- 532 24. Op de Coul ELM, Joore IK, van Sighem A, Bom BCJ, Hillebregt M, Prins JM, et al.
533 [Mapping HIV prevalence in the Netherlands with geographic information systems]. Ned
534 Tijdschr Geneesk. 2017;161:D965.
- 535 25. Richards G, Wilson J. Tourism, creativity and development. London: Routledge; 2007.
- 536 26. Keuzenkamp S. Acceptance of Homosexuality in the Netherlands. International
537 Comparison, Trends and Current Situation.: Netherlands Institute for Social Research; 2011.
- 538 27. Twisk DE, Meima B, Nieboer D, Richardus JH, Götz HM. Distance as explanatory factor
539 for sexual health centre utilization: an urban population-based study in the Netherlands.
540 European Journal of Public Health. 2021;31(6):1241-8.
- 541 28. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - A Bayesian modelling
542 framework: Concepts, structure, and extensibility. Statistics and Computing. 2000;10(4):325-
543 37.
- 544 29. Li ZR, Martin BD, Dong TQ, Fuglstad G-A, Paige J, Riebler A, et al. Space-Time
545 Smoothing of Demographic and Health Indicators using the R Package SUMMER. arXiv
546 preprint arXiv:200705117. 2020.
- 547 30. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure
548 prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from
549 the pilot phase of a pragmatic open-label randomised trial. Lancet. 2016;387(10013):53-60.
- 550 31. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure
551 Chemoprophylaxis for HIV Prevention in Men Who Have Sex with Men. New England Journal
552 of Medicine. 2010;363(27):2587-99.
- 553 32. Hoornenborg E, Krakower DS, Prins M, Mayer KH. Pre-exposure prophylaxis for MSM
554 and transgender persons in early adopting countries. AIDS (London, England).
555 2017;31(16):2179-91.
- 556 33. Hoornenborg E, Coyer L, Achterbergh RCA, Matser A, Schim van der Loeff MF, Boyd A,
et al. Sexual behaviour and incidence of HIV and sexually transmitted infections among men
who have sex with men using daily and event-driven pre-exposure prophylaxis in AMPREP: 2
year results from a demonstration study. The lancet HIV. 2019;6(7):e447-e55.

- 557 34. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, et al. Uptake of
558 pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender
559 women who have sex with men: a cohort study. *Lancet Infect Dis.* 2014;14(9):820-9.
- 560 35. Goldstein ND, Wheeler DC, Gustafson P, Burstyn I. A Bayesian approach to improving
561 spatial estimates of prevalence of COVID-19 after accounting for misclassification bias in
562 surveillance data in Philadelphia, PA. *Spat Spatiotemporal Epidemiol.* 2021;36:100401.
- 563 36. Lemoine NP. Moving beyond noninformative priors: why and how to choose weakly
564 informative priors in Bayesian analyses. *Oikos.* 2019;128(7):912-28.
- 565 37. Simpson D, Rue H, Riebler A, Martins TG, Sorbye SH. Penalising Model Component
566 Complexity: A Principled, Practical Approach to Constructing Priors. *Statist Sci.* 2017;32(1):1-
567 28.
568
- 569

570 Statements & Declarations

571 Funding

572 The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

573

574 Competing Interest

575 The authors have no relevant financial or non-financial interests to disclose.

576

577 Author Contributions

578 All authors contributed to the study conception and design. Material preparation, data analysis were performed
579 by [Kai J Jonas], [Chantal den Daas], and [Haoyi Wang]. The first draft of the manuscript was written by [Haoyi
580 Wang] and all authors commented on previous versions of the manuscript. All authors read and approved the final
581 manuscript.

582

583 Ethics approval

584 This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the
585 Observational Research Ethics Committee at the London School of Hygiene & Tropical Medicine (review
586 reference 14421/RR/8805) and the Ethics Committee of the Faculty of Social and Behavioural Sciences, Utrecht
587 University (FETC17-131).

588

589 Consent to participate

590 Informed consent was obtained from all individual participants included in the studies.