***Science Translational Medicine* Manuscript Template**

**General Instructions on using this template and submitting a manuscript to *Science Translational Medicine*:** Using this template will help to speed the processing of your paper and completing the manuscript’s record in our system. Our goal is to automatically identify each section of your manuscript so that we can accurately extract title, authors, abstract, etc. and also be able to enrich it by including reference links and an accurate layout.

Please use the actual template, which starts on page 2. When you are ready to submit, please delete the text on this cover page.

You can submit your paper at <https://cts.sciencemag.org>

Additional instructions are available at https://www.science.org/content/page/stm-instructions-research-articles-initial-submission

So that we can extract parts of your paper (even if you do not use this template), begin each section with the specific words listed below, some of which are followed by a colon. Do not use paragraph breaks in the title, author list, or abstract. The author list, corresponding author email(s), and affiliation(s) should be checked carefully because they will be published as listed in the manuscript.

**Title**: No more than 135 characters and spaces, lacking jargon and abbreviations where possible

**Title 1:** Quantifying nosocomial transmission of pathogenic bacteria in hospital settings using patient records and culture data.

**Title 2:** Hospital traffic and surveillance determine nosocomial transmission and detection of pathogenic bacteria in hospital settings in New York City.

**Authors:** Jaime Cascante Vega1, Rami Yaari1, Tal Robin1, Lingsheng Wen2, Jason Zucker2, Anne-Catrin Uhlemann2, Sen Pei3, Jeffrey Shaman1,3

**Affiliations:**

1. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA

2. Division of Infectious Diseases, Department of Medicine, Columbia University, College of Physicians and Surgeons, New York, NY, USA

3. Columbia Climate School, Columbia University, New York, NY, USA.

**Abstract:** 250 words or less.

**One Sentence Summary:** No more than 150 characters and spaces.

**Main Text:** Total number of words (text + captions + references) should be less than 10,000.

**References and Notes:** Only a single numbered list should be provided for all references cited in the main text and in the supplementary materials.

**Acknowledgments:** Split into general, Funding, Author contributions, Competing interests, and Data and materials availability, as described in the template below.

**Supplementary Materials:** Include a list, noting which references are only cited in the SM.

**Fig. #.** (Begin each figure caption with a label, “**Fig. 1.**”, for example, as a new paragraph.)

**Table #.** (Begin each table caption with a label “**Table 1.**”, for example, as a new paragraph.)

Please use the .docx format (all versions after Word 2007 for PC and Word 2011 for Mac) and include page numbers in your submitted file. We also encourage use of line numbers. Supplementary Materials (comprising Supplementary Materials and Methods, figures, and tables) should be in a separate file.

More specific formatting instructions are provided in the actual template, which follows.

Title: Insert your title here of not more than 135 characters including spaces

**Authors:** Jaime. Cascante Vega1, Rami. Yari1, Tal. Robin1, Lingsheng. Wen2, Jason. Zucker2, Anne-Catrin. Uhlemann2, Sen. Pei1,\*, Jeffrey. Shaman1,3,\*

**Affiliations:**

1Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University; New York, NY, 10032, USA.

2Division of Infectious Diseases, Department of Medicine, Columbia University, College of Physicians and Surgeons; New York, 10032, NY, USA.

3Columbia Climate School, Columbia University; New York, NY, 10032, USA.

\*Corresponding authors. Email: sp3449@cumc.columbia.edu, jls106@cumc.columbia.edu

**One Sentence Summary:** A brief summary of the main result of your paper (150 characters, including spaces) without excessive jargon.

**Abstract (255/250 words):** The abstract should be one paragraph and may not exceed 250 words. It should have the following structure: An opening sentence that sets the question that you address and is comprehensible to the general reader, background content specific to this study, results, and a concluding sentence. It should be one paragraph only.

Pathogenic bacteria are a major threat to patients' health in hospitals. In this work, we leverage electronic health records from a major New York City hospital system collected during 2020-2021 to support simulation-based inference of nosocomial transmission for eight pathogens. We develop an agent-based model informed by patient hospitalization records to simulate importation from the community, nosocomial transmission, and patient spontaneous decolonization of bacteria. The model is coupled with a Bayesian inference algorithm to estimate the likelihood of detection upon testing and nosocomial transmission rates. We evaluate parameter identifiability for this model-inference system and find that it is able to discriminate nosocomial transmission and effective sensitivity. We apply the framework to estimate both quantities for eight prevalent bacterial pathogens: *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, MSSA, MRSA, *Staphylococcus epidermidis*, *Enterococcus faecium* and *Enterococcus faecalis*. We find that nosocomial transmission for E. coli is negligible, and MSSA has a lower nosocomial transmission rate than MRSA. We show that while bacterial pathogens have different levels of importation rates, nosocomial transmission rates were similar suggesting similar levels of transmission for all, except *E. coli*. We also find detection is similar for all pathogens and proportional to their abundance. This work highlights how fine-scale patient data can support simulation-based inference of epidemiological properties of micro-organisms and how hospital traffic, patient contact and surveillance determine epidemiological features. Evaluation of the surveillance and transmission potential for different pathogens could ultimately support the development of in-hospital control measures that limit the spread of these pathogens as well as design of surveillance strategies.

**Main Text:**

**INTRODUCTION**

This should include introductory information that lays out the clinical problem addressed by the research and that explains other background necessary for understanding the study.

Antimicrobial resistance micro-organisms (AMRO) are a major threat to human health worldwide and has emerged as one of the leading public threats of the 21st century 1. An estimated 4.95 million deaths were associated with bacterial AMR in 2019 globally, and mortality caused by AMR is projected to reach 10 million by 2050 2. Hospital-acquired (nosocomial) infections by bacterial pathogens including those resistant to antibiotics are a major contributor to mortality, length of stay in hospital and health-care associated costs 2. Understanding the burden and spread of AMR pathogens, and micro-organisms in general, within hospital settings, is critical for effective control planning and design of testing/culture protocols. Quantification of these characteristics remains challenging due to limited observation of micro-organisms carriage, difficulty assessing interventions in real-world hospital settings and incomplete understanding of the underlying data generation processes 3–5.  
  
To circumvent these difficulties, mathematical models have been applied to study pathogen transmission in hospital settings, to quantify and understand the relative roles of different routes of transmission 4, and to characterize the condition of the hospital settings to sustain transmission of both resistant and sensitive strains 6. In the context of AMRO, theory has been used to understand the emergence of resistance and its interplay with community-acquired infections 7,8, to evaluate antibiotic treatment protocols 9, to assess control measures to reduce nosocomial transmission 6,10–13. More recently, models have been used in pair with empirical observations to assess the role of competition of different strains at the between-host level and the role of within-host microbiome pathogen interactions 14.  
  
Most existing modeling studies focus on general theoretical frameworks of AMR organisms 6–9,14–18 providing insights on possible mechanisms behind transmission, co-existence and other processes in the hospital settings. Process-based models need to be interrogated with data before being used as reliable tools for public health 19. Investigations using simulation-based inference 20 tools usually focused on a single pathogen of interest such as (MRSA) 10,13,21,22, *P. aeuruginosa* 4*, C. difficil* 23*,* and *Vancomycin-resistant enterococci* (VRE)24using carriage detection 10,13,22,24, intensive care unit (ICU) clinical culture data 4,22 or sequencing data 23. Studies on the epidemiological characteristics of multiple co-circulating organisms in a single hospital network supported by real-world data are absent.  
  
Understanding the transmission dynamics of healthcare-associated (HA) infections is challenging. The epidemiology and transmission HA infections differ from community-acquired (CA) infections in a number of important ways. First, hospital networks are open systems with significant variations in the daily and weekly admission rate ranging from 23% to 52% (See Figure 1A). Variations are explained by heterogeneous patient traffic at the building and ward scale (Results). In contrast, communities are relatively closed systems with individuals moving in or out at much lower rates. Second, as many bacterial species exist as commensals on the human hosts, detection of colonization is difficult and differential by body site. Indeed, patients with infections of the bloodstream or lower respiratory tract are more likely to be detected in hospitals than those colonized at other sites 7,25. Third, facilities within a single hospital system (e.g., infusion, pediatric, emergency wards, surgical, among others) may differ substantially in their control and surveillance of micro-organisms as well as in the hospital traffic features, a complication compounded by patient transfers between wards and hospital buildings (see SI Figure S1 for the patient transfer matrix at both spatial scales) 26,27. Fourth, a lack of observational data for communities of micro-organisms spreading among patients and the environment in the same hospital system often imposes further challenges in comparing its epidemiological features. Lastly, the emergence of strains with different genetic backgrounds that were once confined to hospital circulation such as CA-MRSA clone-USA300 in the United States with enhanced features compared to HA-MRSA makes epidemiology in communities and healthcare settings different 8,25.  
  
In this study, we use a process agent-based model (ABM) and patient clinical culture data for seven prevalent pathogenic bacterial species collected in a major New York City (NYC) hospital system to address these challenges. The ABM is informed by real-world patient movement in the hospital system and incorporates importation of micro-organisms from the community, nosocomial transmission, decolonization via host clearance and patient transfer across hospital wards. To account for the heterogeneity of testing frequency and micro-organism prevalence among facilities, we used clinical testing records to design a patient-level observational model that mimics the detection of micro-organisms in hospitals. We fix the importation rates by surveying the literature and estimate nosocomial transmission and the likelihood of detection given carriage upon testing. We couple the ABM with a Bayesian inference algorithm and explore the identifiability of the system against simulated data. We estimate the two epidemiological features for eight pathogenic bacteria, which cause substantial mortality associated with AMR worldwide 2. The ABM captures multiple sources of heterogeneity (e.g., patient length of stay, contact patterns, individual observational model, etc.) that otherwise cannot be represented by compartmental models. To simulate and quantify the transmission and detection of different micro-organisms using the same modeling-inference framework, we simplify the underlying biological details that are unique to each microparasitic infection. Our aim is, in consequence, to understand the general similarities and quantify epidemiological properties of communities of circulating bacteria among patients in hospitals. We show that while bacterial pathogens have different levels of importation rates, nosocomial transmission rates were similar suggesting similar levels of transmission for all. We also found the likelihood of detection is similar for all pathogens and its ordering is proportional to the reported abundance of bacteria.

**RESULTS**

***Empirical patterns of hospital traffic and heterogeneity of nosocomial infections burden***

In Figure 1A, we plotted weekly and monthly incidence using faded and solid lines respectively for E. coli (total positives cultures, n=2890), K. pneumoniae (n=1139), P. aeruginosa (n=809), MSSA (n=773), MRSA (n=486), S. epidermidis (n=694), E. faecalis (n=596) and E. faecium (n=263). We deduplicated multiple positive clinical culture results during a patient visit, and consistently deduplicated them in the observational model. During the study period, the daily number of hospitalized patients fluctuated between 1,000 and 2,500 with most of the variation explained by differences in the day of the week (See solid line for daily and dashed line for weekly in Figure 1B). Daily numbers of new admissions ranged between 100 and 1,000, including outpatients. During the first COVID-19 wave in New York City, the numbers of in-patients and admissions were generally lower; after June 2020, patient traffic was higher and relatively stationary. Ward size, defined as the average ward occupancy per day during the study period, was heterogeneous (SI Figure S2A) with the majority of wards experiencing an occupancy below 10 patients. However, a few wards (e.g., emergency rooms) could admit over 100 patients each day. Average ward size was 9 considering all the wards and 20 excluding wards with ward size equal to 1.

Subheadings should be brief, ideally less than 10 words. Subheadings should not end in a period. Your paper may have as many subheadings as are necessary.

**Subhead 2: All figures and tables cited in order**

All sections of the results should refer to a figure (for example, Fig. 1) or table (for example, Table 1), preferably a data display element that is part of the main text. All figures, figure panels and tables must be presented in order. For example, the description of panel A of figure 3 cannot come before that of panel B of figure 2. The supplementary figures (for example, fig. S1) and tables (table S1) must also be in presented in order.

You may include page breaks if you would like to embed the figures within the text instead of at the end of the paper.

**Subhead 3: All data presented first in the Results**

All data should be presented in the Results. No data should be presented for the first time in the Discussion. Data (such as from Western blots) should be appropriately quantified.

**Subhead 4: Formatting in-text reference callouts**

References should be cited in parentheses with an italic number (*1*). Multiple reference citations are separated by commas (*2*, *3*) or if a series, dashes (*4*-*6*). References are cited in order by where they first are called out, through the text, then the figure captions and tables or table captions, and then through the supplementary materials.

**Subhead 5: Handling of equations**

Equations can be included. Use MathType (recommended) or use the legacy equation editor in Word (Chose Insert > Insert Object > Word Equation). We do not recommend using the native equation editor. This can in some cases produce less reliable MathML, the online markup language we use, which may result in display errors. If you enter equations in simple LaTeX, check that they will convert accurately (Word 2007 and higher can convert simple LaTeX equations).

**DISCUSSION**

Include a Discussion that summarizes your conclusions and elaborates on their implications. There should be a paragraph outlining the limitations of your results and interpretation, as well as a discussion of the steps that need to be taken for the findings to be applied in the clinic. Please avoid claims of priority and avoid repeating the conclusions at the end.

**MATERIALS AND METHODS**

All descriptions of Materials and Methods should be included in the main paper. The Materials and Methods should be broken up into sections, each with a short subheading. Please include a study design paragraph at the start of the Materials and Methods and a statistical paragraph at the end of the Materials and Methods in the main text. If the Materials and Methods make the paper exceed the length limitations, less important sections of the Materials and Methods can be moved to the Supplementary Materials.

**List of Supplementary Materials**

Present a list of the Supplementary Materials in the following format.

Materials and Methods

Fig. S1 or Fig S1 to Sx for multiple supplementary figures

Table S1 or Tables S1 to Sx for multiple supplementary tables

Data file S1 or Data files S1 to Sx (Excel files)

Movie S1 or Movies S1 to Sx

References (*xx*–*xx*) (numbers for references only cited in SM)

References and Notes

1. There is only one reference list spanning the text, figure captions and Supplementary Materials. Do not include a second reference list in the Supplementary Materials section. References only cited in the Supplementary Materials section are not counted toward length guidelines.
2. Each reference should be on a separate line ending in a period. For a style guide, see <https://www.science.org/content/page/stm-instructions-research-articles-initial-submission#citation-style>.
3. Include all author names, full titles, and full page ranges: J. D. Crane, D. I. Ogborn, C. Cupido, S. Melov, Massage Therapy Attenuates Inflammatory Signaling After Muscle Damage. *Sci. Transl. Med.* **4**, 119–125 (2012).
4. You can use an automatically numbered list in Word.
5. Each reference should have a separate number.
6. Please do not combine references and explanatory notes under the same reference number.

**Acknowledgments:** Acknowledgments follow the references and notes but are not numbered. Start with text that acknowledges non-author contributions and then complete each of the sections below as separate paragraphs.

**Funding:** Include all funding sources here, including grant numbers, complete funding agency names, and recipient’s initials. Each funding source should be listed in a separate paragraph.

Examples:

National Institutes of Health grant U12AB123456 (PV, CHO)

National Institutes of Health grant R01AB123456 (PV, GS)

William K. Bowes Jr Foundation (PV)

German Research Foundation grant AB 1234/1-1

Office of Biological and Environmental Research of the U.S. Department of Energy Atmospheric System Research Program Interagency Agreement grant DE-SC0000001

National Institute of Health Research UK

UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund

**Author contributions:** Each author’s contribution(s) to the paper should be listed [we encourage you to follow the CRediT model]. Each CRediT role should have its own line, and there should not be any punctuation in the initials.

Examples:

Conceptualization: SBB, DLA, MPW

Methodology: HP, FTGS, CW, JRK, NJB, PRB, JLS, EH

Investigation: SBB, DLA, MPW, WCB

Visualization: SFB, MJM, JLS, EH

Funding acquisition: SJE, MJM, JLS, EH

Project administration: JLS, EH

Supervision: SJE, MJM, JLS, EH

Writing – original draft: SBB, DLA, WCB, JLS, EH

Writing – review & editing: SBB, DLA, PRB, JLS, EH

**Competing interests:** Include any financial interests for each author that could be perceived as being a conflict of interest (including but not limited to financial holdings, professional affiliations, advisory positions, and board memberships). Here, also include any awarded or filed patents pertaining to the results presented in the paper, stating patent filing number and title and which authors are co-inventors. When authors have no competing interests, this should also be declared (e.g., “Authors declare that they have no competing interests.”).

**Data and materials availability:** All data, code, and materials used in the analysis must be available in some form to any researcher for purposes of reproducing or extending the analysis. Include a note explaining any restrictions on materials, such as materials transfer agreements (MTAs). Note accession numbers to any data relating to the paper and deposited in a public database; include a brief description of the data set or model with the number. If all data are in the paper and supplementary materials, include the sentence “All data are available in the main text or the supplementary materials.”

**Figures**

[If possible, embed the figures within the Word file, either within results or after the acknowledgements, with each figure’s caption immediately below it. This will facilitate evaluation of the paper.]

**Fig. 1**. **Short title of the first figure.** The figure caption should begin with a title (an overall descriptive statement of the figure) followed by additional text. The caption should be placed immediately after each figure. The primary callout of each figure part is indicated with a bold capital letter enclosed in parentheses [e.g., (A)]. Additional callouts are indicated the same way, but without the bold format. If a paragraph in the main text begins with the name of a figure, write out “Figure” in full (e.g., <para>“Figure 1 shows….”).

**Fig. 2.** **Short title of the second figure.** Indicate figure parts with bold capital letters (**A**), (**B**) . If you prefer, you can place both the actual figures and captions logically through the text near where they are cited rather than at the end of the file (but not both). Please do not use text boxes to arrange figures. High-resolution (preferably editable PDF or Adobe Illustrator format) figure files will be requested following review.

**Table 1.** **Short title of the first table.** Start table captions with a title (short description of the table). Format tables using the Word Table commands and structures. Do not create tables using spaces or tab characters.