Zoonotic host diversity increases in human-dominated ecosystems

Rory Gibb[1†], David W. Redding[1†*], Kai Qing Chin[1], Christl A. Donnelly[2,3], Tim M. Blackburn[1,4], Tim Newbold[1] and Kate E. Jones[1,4*]

Affiliations:

[1]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, United Kingdom.

[2]Department of Statistics, University of Oxford, St Giles', Oxford, OX1 3LB, United Kingdom.

[3]MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, W2 1PG, United Kingdom.

[4]Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, United Kingdom.

*Correspondence to: D.W. Redding (d.redding@ucl.ac.uk) or K.E. Jones (kate.e.jones@ucl.ac.uk).

[†]Denotes equal author contributions.

Word count: Summary paragraph 243, Summary + Main text 2025, References 1392, Figure legends 502, Methods 3586.

Main text

Land use change (e.g. agriculture, urbanization) is widely recognised to influence zoonotic disease risk and emergence in humans[1,2], but whether this is underpinned by predictable ecological changes remains unclear[3]. In particular, it has been hypothesised that systematic differences in species resilience to human impacts, linked to traits, life histories and phylogeny, might result in habitat disturbance causing predictable changes in potential reservoir host diversity and species composition[4,5]. Here, we analyse 6801 ecological assemblages and 376 host species worldwide, controlling for research effort, and show that land use has global and systematic effects on local zoonotic host communities. Known wildlife hosts of human-shared pathogens and parasites overall comprise a significantly greater proportion of local species richness (18%-72% increase) and total abundance (21%-144% increase) in sites under substantial human use (secondary, agricultural and urban ecosystems) than in nearby undisturbed habitats. The magnitude of this effect varies taxonomically and is strongest for rodent, bat and passerine bird zoonotic host species, which may be one factor underpinning the global importance of these taxa as zoonotic reservoirs. Crucially, we further show that mammal species that harbor more pathogens overall (either human-shared or non human-shared) are more likely to occur in human-managed ecosystems, suggesting that these trends may be mediated by ecological or life-history traits that influence both host status and human-tolerance[6,7]. Our results suggest that global changes in mode and intensity of land use are creating growing hazardous interfaces between people, livestock and wildlife reservoirs of zoonotic disease.

Anthropogenic environmental change impacts many dimensions of human health and wellbeing, including the incidence and emergence of zoonotic and vector-borne diseases[1]. Although large-scale research into environmental drivers of disease has mostly focused on climate, there is growing consensus that land use change (conversion of natural habitats to agricultural, urban or otherwise anthropogenic ecosystems) is a globally-significant mediator of human infection risk and disease emergence[2,4]. Land use change directly and indirectly drives biodiversity loss, turnover and homogenisation (including through invasions and rare species losses)[8,9], modifies landscape structure in ways that modulate epidemiological processes (e.g. fragmentation[10], resource provisioning[11]) and can increase human-wildlife

contact (e.g. via agricultural practices or hunting)[1]. These processes interact to influence transmission dynamics in reservoir and vector communities and ultimately spillover risk to humans[12,13], with land use change implicated in driving both endemic (e.g. trypanosomiasis[14], malaria[15]) and epidemic (e.g. Nipah[16], West Nile[17]) zoonoses. However, the complexity of these systems (Extended Data Fig. 1) has made it difficult to identify whether land use has consistent effects on the ecological factors underpinning zoonotic disease risk[2], a critical knowledge gap given ongoing global land change trends[18].

Although there is broad evidence for regulatory effects of local species diversity on pathogen transmission[19], such effects are not universal: higher disease risk in depauperate assemblages has been observed for some disease systems (e.g. *Borrelia*[20], West Nile[17], *Ribeiroia*[7]) but not others. One ecological factor underlying these inconsistencies may be differences in host species sensitivity to human pressures[5]. It is often proposed that more effective zoonotic host species might be generally more likely to persist in disturbed ecosystems, since certain trait profiles (e.g. 'fast' life-histories, higher population densities) correlate to both reservoir status and reduced extirpation risk in several vertebrate taxa[21,22]. Alternatively, any such tendencies might be taxonomically or geographically idiosyncratic: for example, mammals that are more closely phylogenetically-related to humans are more likely to be zoonotic reservoirs[23], but may also be highly variable in their sensitivity to human impacts[21]. Reservoir host responses to disturbance have been investigated in certain taxa (e.g. primates[24]) and disease systems[14,20], but to date there has been no comprehensive analysis of the effects of land use on zoonotic host diversity and species composition.

Here, we use a global dataset of 6801 ecological assemblages derived from the Projecting Responses of Ecological Diversity in Changing Terrestrial Systems (PREDICTS) biodiversity database[25], to test whether land use has systematic effects on the zoonotic potential of wildlife communities. We identified records of wildlife hosts of known human pathogens and endoparasites (henceforth *'pathogens'*) within PREDICTS using a comprehensive host-pathogen associations database, and classified species as zoonotic hosts (henceforth *'hosts'*) based on evidence of association with at least one human-shared pathogen (Methods). PREDICTS compiles >3.2 million species records from 666 published studies that sampled biodiversity across land use gradients using consistent protocols, enabling global comparison of local assemblages in primary vegetation (minimally-disturbed baseline) to nearby secondary (recovering from past disturbance), managed (cropland, pasture, plantation) and urban sites, of varying use intensities (here, minimal or substantial-use)[25]. We identified records of 376 host species in a dataset of 6801 survey sites from 184

studies across 6 continents, with a taxonomic distribution broadly representative of known zoonotic host diversity (Figure 1, Supp. Tables 1-2; Methods). We compared host responses to land use to those of all other species at the same locations (*'non-hosts'*, approximating the response of background biodiversity; n=6512 species), using Bayesian mixed-effects models to control for study methods and sampling design (Methods). Pathogen detection is sensitive to research effort, such that some poorly studied species might be misclassified as non-hosts. We account for this uncertainty in our models using a bootstrap approach, with each iteration transitioning a proportion of non-host species to host status, with species-level transition rates determined by both publication effort and taxonomic order (Supp. Methods 1, Extended Data Fig. 2). All parameter estimates are obtained across each full bootstrap ensemble (Methods).

We first estimated the effects of land use type and intensity on two community metrics: site-level host species richness (number of host species; related to potential pathogen richness) and host total abundance (total number of host individuals; a more epidemiologically-relevant metric related to opportunities for transmission)[26]. Both host richness and total abundance either persist or increase in response to land use, against a background of consistent declines in all other (non-host) species in human-dominated habitats (Figure 2a-b). Together these changes lead to hosts comprising an increasing proportion of ecological assemblages in secondary, managed and urban land (Figure 2c-d, Supp. Tables 3-5). Notably, land use intensity has clear positive effects on community zoonotic potential both within and between land use types, with largest increases in substantial-use secondary and managed (posterior median: +18-21% host proportion richness, +21-26% proportion abundance) and urban sites (+62-72% proportion richness, +136-144% proportion abundance; but with higher uncertainty due to sparser sampling). These results are robust to testing for sensitivity to random study-level variability (Extended Data Fig. 3a), geographical biases in data coverage[25] (Extended Data Fig. 3b), and strictness of host status definition (Extended Data Fig. 4). The last of these is crucial to understanding disease risk, since species capable of being infected by a given pathogen may not contribute substantially to transmission dynamics or zoonotic spillover risk. We therefore repeated analyses for mammals only (the major reservoirs of zoonoses globally) with reservoir status strictly-defined as an association with at least one zoonotic agent (aetiologic agent of a specific human disease with a known animal reservoir), based on pathogen detection, isolation or confirmed reservoir status (143 host species, 2026 sites, 63 studies). Overall trends remain consistent, although with notably stronger effects on host proportion of total abundance (+42-52% in secondary and managed land), and weaker effects on host richness that may reflect

underlying variability in responses between mammal taxa (Extended Data Fig. 4).

To examine the possibility of such taxonomic variability in host responses to land use, we analysed mean land use effects on species-level occurrence and abundance of zoonotic host (strictly-defined) and non-host species, for several mammalian (Carnivora, Cetartiodactyla, Chiroptera, Primates, Rodentia) and avian orders (Passeriformes, Psittaciformes) that are well-sampled in PREDICTS and harbour the majority of known zoonoses (Methods). Within most orders, non-host species tend to decline more strongly in response to land disturbance than host species, but with substantial between-order variation in the direction and clarity of effects (Figure 3, Extended Data 5, Supp. Table 6). Notably, within passerine birds, bats and rodents, hosts and non-hosts show clear divergent responses to land use, with host species abundances on average increasing (+14-96% Passeriformes, +45% Chiroptera, +52% Rodentia) while non-host abundances decline (-28-43% Passeriformes, -13% Chiroptera, -53% Rodentia) in human-dominated relative to primary sites (Figure 3). Although such a tendency has been observed in some disease systems, our results suggest this is a more general phenomenon in these taxa, which may contribute to numerous documented links between anthropogenic ecosystems and bat-, rodent- and bird-borne emerging infections (e.g. corona-, henipa-, arena- and flaviviruses, *Borrelia* and *Leptospira* spp.)[16,17,20]. In contrast, primate and carnivore host responses are not clearly distinguishable from overall species declines in these orders, consistent with past studies showing no consistent links between land disturbance and disease in primates[24] and highlighting the importance of ecotonal or edge habitats as human-primate epidemiological interfaces[15] (although sparser urban sampling means that urban-adapted primates, such as macaques, are likely underrepresented).

The differing responses of host and non-host species may be mediated by covariance between traits influencing both host status and human-tolerance[27], but could also reflect histories of human-wildlife contact and coevolution of shared pathogens[12]. If the former is the case we hypothesise that harbouring a higher number of pathogens overall (richness of either zoonotic or non-zoonotic pathogens; a metric often correlated to species traits[28]), would be associated with more positive species responses to land use. We tested this across all mammals in our dataset (due to more complete pathogen data availability than for other taxa; 546 species, 1950 sites), here controlling for species-level differences in research effort by analysing residual pathogen richness not explained by publication effort (Methods, Extended Data Fig. 6). We find that pathogen richness is associated with increasing species probability of occurrence in managed sites but not in primary habitat, and that this result is consistent for

either human-shared or non-human-shared pathogens (no documented infection of either people or domestic animals; Extended Data Fig. 7, Supp. Table 7). This suggests that the net increase in zoonotic host diversity in disturbed sites is at least partly trait-mediated; in particular, species traits associated with a faster pace-of-life are often correlated both with reservoir status and infection outcomes[6,27] (potentially owing to life-history trade-offs between reproductive rate and immune investment[29]) and with population resilience to anthropogenic pressures[21]. A trait-mediated explanation is also supported by our finding that differential host and non-host species responses to land use are most clearly detected when comparing across large clades with a wide diversity of life-histories, such as rodents, passerines and, notably, mammals overall (Extended Data Fig. 5). In contrast, generally longer-lived, large-bodied clades (e.g. primates, carnivores) show more idiosyncratic or negative responses to landscape disturbance (Figure 3).

Overall, our results indicate that the homogenising impacts of land use on biodiversity globally[9] have produced systematic changes to local zoonotic host communities, which may be one factor underpinning links between human-disturbed ecosystems and disease emergence. By leveraging site-level survey data, our analyses reflect community changes at the epidemiologically-relevant local landscape scale[22], negating the need to ignore community interactions or generalise ecological processes to coarser spatial scales (a typical limitation of global studies that can confound or mask biodiversity-disease relationships[3]). Our results reflect potential zoonotic hazard, since proximity to reservoir hosts is not sufficient for spillover[30], and emergent disease risk will depend on contextual factors (e.g. pathogen prevalence, intermediate host/vector populations, landscape structure, socioeconomics) that may synergistically or antagonistically affect transmission dynamics and exposure rates[12]. Nonetheless, land use also predictably impacts other factors that can amplify within- and cross-species transmission[31] (e.g. resource provisioning[11], vector diversity[32]), and increases potential for human-wildlife contact[13]: for example, human populations are consistently higher at disturbed sites in our dataset (Extended Data Fig. 8). Global expansion of agricultural and urban land forecast for the coming decades, much of which is expected to occur in low-and middle-income countries with existing vulnerabilities to natural hazards[18], thus have the potential to create growing hazardous interfaces for zoonotic pathogen exposure. In particular, the large effect sizes but sparser data availability for urban ecosystems (especially for mammals; Extended Data Fig. 4) highlight a key knowledge gap for anticipating urbanisation effects on public health and biodiversity. Our findings strongly support calls to enhance proactive human and animal surveillance within

agricultural, pastoral and urbanising ecosystems[33,34], and highlight the need to consider disease-related health costs in land use and conservation planning.

References

1. Myers, S. S. *et al.* Human health impacts of ecosystem alteration. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18753–60 (2013).
2. Gottdenker, N. L., Streicker, D. G., Faust, C. L. & Carroll, C. R. Anthropogenic land use change and infectious diseases: a review of the evidence. *Ecohealth* **11**, 619–632 (2014).
3. Keesing, F. *et al.* Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).
4. Ostfeld, R. S. & LoGiudice, K. Community disassembly, biodiversity loss, and the erosion of an ecosystem service. *Ecology* **84**, 1421–1427 (2003).
5. Johnson, P. T. J. *et al.* Living fast and dying of infection: host life history drives interspecific variation in infection and disease risk. *Ecol. Lett.* **15**, 235–242 (2012).
6. Johnson, P. T. J., Preston, D. L., Hoverman, J. T. & Richgels, K. L. D. Biodiversity decreases disease through predictable changes in host community competence. *Nature* **494**, 230–233 (2013).
7. Newbold, T. *et al.* Global effects of land use on local terrestrial biodiversity. *Nature* **520**, 45–50 (2015).
8. Newbold, T. *et al.* Widespread winners and narrow-ranged losers: land use homogenises biodiversity in local assemblages worldwide. *PLoS Biol.* **16**, e2006841 (2018).
9. Faust, C. L. *et al.* Pathogen spillover during land conversion. *Ecol. Lett.* **21**, 471–483 (2018).
10. Becker, D. J., Streicker, D. G. & Altizer, S. Using host species traits to understand the consequences of resource provisioning for host–parasite interactions. *J. Anim. Ecol.* **87**, 511–525 (2018).
11. Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
12. Shah, H. A., Huxley, P., Elmes, J. & Murray, K. A. Agricultural land-uses consistently exacerbate infectious disease risks in Southeast Asia. *Nat. Commun.* **10**, 4299 (2019).
13. Gottdenker, N. L., Chaves, L. F., Calzada, J. E., Saldaña, A. & Carroll, C. R. Host life history strategy, species diversity, and habitat influence Trypanosoma cruzi vector infection in changing landscapes. *PLoS Negl. Trop. Dis.* **6**, 5–7 (2012).
14. Fornace, K. M. *et al.* Association between landscape factors and spatial patterns of

Plasmodium knowlesi infections in Sabah, Malaysia. *Emerg. Infect. Dis.* **22**, 3–10 (2016).

15. Pulliam, J. R. C. *et al.* Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface* **9**, 89–101 (2012).
16. Kilpatrick, A. M. Globalization, land use, and the invasion of West Nile virus. *Science (80-.).* **334**, 323–327 (2011).
17. Popp, A. *et al.* Land-use futures in the shared socio-economic pathways. *Glob. Environ. Chang.* **42**, 331–345 (2017).
18. Civitello, D. J. *et al.* Biodiversity inhibits parasites: broad evidence for the dilution effect. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8667–8671 (2015).
19. LoGiudice, K., Ostfeld, R. S., Schmidt, K. A. & Keesing, F. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 567–71 (2003).
20. Purvis, A., Gittleman, J. L., Cowlishaw, G. & Mace, G. M. Predicting extinction risk in declining species. *Proc. R. Soc. B Biol. Sci.* **267**, 1947–1952 (2000).
21. Johnson, P. T. J., Ostfeld, R. S. & Keesing, F. Frontiers in research on biodiversity and disease. *Ecol. Lett.* **18**, 1119–1133 (2015).
22. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
23. Young, H., Griffin, R. H., Wood, C. L. & Nunn, C. L. Does habitat disturbance increase infectious disease risk for primates? *Ecol. Lett.* **16**, 656–663 (2013).
24. Hudson, L. N. *et al.* The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecol. Evol.* **7**, 145–188 (2017).
25. Lloyd-Smith, J. O. *et al.* Epidemic Dynamics at the Human-Animal Interface. *Science (80-.).* **326**, 1362–1367 (2009).
26. Joseph, M. B., Mihaljevic, J. R., Orlofske, S. A. & Paull, S. H. Does life history mediate changing disease risk when communities disassemble? *Ecol. Lett.* **16**, 1405–1412 (2013).
27. Kamiya, T., O'Dwyer, K., Nakagawa, S. & Poulin, R. What determines species richness of parasitic organisms? A meta-analysis across animal, plant and fungal hosts. *Biol. Rev.* **89**, 123–134 (2014).
28. Lee, K. A., Wikelski, M., Robinson, W. D., Robinson, T. R. & Klasing, K. C.

303 Constitutive immune defences correlate with life-history variables in tropical birds. *J.*
304 *Anim. Ecol.* **77**, 356–363 (2008).
305 29. Rohr, J. R. *et al.* Towards common ground in the biodiversity–disease debate. *Nat.*
306 *Ecol. Evol.* **4**, 24–33 (2020).
307 30. Hosseini, P. *et al.* Does the impact of biodiversity differ between emerging and
308 endemic pathogens? The need to separate the concepts of hazard and risk. *Philos.*
309 *Trans. R. Soc. B Biol. Sci.* **372**, (2017).
310 31. Brearley, G. *et al.* Wildlife disease prevalence in human-modified landscapes. *Biol.*
311 *Rev.* **88**, 427–442 (2013).
312 32. Burkett-Cadena, N. D. & Vittor, A. Y. Deforestation and vector-borne disease: Forest
313 conversion favors important mosquito vectors of human pathogens. *Basic Appl. Ecol.*
314 **26**, 101–110 (2018).
315 33. Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and Disease
316 Emergence: Dynamics at the Wildlife–Livestock–Human Interface. *Trends Ecol. Evol.*
317 **32**, 55–67 (2016).
318 34. Holmes, E. C., Rambaut, A. & Andersen, K. G. Pandemics: Spend on surveillance, not
319 prediction. *Nature* **558**, 180–182 (2018).
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335

Figure legends

Figure 1: Combined ecological communities and zoonotic host species dataset. Map points show the geographical locations of surveyed assemblages (n=6801 sites), with mammal survey locations in black and all other sites in red, and countries containing sites shaded in blue. Inset chart shows the taxonomic distribution of hosts of human-shared pathogens (birds, invertebrates, mammals, reptiles and amphibians; see Methods). Boxplots and points show, for each study, host species richness as a percentage of the total per-study sampled richness, split across temperate and tropical biomes (n=184 studies; boxes show median and interquartile range, whiskers show values within 1.5*IQR from quartiles).

Figure 2: Effects of land use on site-level host species richness and total abundance. Points, wide and narrow error bars show modelled percentage difference in diversity metrics (posterior marginal median, 67% and 95% quantile ranges respectively, across 1000 bootstrap models) relative to a baseline of primary land under minimal use (dashed line) (n=6801 sites: primary (1423 and 1457 for minimal and substantial use, respectively), secondary (1044, 629), managed (565, 1314), urban (136, 233)). Models are of species richness (A) and total abundance (B) of host species and of all other (non-host) species, and of hosts as a proportion of total site-level richness and abundance (C-D). Point shape denotes land use intensity (minimal or substantial) and colour denotes host (brown) or non-host (green). All posterior estimates were calculated across an ensemble of 1000 bootstrapped models, each with a proportion of non-hosts probabilistically transitioned to host status (median 121, range 90–150; Extended Data Fig. 2) to account for variability in species-level research effort (Methods, Supp. Methods 1). Models also included fixed effects for human population density and random effects for study methods and biome (Methods). Parameter estimates represent averaged effect sizes across multiple studies with differing survey methods and taxonomic focus, so do not have an absolute numerical interpretation.

Figure 3: Effects of land use on species abundance of mammalian and avian zoonotic hosts and non-hosts. Points, wide and narrow error bars show average difference in species abundance (posterior median, 67% and 95% quantile ranges respectively, across 500 bootstrap models to account for host status uncertainty) in secondary (Sec.), managed and urban sites relative to a primary land baseline (dashed line). Differences are estimated across

all host (brown) and non-host (green) species in each mammalian or avian order. For mammals, zoonotic host status was defined strictly (direct pathogen detection, isolation or confirmed reservoir status), and urban sites were excluded owing to sparse urban sampling (only 2 studies; additionally, no non-host primates were recorded in managed land, and urban 95% quantile range for Psittaciformes is not shown due to high uncertainty). Abundance differences were predicted using a hurdle model-based approach to account for zero-inflation (combining separately-fitted occurrence and zero-truncated abundance models; see Extended Data Fig. 5, Methods). The inset table show per-order numbers of species in the dataset (between 8% and 35% of total described species in each order), known zoonotic hosts (prior to bootstrap), and sampled sites. Silhouettes are from PhyloPic (http://phylopic.org/).

Methods

We combined a global database of ecological assemblages (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems, PREDICTS)[25] with data on host-pathogen and host-parasite associations, to create a global, spatially-explicit dataset of local zoonotic host diversity. We define pathogens and parasites (henceforth *'pathogens'*) as including bacteria, viruses, protozoa, helminths and fungi (excluding ectoparasites). PREDICTS contains species records compiled from 666 published studies that sampled local biodiversity across land use type and intensity gradients, allowing global space-for-time analysis of land use effects on local species assemblages (i.e. comparison between sites with natural vegetation considered to be a baseline). We analysed relative differences in wildlife host community metrics (zoonotic host species richness and abundance) between undisturbed (primary) land and nearby sites under varying degrees of anthropogenic disturbance. We subsequently conducted further analyses to examine how host species responses to land use vary across different mammalian and avian orders, and to test whether mammal pathogen richness (including both human and non-human pathogens) covaries with tolerance to land use.

Datasets

Ecological community and land use data. Each of the >3.2 million records in PREDICTS is a per-species, per-site measure of either occurrence (including absences) or abundance, alongside metadata on site location, land use type and use intensity. The database provides as representative a sample as possible of local biodiversity responses to human pressure, containing 47,000 species in a taxonomic distribution broadly proportional to the numbers of described species in major terrestrial taxonomic groups[25]. We first pre-processed PREDICTS following previous studies[8]: records collected during multiple sampling events at one survey site (e.g. multiple transects) were combined into a single site record, and for studies whose methods were sensitive to sampling effort (e.g. area sampled), species abundances were adjusted to standardise sampling effort across all sites within each study, by assuming a linear relationship between sampling effort and recorded abundance measures (both following ref.[8]). Our analyses of species occurrence and richness are therefore based on discrete count data, whereas abundances are pseudo-continuous (counts adjusted for survey effort). Due to the multi-source structure of PREDICTS (multiple studies with differing methods and scope),

434 the absolute species richness and abundance measures are non-comparable between studies[25],
435 so our analyses necessarily measure relative differences across land use classes.
436
437 *Host-pathogen association data.* We compiled animal host-pathogen associations from
438 several source databases, to provide as comprehensive a dataset as possible of zoonotic host
439 species and their pathogens: the Enhanced Infectious Diseases (EID2) database[35]; the Global
440 Mammal Parasite Database v2.0 (GMPD2) which collates records of parasites of
441 cetartiodactyls, carnivores and primates[36]; Plourde *et al.*'s reservoir hosts database[37] ; Olival
442 *et al.*'s mammal-virus associations database[23]; and Han *et al.*'s rodent zoonotic reservoirs
443 database[38] augmented with pathogen data from the Global Infectious Disease and
444 Epidemiology Network (GIDEON) (Supp. Table 8). We harmonised species names across all
445 databases, excluding instances where either hosts or pathogens could not be classified to
446 species level. To prevent erroneous matches due to misspelling or taxonomic revision, all
447 host species synonyms were accessed from Catalogue Of Life using 'taxize' v.0.8.9[39].
448 Combined, the dataset contained 20,382 associations between 3883 animal host species and
449 5694 pathogen species.
450 Each source database applies different methods and taxonomic scope. EID2 defines
451 associations broadly, based on evidence of a cargo species being found in association with a
452 carrier (host) species, rather than strict evidence of a pathogenic relationship or reservoir
453 status[35]. The other 4 databases were developed using targeted searches of literature and/or
454 surveillance reports, focus mainly on mammals, and provide more specific information on
455 strength of evidence for host status (either serology, pathogen detection/isolation, and/or
456 evidence of acting as reservoir for cross-species transmission). We therefore harmonised
457 definitions of host-pathogen associations across the full combined database. Across all animal
458 taxa we broadly defined associations based on any documented evidence (cargo-carrier or
459 stronger, i.e. including all datasets). Additionally, for mammals only (due to more
460 comprehensive pathogen data availability), we were able to define two further tiers based on
461 progressively stronger evidence: firstly, serological or stronger evidence of infection, and
462 secondly, either direct pathogen detection, isolation or reservoir status. Across all pathogens,
463 we also harmonised definitions of zoonotic status. Each pathogen was classified as *human-*
464 *shared* if recorded as infecting humans within either one of the source host-pathogen
465 databases or an external human pathogens list collated from multiple sources (Supp. Table 8).
466 Because the source datasets contain some organisms that infect humans and animals rarely or
467 opportunistically, or that may not strictly be zoonotic (e.g. pathogens with an environmental

or anthroponotic reservoir), pathogens were also more specifically defined as *zoonotic agents* (aetiologic agent of a specific human disease with a known animal reservoir) if classed as such in GIDEON, Wertheim *et al.*'s Atlas of Human Infectious Diseases[40] or Taylor *et al.*'s human pathogens database[41].

Combined datasets of hosts and land use. We combined PREDICTS with the compiled host-pathogen database by matching records by species binomial, and each species record was given a binary classification of 'host' or 'non-host' of human-shared pathogens. We adopted a two-tiered definition of host status, to examine the impact of making more or less conservative assumptions about the likelihood of a species contributing to pathogen transmission dynamics and spillover to humans. Firstly, we defined host status broadly: as any species with an association with at least one human-shared pathogen (as defined above), which for mammals must be based on serological or stronger evidence of infection (henceforth referred to as the *'full dataset'*). 177 studies in PREDICTS contained host species matches (190 mammals, 146 birds, 1 reptile, 2 amphibians, 37 invertebrates, listed in Supp. Table 1). Secondly, since mammals are the predominant reservoirs of both endemic and emerging zoonotic infections due to their phylogenetic proximity to humans[42,43], we also defined mammal species as zoonotic reservoir hosts based on stricter criteria: an association with at least one zoonotic agent (as defined above) which must be based on direct pathogen detection, isolation or confirmed reservoir status (henceforth referred to as *'mammal reservoirs subset'*). Within PREDICTS, 63 studies contained host matches based on this narrower definition (143 mammal reservoir hosts; Extended Fig. Data 4, Supp. Table 1).

Prior to analysis, we filtered PREDICTS to include only studies that sampled taxa relevant to zoonotic transmission, since the full database includes many studies with a different taxonomic scope (e.g. plants or non-vector invertebrates)[25]. We retained all studies that sampled any mammal or bird species, as these groups are the main reservoir hosts of zoonoses. For all other taxa, given that zoonoses and their hosts occur globally, we made the more conservative assumption that studies with no sampled hosts represent false absences (i.e. resulting from study aims and methodology) rather than true absences (i.e. no hosts are present), and only included studies with at least one host match in one sampled site in community models. This resulted in a final dataset of 530,161 records from 6801 sites in 184 studies (full dataset) and 51,801 records from 2066 sites within 66 studies (mammal reservoirs dataset; including mammal studies only) (Figure 1). Some host records were of arthropod vectors, but as these are a small proportion of records (around 2%; Supp. Table 1)

we generically refer to all matched species as *'hosts'*. By matching on species binomial we assume that pathogens are equally likely to occur anywhere within their hosts' geographical range; evidence from terrestrial mammal orders suggests that this assumption is reasonable globally[44,45]. Although overlooking geographical variation in pathogen occurrence, pathogen geographical distributions are poorly understood and subject to change, making it difficult to define geographical constraints on host status.

We aggregated land use classes in PREDICTS to ensure a more even distribution of sampled sites. We assigned each survey site's land use type to one of four categories: primary vegetation, secondary vegetation, managed ecosystems (plantation forest, pasture and cropland) and urban. Land use intensity was assigned to either minimal, substantial (combining light and intense use), or cannot decide (the latter were excluded from models). Original use intensity definitions[8] reflect gradation of potential human impacts within land use types; for example urban sites range from minimal (villages, large managed green spaces) to high intensity (impervious with few green areas). Land use categories simplify complex landscape processes, so our aggregation might mask subtle differences in disturbance mode and intensity. However, although some local studies have found differences in zoonotic host abundance and pathogen prevalence between different management regimes[46], we had no *a priori* reason to hypothesise differences between managed ecosystem types globally. Study regions were categorised as temperate or tropical, following ref.[47].

Statistical analysis

Accounting for species-level differences in pathogen discovery effort. The probability of identifying zoonotic pathogens within a species is strongly influenced by effort, meaning that poorly-studied species in our data could be falsely classified as non-hosts. Since research effort might also positively correlate with species' abundance in anthropogenic landscapes, accounting for this uncertainty is crucial. In statistical models we therefore consider host status (and derived metrics such as host richness) to be an uncertain variable, by assuming that all known hosts in our dataset are true hosts (true positives), and that non-hosts comprise a mixture of true non-hosts and an unknown number of misclassified species. We propagate this uncertainty into all model estimates using a bootstrapping approach, in which each iteration transitions a proportion of non-host species to host status with a probability influenced by research effort and taxonomic group (with poorly-researched species in taxonomic orders known to host more zoonoses having the highest transition rates; Extended Data Fig. 2, Supp. Methods 1).

We estimate disease-related research effort using species publication counts extracted from the PubMed biomedical database (1950–2018) for every species within our dataset (n=7285; Extended Data Fig. 2c), following other studies in disease macroecology in which publication effort often explains much of the variation in response variables[23,48]. Across 100 randomly-sampled mammal species from PREDICTS, PubMed publication counts were highly correlated to those from Web of Science and Google Scholar (both Pearson $r = 0.93$), indicating robustness to choice of publications database. Using publication counts directly to index species misclassification probability is problematic, since the relationship between publication effort and host status is both nonlinear (e.g. due to positive feedback, where pathogen detection drives increasing research towards a species or taxon) and taxon-specific (e.g. because some taxa are more intensely targeted for surveillance). We therefore calculate a trait-free approximation of false classification probability for non-host species (detailed in Supp. Methods 1) by assuming, first, that a species' relative likelihood of being a zoonotic host is proportional to the number of known hosts in the same taxonomic order (i.e. a poorly-studied primate is more likely to be a zoonotic host than a poorly-studied moth), and second, that confidence in non-host status accrues and saturates with increasing publication effort (following the cumulative curve of publication effort for known hosts within the same order; Extended Data Fig. 2a-b). Therefore, under-researched mammals, followed by birds, have the highest estimated false classification probabilities, but with substantial variation among mammalian and avian orders (Extended Data Fig. 2d-e).

Since data constraints prevent direct observation of how host detections accrue with discovery effort, our trait-free approximation leverages current knowledge of the distribution of zoonotic hosts and publication effort across broad taxonomic groups, and thus might over- or underestimate absolute host potential in any particular species. For example, because species traits and research effort are autocorrelated, our assumption that all non-host species per taxonomic group are equally likely to host zoonoses may conservatively overestimate host potential in less-researched species: many ecological traits that make species more likely to be poorly-studied (e.g. lower population densities, smaller range sizes[49,50]) would often be expected to reduce their relative importance in multi-host pathogen systems[51]. Nonetheless, our approach is sufficient to address our study's main confounding factor, i.e. the potential for biased distribution of research across land use types and biomes globally.

Community models of host species richness and total abundance. All modelling was conducted using mixed-effects regression in a Bayesian inference framework (Integrated

570 Nested Laplace Approximation (INLA)[52]. We aggregated ecological communities data to
571 site-level by calculating the per-site species richness (number of species) and total abundance
572 (total number of sampled individuals, adjusted for survey effort) of host and non-host species.
573 Land use type and intensity were combined into a categorical variable with 8 factor levels
574 (type+intensity, for 4 types and 2 intensity levels). During model selection we considered
575 fixed effects for land use and log-transformed 2015 human population density extracted from
576 CIESIN (because synanthropic species diversity might respond to changes in human
577 population density independently of land use; Extended Data Fig. 8). All models included
578 random intercept for study to account for between-study variation, and we additionally
579 considered random intercepts for spatial block within study (to account for the local spatial
580 arrangement of sites), site ID (to account for overdispersion caused by site-level differences)[8]
581 and biome (as defined in PREDICTS).

582 We modelled the effects of land use on the richness and total abundance of host and
583 non-host species separately, using a Poisson likelihood (log-link) to model species richness
584 (discrete counts). Since abundance data were continuous following adjustment for survey
585 effort, we followed other PREDICTS studies[8] and modelled log-transformed abundance with
586 a Gaussian likelihood; log-transformation both reduces overdispersion and harmonises
587 interpretation of the fixed effects with the species richness models (i.e. both measure relative
588 changes in geometric mean diversity from primary land under minimal use). We also
589 modelled the effects of land use on host richness and abundance as a proportion of overall
590 site-level sampled species richness or abundance, by including log total species richness as an
591 offset in Poisson models, and log total abundance as a continuous fixed effect (effectively an
592 offset) in abundance models.

593 For each response variable we first selected among candidate model structures,
594 comparing all combinations of random effects with all fixed effects included, and
595 subsequently comparing all possible fixed effects combinations using the best-fitting random
596 effects structure. In all cases we selected among models using the Bayesian pointwise
597 diagnostic metric Watanabe-Akaike Information Criterion (WAIC)[53] (Supp. Table 3-4). The
598 final models were subsequently checked for fit and adherence to model assumptions,
599 including testing for spatial autocorrelation in residuals (Extended Data Fig. 9). We then
600 bootstrapped each final model for 1000 iterations to incorporate research effort. For each
601 iteration, each non-host species was randomly transitioned to host status as a Bernoulli trial
602 with success probability p equal to estimated false negative probability (as described above;
603 Supp. Methods 1, Extended Data Fig. 2), all community response variables were recalculated,

the model was fitted and 2500 samples were drawn from the approximated joint posterior distribution. We then calculated posterior marginal parameter estimates (median and quantile ranges) across all samples from the bootstrap ensemble (Figure 2, Supp. Table 5). Between 90 and 150 non-host species (median 121) were selected to transition per iteration, increasing the total number of hosts by 24–40% (median 32%; Extended Data Fig. 2e). Because study coverage is heterogeneous globally, we subjected the full model ensembles to random and geographical cross-validation (Extended Data Fig. 3). We also conducted the same modelling procedure using only the strictly-defined mammal reservoirs subset (Extended Data Fig. 4).

Species-level estimates of land use effects on mammalian and avian zoonotic hosts. Because aggregate community diversity metrics may mask important variation between taxonomic groups, we separately modelled the average effects of land use type on the occupancy and abundance of all hosts and non-hosts of zoonotic agents within five mammalian (Carnivora, Cetartiodactyla, Chiroptera, Primates, Rodentia) and two avian orders (Passeriformes, Psittaciformes). For mammals we defined zoonotic host status strictly (pathogen detection, isolation or confirmed reservoir status, as described above) and excluded urban sites due to sparse urban sampling for mammals in PREDICTS (only 2 studies). All models included an interaction term between land use type and zoonotic host status (host or non-host) and random intercepts for each species-study combination and for taxonomic family (to account for gross phylogenetic differences). We again accounted for variable research effort per species as described above, fitting 500 models per order, and calculating posterior marginal estimates across samples drawn from the whole ensemble (Supp. Table 6).

Abundance data were overdispersed and zero-inflated due to the high proportion of absence records (i.e. sites where species were not found despite being sampled for). We therefore used a hurdle model-based approach[54] to estimate effects of land use on abundance, by separately fitting occurrence models (presence-absence; binomial likelihood, logit-link) to the complete dataset for each mammalian order, and zero-truncated abundance models (ZTA, log-abundance with Gaussian likelihood) to the dataset with absences removed (Extended Data Fig. 5). Mean differences in abundance across land uses are then calculated as the product of the proportional differences in predicted occurrence probability and ZTA relative to primary land[54]. We used posterior samples from paired occurrence (transformed to probability scale) and ZTA models (transformed to linear scale) to calculate a distribution of hurdle predictions separately for each bootstrap iteration (i.e. with the same non-hosts reclassified). We then summarised predicted changes per land use type across samples from

638 the entire bootstrap ensemble (median and quantile ranges; Figure 3). Due to the complex
639 nested structure of PREDICTS, our hurdle predictions assume independence between
640 occurrence and ZTA processes[54], so do not formally account for the possibility of covariance
641 at random effects (species or family) level. For clarity, we therefore show the contributions of
642 each separate model for each order (Extended Data Fig. 5, Supp. Table 6). In most orders,
643 and when fitting models across all mammal species, land use often appears to act most
644 consistently on species occurrence, with more variable effects on ZTA, suggesting that the
645 independence assumption may be broadly reasonable at this global and cross-taxa scale.
646
647 *Relationship between pathogen richness and responses to land use across mammal species*.
648 Pathogen richness (the number of pathogens hosted by a species) is a widely-analysed trait in
649 disease macroecology, with both overall pathogen richness, shared pathogen richness (i.e.
650 number of pathogens shared between focal species) and zoonotic pathogen richness often
651 correlated to species traits such as intrinsic population density, life history strategy and
652 geographic range size[6,23,28,55]. If human-disturbed landscapes systematically select for species
653 trait profiles that facilitate host status, we might expect to observe positive responses to land
654 use in species with higher richness of either human-shared or non human-shared pathogens[24].
655 We tested this hypothesis for mammals, due to availability of much more comprehensive
656 pathogen data than for other taxa, by analysing the relationship between species pathogen
657 richness and probability of occurrence across three land use types (primary, secondary and
658 managed; urban sites excluded due to limited sampling).
659 Within the subset of PREDICTS studies that sampled for mammals, containing
660 26,569 records of 546 mammal species (1950 sites, 66 studies), we used the host-pathogen
661 association dataset to calculate, firstly, each mammal species' richness of human-shared
662 pathogens, and secondly its richness of pathogens with no evidence of infecting either
663 humans or domestic animals (*'non human-shared'*), defining associations based on
664 serological evidence or stronger. Of the 546 mammals, 190 species had at least one known
665 human-shared pathogen (human-shared pathogen richness mean 1.92, sd 6.07) and 96 species
666 had at least one non human-shared pathogen (non human-shared pathogen richness mean
667 0.81, sd 4.16). We account for research effort differently than in the binary host status models
668 above, since pathogen richness is a continuous variable that is influenced by magnitude of
669 effort (i.e. more effort would be expected to increase the number of detected pathogens;
670 Extended Data Fig. 6b-c). Therefore, we account for effort by estimating per-species residual
671 pathogen richness not explained by publication effort (i.e. the difference between observed

672 pathogen richness and expected pathogen richness given publication effort and taxonomic
673 group). To do this, we modelled the effect of publication effort on pathogen richness (discrete
674 counts) separately for human-shared and non human-shared pathogens, using a Poisson
675 likelihood with a continuous fixed effect of log-publications and random intercepts and
676 slopes for each mammalian Order and Family (to account for broad taxonomic differences in
677 host-pathogen ecology between orders[23]). We fitted the model to data from all mammal
678 species in our host-pathogen database (n=780) and predicted expected mean pathogen
679 richness for all mammals in PREDICTS. We calculated residuals from observed values for
680 these species (Extended Data Fig. 6), which we expect represent trait-mediated variation,
681 given the evidence that mammal pathogen richness covaries with species traits after
682 accounting for phylogeny and research effort[23].

683 We then modelled the relationship between residual pathogen richness (scaled to
684 mean 0, sd 1) and species probability of occurrence across land use types, separately for
685 human-shared and non-human-shared pathogens (Extended Data Fig. 7). Species occurrence
686 was modelled using a binomial (logit-link) likelihood, with fixed effects for the interaction
687 between residual pathogen richness and land use type, and random intercepts for species,
688 order, study and spatial block within study. As with prior analyses, models were checked for
689 fit and adherence to assumptions. Pathogen surveillance in animals is often focused on
690 species of zoonotic concern, meaning that pathogen inventories (especially of non-human-
691 shared pathogens) may be more complete for some taxonomic groups than others. We
692 therefore tested model sensitivity to separately fitting models containing, firstly, only species
693 from the four most comprehensively-sampled mammalian orders for parasites and pathogens
694 (Primates, Cetartiodactyla, Perissodactyla and Carnivora; the focal taxa of the Global
695 Mammal Parasite Database[36]), and secondly, species from all other mammal orders. We also
696 tested for sensitivity to uncertainty in the publications-pathogen richness relationship, by
697 separately fitting the land use model to 400 sets of residuals derived using posterior samples
698 from the fitted publication effort model (Extended Data Fig. 6g-h), and summarising
699 parameters across the full ensemble. Fixed effects directions and strength of evidence were
700 consistent across all models (Supp. Table 7). Data processing and analyses were conducted in
701 R v. 3.4.1[56], with model inference conducted in R-INLA[52].
702
703
704
705

End notes

Acknowledgements: The authors thank L. Enright, A. Etard, L. Franklinos, R. Freeman, R. Lowe and R. Pearson for discussion on previous versions of the manuscript. This research was supported by a University College London Graduate Research Scholarship (RG); the Ecosystem Services for Poverty Alleviation Programme (ESPA), Dynamic Drivers of Disease in Africa Consortium, NERC project no. NE-J001570-1 (DWR and KEJ); an MRC UKRI/Rutherford Fellowship (MR/R02491X/1) and Wellcome Trust Institutional Strategic Support Fund (204841/Z/16/Z) (both DWR); and a Royal Society University Research Fellowship (TN). CAD thanks the UK MRC and DFID for Centre funding (MR/R015600/1), and thanks the UK National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Modelling Methodology at Imperial College London in partnership with Public Health England (PHE) for funding (grant HPRU-2012–10080).

Author contributions: RG, DWR, KEJ and TN conceived and designed the study. CAD contributed to the design of statistical analyses. RG collated and processed the data, and led and conducted the analyses with DWR, KQC and TN. All authors contributed to writing the manuscript.

Competing interests: The authors declare no competing interests.

Supplementary Information is available for this paper.

Correspondence: Correspondence and requests for materials should be addressed to Kate E. Jones (kate.e.jones@ucl.ac.uk).

Data and code availability

All data and code for this study, where not freely available online, are archived at Figshare (doi: 10.6084/m9.figshare.7624289). Data sources are listed, with links to freely-available online sources, in Supp. Table 8.

References (additional)

35. Wardeh, M., Risley, C., McIntyre, M. K., Setzkorn, C. & Baylis, M. Database of host-pathogen and related species interactions, and their global distribution. *Sci. Data* **2**, 150049 (2015).
36. Stephens, P. R. *et al.* Global Mammal Parasite Database version 2.0. *Ecology* **98**, 1476 (2017).
37. Plourde, B. T. *et al.* Are disease reservoirs special? Taxonomic and life history characteristics. *PLoS One* **12**, 1–23 (2017).
38. Han, B. A., Schmidt, J. P., Bowden, S. E. & Drake, J. M. Rodent reservoirs of future zoonotic diseases. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7039–44 (2015).
39. Chamberlain, S. & Szocs, E. Taxize: taxonomic search and retrival in R. (2013).
40. Wertheim, H. F. ., Horby, P. & Woodall, J. . *Atlas of Human Infectious Diseases*. (UK: Wiley-Blackwell, 2012).
41. Taylor, L. H., Latham, S. M. & Woolhouse, M. Risk factors for human disease emergence. *Philos. Trans. R. Soc. B Biol. Sci.* **356**, 983–989 (2001).
42. Han, B. A., Kramer, A. M. & Drake, J. M. Global Patterns of Zoonotic Disease in Mammals. *Trends Parasitol.* **32**, 565–577 (2016).
43. Rottingen, J. *et al.* New vaccines against epidemic infectious diseases. *New Engl. J. Med.* **376**, 610–613 (2017).
44. Harris, N. C. & Dunn, R. R. Using host associations to predict spatial patterns in the species richness of the parasites of North American carnivores. *Ecol. Lett.* **13**, 1411–1418 (2010).
45. Cooper, N., Griffin, R., Franz, M., Omotayo, M. & Nunn, C. L. Phylogenetic host specificity and understanding parasite sharing in primates. *Ecol. Lett.* **15**, 1370–1377 (2012).
46. Young, H. *et al.* Interacting effects of land use and climate on rodent-borne pathogens in central Kenya. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160116 (2017).
47. Newbold, T. *et al.* Global patterns of terrestrial assemblage turnover within and among land uses. *Ecography (Cop.).* **39**, 1151–1163 (2016).
48. Cooper, N., Kamilar, J. M. & Nunn, C. L. Host longevity and parasite species richness in mammals. *PLoS One* **7**, e42190 (2012).
49. González-Suárez, M., Lucas, P. M. & Revilla, E. Biases in comparative analyses of

772 extinction risk: Mind the gap. *J. Anim. Ecol.* **81**, 1211–1222 (2012).
773 50. Ducatez, S. & Lefebvre, L. Patterns of research effort in birds. *PLoS One* **9**, (2014).
774 51. Paull, S. H. *et al.* From superspreaders to disease hotspots: Linking transmission across
775 hosts and space. *Front. Ecol. Environ.* **10**, 75–82 (2012).
776 52. Blangiardo, M., Cameletti, M., Baio, G. & Rue, H. Spatial and spatio-temporal models
777 with R-INLA. *Spat. Spatiotemporal. Epidemiol.* **7**, 33–49 (2013).
778 53. Hooten, M. B. & Hobbs, N. T. A guide to Bayesian model selection for ecologists.
779 *Ecol. Monogr.* **85**, 3–28 (2015).
780 54. Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A. & Smith, G. M. *Mixed effects*
781 *models and extensions in ecology with R*. (Springer, New York, NY, 2010).
782 55. Dallas, T. A. *et al.* Host traits associated with species roles in parasite sharing
783 networks. *Oikos* **128**, 23–32 (2018).
784 56. R Core Team. R: A Language & Environment for Statistical Computing. R Foundation
785 for Statistical Computing, Vienna, Austria. (2017).
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803

Extended Data

Extended Data Fig. 1: Conceptual framework for the effects of land use change on zoonotic disease transmission. Pathogen transmission between potential hosts is shown as black arrows. Land use change (green driver) acts on ecological community composition and human populations (white boxes), and on environmental features that influence contact and transmission both locally (light blue box) and at broader geographical scales (dark blue box). These processes occur within a broader socio-ecological system context also influenced by additional environmental (e.g. climatic), socioeconomic and demographic factors. Unpicking the relative influence of these different processes on disease outcomes is challenging in local disease system studies, where multiple processes may be acting on pathogen prevalence and transmission intensity. The aim of this analysis was therefore to specifically examine, at a global scale, the effects of land use change on the composition of the potential host community (excluding domestic species), denoted below by the red box.

Extended Data Fig. 2: Approximating research effort bias for non-host species within the PREDICTS dataset. For all non-host species, we approximated the likelihood of false classification given research effort (i.e. probability of being a host, but not detected), based on the distribution of publication effort across known zoonotic hosts within the same taxonomic order (Supp. Methods 1). Line graphs show, for several orders, the cumulative curve of publication counts for known zoonotic hosts (A; shown on log-scale), and approximated false classification probability, which declines and asymptotes with increasing levels of research effort (B) (line colours denote taxonomic order). Points and boxplots show the distribution of PubMed publications for all host and non-host species in PREDICTS (C; total n=6921), and false classification probabilities (used as bootstrap transition rates) for all non-host species per taxonomic class in PREDICTS (D; total n=3665), and per key mammalian and avian order (E; total n=2927) (bracketed numbers denote number of species per-group; boxes show median and interquartile range, whiskers show values within 1.5*IQR from quartile). Histogram shows the number of non-host species transitioned to host status for each of 1000 bootstrapped models of the full dataset (F; median 121, 95% quantile range 102–142).

836 **Extended Data Fig. 3: Random (study-level) and geographical cross-validation of**
837 **community models (full dataset).** We tested the sensitivity of fixed effects estimates to both
838 random and geographically-structured (biome-level) subsampling. For random tests we fitted
839 8 hold-out models, excluding all sites from 12.5% of studies at a time (mean 12.5% of total
840 sites excluded per model, range 4%-19%; results in A). For geographical tests we fitted 14
841 hold-out models, with each excluding all sites from one biome (mean 7% of sites excluded
842 per model, range 0.07%-32%; results in B). Points and error bars show posterior marginal
843 parameter distributions for each hold-out model (median and 95% quantile range, with colour
844 denoting hold-out group or biome), calculated across samples from 500 bootstrap iterations
845 per-model to account for variable research effort across species. Directionality and evidence
846 for fixed-effects estimates are robust to both tests, suggesting that our results are not driven
847 by data from any particular subset of studies or regions. Urban parameters are however the
848 most sensitive to exclusion of data, likely due to the relatively sparse representation of urban
849 vertebrate diversity in the PREDICTS database (17 studies in our full dataset).
850
851 **Extended Data Fig. 4: Effects of land use on site-level mammalian reservoir host species**
852 **richness and total abundance.** Points, wide and narrow error bars show differences in
853 diversity metrics from primary minimal use baseline (posterior marginal median, 67% and
854 95% quantile ranges respectively, across 1000 bootstrap models). Models are of species
855 richness (A) and total abundance (B) of reservoir host and all other (non-host) species, and of
856 hosts as a proportion of site-level richness (C) and total abundance (D). For managed and
857 urban sites, use intensities were combined to improve evenness of sampling (n=2026 sites
858 from 63 studies: primary (589 and 572 for minimal and substantial use respectively),
859 secondary (144, 257), managed (348) and urban (116)). Posterior estimates were calculated
860 across an ensemble of 1000 bootstrapped models (median 51, range 38–62 non-hosts
861 transitioned to host status, i.e. increasing host number by 28–46%) (Methods). Urban sites
862 results show the same trend as the full dataset (Figure 2), but are not visualised due to wide
863 uncertainty: 88.7% (-2.1, 252.3) proportion richness, 307% (78.8, 500.7) proportion
864 abundance (posterior median and 95% quantile range; see Supp. Table 4). Point shape
865 indicates use intensity (minimal, substantial or both combined) and colour indicates host
866 (brown) or non-host (green). Reservoir species are listed in Supp. Table 1 (mammal species
867 listed as ‘Detection/reservoir’ in the ‘Evidence of host status’ column).
868

Extended Data Fig. 5: Effects of land use on occurrence and zero-truncated abundance (abundance given presence) of mammalian and avian hosts and non-hosts of zoonotic agents. Each row of three plots shows the results of species-level modelling for each of 5 mammalian and 2 avian orders, and for mammals overall. Points, wide and narrow error bars show average difference in species occurrence probability (left column) and zero-truncated abundance (ZTA; middle column) (posterior median, 67% and 95% quantile ranges across 500 and 750 bootstrap iterations, for each order and all mammals respectively). Differences are shown in secondary (Sec), managed and urban sites relative to a primary land baseline (dashed line), across all host (brown) and non-host (green) species. Histograms show, for each taxonomic group, the distribution of host species counts across all bootstrap models (i.e. after reclassifying non-hosts) compared to current number of known hosts (red vertical line), and the total number of species included in models (brackets in plot title). Estimates from occupancy and ZTA models (Supp. Table 6) were combined, assuming independence of processes, to give the hurdle predictions in Figure 3. Mammal reservoir status was defined based on strict criteria (pathogen detection or isolation), and the full list of host species included in these estimates is provided in Supp. Table 1 (scored '1' in the' zoonotic agent host' column). Silhouettes are from PhyloPic (http://phylopic.org/).

Extended Data Fig. 6: Residual human-shared and non human-shared pathogen richness across mammals. Distribution of human and non human-shared pathogen richness (A) and relationship to publication counts (B-C) are shown for mammals in our host-pathogen association dataset (n=780 species; points represent species shaded by Order, associations defined on serological or stronger evidence). Observed versus fitted plots (D-E) show where observed deviates from expected pathogen richness given log-publications and taxonomic group (Poisson likelihood with random intercepts and slopes for Order and Family; slope estimates for log-publications are similar for both human and non human-shared pathogens, β of 0.298 and 0.248 respectively). We used the fitted models to predict expected pathogen richness for mammals in PREDICTS (n=546) and derived residuals from observed values (shown in F), which were used in land use models (Extended Data Fig. 7). Calculating per-species residual quantile ranges across 2500 posterior parameter samples shows that within-species residual variance is generally small relative to residual size (G-H, points and error-bars show posterior median, 67% and 95% intervals, scaled to unit variance), and land use model results are robust to including this uncertainty (Methods, Supp. Table 7).

Extended Data Fig. 7: Effects of land use on the relationship between mammal species pathogen richness and occurrence probability. Points and error bars show intercept (A-B) and slope parameters (C-D) of the relationship between residual pathogen richness (scaled to mean 0 and unit variance) and mammal species occurrence probability (on the log odds scale; median and 95% credible interval). Model was fitted to occurrence data for all mammals in the database (n=29,569 records of 546 species, 1950 sites, 66 studies). Intercept parameters represent the average occurrence probability of a species with residual pathogen richness of 0 (i.e. with average pathogen richness given research effort and taxonomy), and slope parameters represent the change in occurrence probability for one scaled unit (standard deviation) increase in residual pathogen richness (Extended Data Fig. 6g-h). Intercept and slope parameters for primary and secondary land measure the differences relative to managed land (i.e. delta-intercept or delta-slope; B, D). Plotted lines show these relationships on the probability scale (E-F), showing the median (black line), 67% (dark shading) and 95% (light shading) quantile range, based on 3000 samples from the joint posterior distribution. For both human-shared and non human-shared pathogens, there is a positive relationship between a species' residual pathogen richness and its probability of occurrence in human-managed land. For human-shared pathogens, the strength of this relationship (slope parameter) is significantly larger in managed sites than in both primary and secondary land, and for non human-shared pathogens significantly larger in managed than in primary land (D; slopes for primary land not significantly different from 0). Full model summaries and results of sensitivity analyses are in Supp. Table 7.

Extended Data Fig. 8: Differences in human population density between land use types, for all sites within the full dataset. Points and boxplots show the distributions of log-transformed human population density by land use type and intensity, across all sites included in community models (n=6801). Boxes show median and interquartile range with whiskers showing values within 1.5*IQR from quartile, and are coloured by land use type, and numbers denote the number of sites in each category. Human population density estimates were extracted from CIESIN Gridded Population of the World 4, for 2005, the median year of studies included in the dataset. Per-site log human density estimates were considered as fixed effects in community models of host diversity, since human-tolerant or synanthropic species might respond to human population change independently of land use (Methods).

Extended Data Fig. 9: Diagnostic plots for all community models (full dataset and mammal reservoirs subset). Species richness counts were modelled with a Poisson likelihood, and abundance (adjusted counts) were log-transformed and modelled with a Gaussian likelihood (see Methods). Plot titles refer to model response variables: species richness (SR), total abundance (Abundance), for hosts, non-hosts, and for hosts as a proportion of the community (Prop). Points in (A) show observed data against model-fitted values, and the red line shows the expectation if observed equals fitted (n=6801 for full SR; n=6093 for full Abundance; n=2026 for mammals SR; n=1963 for mammals Abundance). We also tested for spatial autocorrelation of residuals across all sites within each study, with histograms (B) showing the distribution of per-study Moran's *I* *p*-values (indicating significance of spatial autocorrelation among sites within that study) for each model (n=184 for full SR; n=164 for full Abundance; n=63 for mammals SR; n=60 for mammals Abundance). Numbers in brackets are the percentage of studies that contained significant spatial autocorrelation ($p<0.05$, shown as a red line). Overall, spatial autocorrelation was fairly low across the dataset (statistically significant in 14%-30% of studies, with maximum 26% for models with host metrics as response variables). Residuals and statistics were derived from a single fitted model including community mean false classification probability as a linear covariate to account for research effort (with known hosts given a false classification probability of 0), rather than the full bootstrap ensemble.