UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Modeling Environmental Exposure and Disease at the Scale of Microbes, Hospital Patients, and Geographic Regions

Permalink
https://escholarship.org/uc/item/5wz4t5pj

Author
Greenfield, Ben K.

Publication Date
2016

Peer reviewed|Thesis/dissertation

Modeling Environmental Exposure and Disease at the Scale of Microbes, Hospital Patients, and Geographic Regions

by

Benjamin Kweskin Greenfield

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Health Sciences

in the

Graduate Division of the

University of California, Berkeley

Committee in charge:

Professor Thomas E. McKone, Chair

Professor John R. Balmes

Professor Lee W. Riley

Spring 2016

Abstract

Modeling Environmental Exposure and Disease at the Scale of Microbes, Hospital Patients, and Geographic Regions

by

Benjamin Kweskin Greenfield

Doctor of Philosophy in Environmental Health Sciences

University of California, Berkeley

Professor Thomas E. McKone, Chair

This thesis presents the application of three mathematical models to problems linking environmental exposures to human health. The models differ in spatial and temporal analysis scale. The premise underlying this work is that reliable models follow from careful matching of model scale to the specific research question.

Chapter 1 models bacterial competition at a cellular scale, to study the factors that may result in environmental antimicrobial resistance. A simple analytical solution for the antibiotic minimum selection concentration (MSC) is developed. The MSC is the lowest environmental antibiotic concentration at which a resistant bacterial strain will outcompete a sensitive strain. The solution is formulated as the ratio between the MSC and the minimum inhibitory concentration (MIC), which is a widely available laboratory measurement of the antibiotic concentration at which the growth of a sensitive strain is inhibited. Model equations were fitted to published experimental growth rate competition results. The model fit varied among nine compound-taxa combinations examined, but predicted the experimentally observed MSC/MIC ratio well ($R^2 \geq$ 0.95). Sensitivity analysis indicated that the MSC was sensitive to the shape of the antibiotic versus growth dose–response for the sensitive strain and to the fitness difference between strains. Model findings suggest a benefit of future experimental studies characterizing bacterial competition at low antibiotic concentrations. Employing the model in combination with empirical antibiotic growth curve data, it may be possible to predict environmental antibiotic concentrations at which resistant strains will be selected for. This could be incorporated into risk assessment models, to identify high risk environments for dissemination of antibiotic resistance.

Chapter 2 describes a quantitative model of the relative importance of direct skin-to-skin contact versus indirect transfer via environmental textiles and surfaces for hospital pathogens. The model describes the rate of environmental transfer of pathogenic microbes between patients in a hospital setting. However, the model does not consider the likelihood of infection.

The model was applied to transmission of pathogens between patients residing in separate hospital rooms, via a health-care worker. Simulations were performed to examine the separate contribution of skin, textiles, and nonporous surfaces to the total pathogen number transmitted. The role of elimination (organism death) was considered by comparing literature elimination rates for six pathogens: *Acinetobacter baumannii*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Bordetella pertussis*, sudden acute respiratory syndrome coronavirus (SARS-CoV), and influenza A. Based on model results, all pathogens except influenza A exhibit a high rate of transmission in the model scenario, suggesting that transmission via health-care workers is a valid concern. With the exception of influenza A, there was overlap in literature elimination rates among the pathogens, resulting in similarly high predicted transmission. For all pathogens except SARS-CoV the relative importance for pathogen transmission was nonporous surfaces > textiles > skin, indicating the importance of environmental surfaces as a potential pathway for disease transmission. For SARS-CoV, the order was nonporous surfaces > skin > textiles, due to literature indicating low survival on textiles and porous surfaces. These results, combined with limited data on elimination, suggest a need to perform disease-specific studies on how elimination systematically differs between skin and surfaces. This model application at the scale of individual humans indicates that environmental surfaces are likely important for pathogen transmission in health care settings.

Chapter 3 describes multivariate and geostatistical modeling employed to perform a combined assessment of multiple stressors at a regional scale. The study evaluated a metric of environmental health hazard developed by the California Environmental Protection Agency. The metric, CalEnviroScreen, combines 19 indicators of environmental impact and socioeconomic stress, and is intended to be used to help allocate funding for greenhouse gas amelioration projects within the state of California. Principal component analysis was performed to obtain the predominant multivariate associations in the 19 indicators. The CalEnviroScreen metric was strongly associated with the first principal components, indicating that CalEnviroScreen effectively captures the prevailing gradients in hazard present in the underlying data. However, CalEnviroScreen was poorly associated with agricultural pesticide application, suggesting that hazard from agricultural chemical exposure may not be captured. The first principal components obtained from the environmental pollution measures and the socioeconomic stressor measures were both associated with the rate of hospital visits for several disease diagnoses with an environmental etiology. This suggests that the indicators employed for CalEnviroScreen are associated with the burden of disease. The association was stronger for socioeconomic stressors than for environmental pollutants. The results of this ecological health study suggest a hypothesis that, compared to environmental pollutant exposure, socioeconomic status more greatly impacts overall burden of disease.

Table of Contents

Acknowledgements

I thank Tom McKone, my PhD advisor, for your steadfast support throughout my time at Berkeley. I have appreciated your advising style, challenging me to grow and develop independently as a scientist, while consistently making time to be available when I needed mentoring. I will always remember fondly our many fun, wide ranging conversations. I thank my other dissertation committee members, Lee Riley and John Balmes, for your patient and thoughtful mentoring. I am especially grateful to Dr. Riley for adopting me into your lab research group so that I could further consider the linkages between environmental health and infectious disease, and for your engaging leadership in this vibrant community. To Dr. Balmes, I appreciated your encouragement, in addition to your feedback about the medical relevance of proposed research ideas. In the Environmental Health Sciences faculty, I especially thank Mark Nicas and Robert Spear for maintaining a continued interest in my personal growth, and for always being available to help me strategically in my PhD and professional development. I have appreciated the open door policy among Berkeley faculty, and have particularly appreciated the mentoring of Katherine Hammond, Charlotte Smith, and Marty Mulvihill. I thank Olivier Jolliet (University of Michigan) for spending time to mentor and train me in mathematical modeling, and for your boundless energy and positive enthusiasm. Finally, I thank Allison Luengen (University of San Francisco) and Jason Smith (California State University – East Bay) for taking the time to mentor me on how to be a better and more effective teacher.

The other students during my time at Berkeley have been a pleasure to interact with. EHS students are especially noteworthy for the complete absence of any drama, competition, or conflict. I have felt a strong kinship with many students, including Jenna Hua, Meiling Gao, Tomas Leon, Kat Navarro, Diane Gonzales, Kelsi Perttula, Nandini Parthasarathy, Paul Yousefi, Fraser Gaspar, Alberto Ortega, Marie Tysman, Kate Vavra-Musser, Tina Huang, Veronica Davé, Beverly Shen, Byron Hu, Qu Cheng, Erica Garcia, Swati Rayasam, Julia Varshavsky. Robert Snyder, Bret Stogren, Leah Rubin, Noah Kittner, Molly Davies, Vania Wang, Noriko Kusumi, and Jen Ames. Thank you all for being fun and supportive friends and colleagues.

I thank my friends and family in the "real world". To Ellen Spitalnik, Scott Pesetsky, Jennifer Pesetsky, Amy Franz, April Robinson, Erez Goren, Michelle Lent, Kaamil Bey Isles, Amy Kweskin, Andrew Rodman, Jamie Kass, Karen DiDominicis, Esther Letteney, Tasha Newman, Lester McKee, Lori Roniger, Vicki Goldstone, Scott Hopkins, Megan Hall, and Megan Rundel, thank you supporting me, for keeping me grounded in reality, and for reminding me that there is life beyond the academy. To my siblings Susan and Mark, I have appreciated our connection and I am inspired by your passionate pursuit of your craft and your leadership.

Finally, I thank my parents for your inspiration, your continued and steadfast support, your positive attitudes about my career path, and for making all of this possible. And Abby, I am incredibly grateful to have you in my life. Thank you for being so delightful, sharing my dream, and bringing me peace and joy.

Statement Regarding Collaborators, Data, Human Subjects, and Funding Source

Introduction

> *Now it would be very remarkable if any system existing in the real world could be* <u>*exactly*</u> *represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations.*

> -G. E. P. Box [1]

> *Modeling for me isn't about being beautiful but creating something interesting for people to look at and think about.*

> -K. Bax

This thesis examines the utility of simple mathematical models for current and emerging problems in environmental health science. This Introduction section begins with a brief overview of modeling and model application in exposure science, and then articulates the problem of integrating across disparate exposures. The remainder of the Introduction provides a conceptual overview of how models are employed in the three main chapters to address selection for antibiotic resistance, transmission of microbes, and integration of multiple exposures.

A. The needs of exposure science and the role of models

The recent landmark National Research Council report, *Exposure Science in the 21st Century: A Vision and a Strategy* [2] defines exposure science as data collection and analysis regarding "the nature of contact between receptors (such as people or ecosystems) and physical, chemical, or biologic stressors." Within the field of exposure science, the report articulates a need for research activities for assessing and mitigating unforeseen emerging environmental health threats. In particular, tools must be developed to predict and anticipate human exposures to these threats [2,3]. This requires ongoing effort and resource allocation towards methods development, including improved interpretation of spatiotemporal exposure data (e.g., via sensors) [2,4], development and application of both external and internal markers of exposure [2,5], and computational integration of the increasing amounts of exposure information via exposure modeling [6].

Modeling is an important methodology in exposure science, having traditionally played a complementary role with data collection. Models are defined as simplifications of reality "constructed to gain insights into select attributes of a particular physical, biological, economic, or social system [7,8]." At their core, models are depictions of the dominant processes and relationships governing a system.

Although there are many kinds of models, a useful distinction is conceptual versus computational models. Conceptual models are qualitative system descriptions that may include graphics, hypotheses, narrative information, and flow charts. Though not always explicitly stated or acknowledged, almost all scientific activity includes conceptual models. For example, the null hypothesis that there is no association between two variables is a highly simplified conceptual model. In exposure science, the exposome, the totality of all exposures encountered

by a human from conception onward, has been a useful conceptual model [9,10]. The exposome has helped to frame the range of stressors encountered, the need to measure indicators of these stressors, and the complimentary role of both external environmental measurements (the external exposome) and internal tissue measurements including biomarker responses (the internal exposome) [5]. Another useful conceptual model in exposure science is the exposure science ontology (ExO), which identifies the central components of exposure science as stressor, receptor, exposure event, and outcome [11]. Using human inhalation exposure to a chemical carcinogen as an example, the receptor (human) encounters the stressor (contaminant) in the exposure event (inhalation) resulting in an adverse outcome (cancer). In general, the ExO serves as a conceptual basis for integrating and comparing across exposure studies. In this thesis, the ExO serves as the underlying integration framework across the different studies and scales modeled (Figure 1).
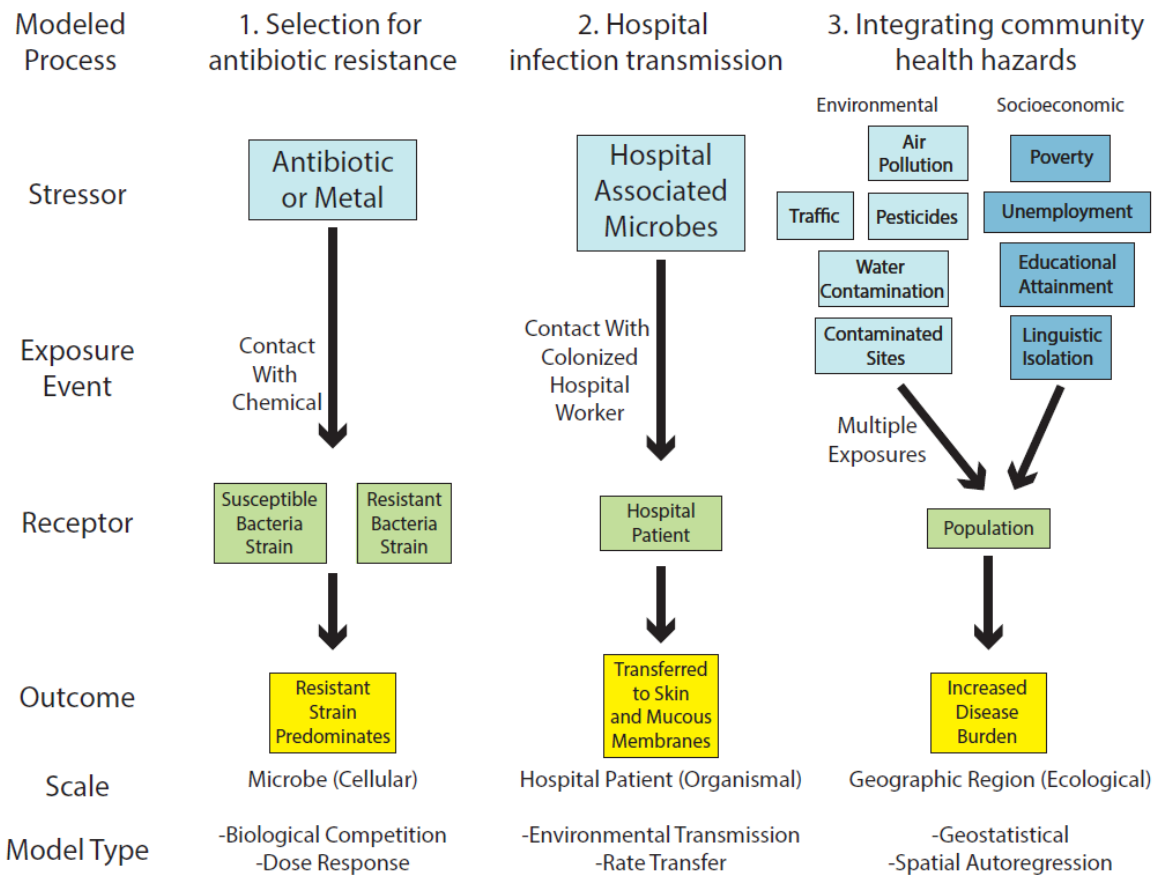

Figure 1. Conceptual model of the three thesis chapters. Each chapter develops and evaluates a computational model of a specific exposure problem. Although the models focus on scales ranging from microbial to ecological, each model fits within the exposure science ontology [11] in depicting an exposure event linking a stressor to a receptor, resulting in an adverse outcome.

Computational models describe a system mathematically, typically employing computers to make quantitative predictions [7,8]. Computational models may be further separated into mechanistic models, which describe underlying environmental relationships and processes via equations, versus statistical models, which fit empirical data to mathematical functions without construing an underlying mechanism [8,12]. Computational models have been developed and used extensively in exposure sciences to characterize environmental exposures to humans, including exposure to chemical contaminants [13–17] and infectious diseases [18,19]. Computational models are also useful for developing integrative assessments that compare outcomes across different systems or pathways within a single system. For example, modeling frameworks have been developed to integrate and interpret the increasing volume of available human tissue biomonitoring exposure data, enabling prioritization based on exposure dose and health hazard [6,20]. One such framework, the ExpoCast model, combines pharmacokinetic-pharmacodynamic models with statistical methods to screen and compare large numbers of compounds based on their tissue concentrations [6].

All models, by definition, are simplifications, attempting to extract the salient components of a system or process [7,8]. Regarding computational models, Naomi Oreskes and others [12,21,22] further caution that environmental models can never be verified because they always contain numerous uncertainties, never perfectly describe all processes in the "real world", and may correspond to environmental observations due to coincidence alone. As a result, rather than "truth generating machines" [7], models in exposure science should be viewed as tools suitable for specific purposes. Models can be employed to help scientists to integrate existing knowledge, interpret linkages between processes, evaluate and forecast possible effects of perturbations, identify and prioritize information gaps, and communicate concepts [7,21]. Models can be especially valuable to evaluate and rank hypothesized system-level effects, and ultimately contrast the anticipated impacts of specific interventions [2].

Because models are always simplified and limited depictions of reality, model evaluation is an ongoing need [7,17,21,22]. Model evaluation entails multiple activities: comparing model predictions to observed data (corroboration), determining how the model responds to changes in different input values (sensitivity analysis), and systematically examining how uncertainty regarding the model parameters or underlying structure affects model output reliability (uncertainty analysis) [8,17]. These model evaluation methods are the subject of considerable study and attention [23–25], all aimed at determining whether a model is suitable for its intended purpose. This thesis demonstrates the application and critical evaluation of three models for emerging problems in exposure assessment. Three processes are modeled: 1. the environmental development of antibiotic resistance; 2. transmission of hospital infections; and 3. integrating multiple health hazards (Figure 1).

B. The challenge of combining exposures

Over its approximately 100 year history, exposure science has exhibited a broadening focus towards integration, which sets the stage for the present work. In the early 20th century, exposure scientists were focused on occupational health, and studied individual stressors or a

small subset of stressors in a given work environment [10]. Similarly, classical epidemiological methods emphasized bivariate relationships between individual stressors and outcomes [26]. Driven by regulation and need, the field expanded to consider air pollutants in the 1950s to 1970s [10]. By the 1980s, the establishment of multimedia fate and transport models, combined with increasing recognition of widespread contamination of the ambient environment, shifted attention to the fate and transport of multiple chemical contaminants, and resulting combined human exposure to those compounds [10,17]. Since the 1990s, epidemiologists, exposure scientists, and risk analysts have increasingly acknowledged the importance of examining mixture effects, and the web of multiple causal factors, in determining health outcomes [9,26–32]. The current widespread availability of information integration via technological advances including cloud computing, high frequency measurements, and nontargeted chemical and biological analyses has built momentum behind the paradigm that integration of multiple exposures is now effectively a requirement in exposure science [2,10,26].

This need for cumulative exposure and risk assessment, including and integrating complex and multiple exposures, is now widely acknowledged, as are the limitations in previous methods and policies to address cumulative risk [32–34]. Relevant conceptual frameworks for health impact of multiple stressors include how environmental exposures intersect with socioeconomic and demographic conditions [28], how individual vulnerability modifies health response to environmental stressors and exposures [29], and how to integrate across the wide multitude of exposures encountered [31]. Consequently, Ted Schettler has articulated the ethical need in medicine to broaden health impacts and benefits assessment beyond frank disease treatment to prevention, community health, and even the ecological health of the biosphere [30]. The proponents of the exposome, the totality of all human external and internal exposures over a lifetime, have argued for a research emphasis on internal measurements of blood and other tissues, in order to simultaneously and efficiently examine the combined effect of all exposures [9,10,26]. Similarly, the idea of environment-wide association studies, which evaluate data across hundreds of different exposure biomarkers or measurements, to identify those most associated with a health outcome, has recently gained favor [31].

With the advent of geographic information systems (GIS), multiple exposures have also been integrated geospatially. Researchers examine and integrate geographic data on exposure hazards in order to identify at-risk populations and regions that warrant greater attention for management [34–38]. In epidemiology, there has been a notable increase in spatial questions and methods after 2000, particularly focusing on health disparities and access to resources [39]. Among statistical modeling techniques, multivariable spatial regression has been important for evaluating hypotheses regarding health predictors in the context of spatial dependence [39].

Despite the widespread trend of increased consideration of combined exposures, there is certainly no consensus regarding methodologies for how to synthesize the diverse drivers of health outcomes, especially on how to evaluate them quantitatively. Given the widespread availability of data on environmental concentrations, biomarkers, and health outcomes, how to integrate and interpret these information sources warrants attention in environmental health

4

research. To confront these current frontiers of exposure and risk, this thesis applies and evaluates statistical and mechanistic models that address exposures of emerging concern.

C. Exposure model development and evaluation at multiple scales

The central theme of this thesis is the use of parsimonious computational models to gain insights on the causes and mitigation of environmental exposures. Each chapter develops and evaluates a mechanistic or statistical model that is scaled to an exposure process of contemporary concern in environmental health (Figure 1). The model performance evaluations include a combination of illustrative simulations and comparisons with existing conceptual models and data. Guiding this work is the prevailing viewpoint [7,8,12,17] that: 1. a model is a simplification of reality that serves as a tool to extract and illustrate the key components of a process; 2. evaluation of model behavior provides insight on drivers underlying that process; and 3. the limitations of a model inform potential future research directions. Understanding model limitations requires particular attention on the relationship between models and data. This work therefore also examines how data uncertainty or variability constrains a model's ability to forecast, and considers what is learned when there is a mismatch between a computational model consistent with current theory and environmental measurements. The underlying principle is that a model's utility is not contingent on how well the model depicts empirical observations, but rather on what knowledge is gained in the evaluation and comparison itself [7,21,22].

The models themselves address environmental health questions occurring at three different scales (Figure 1):

1. **Microbes**[a]: When does antibiotic pollution lead to antibiotic resistance?
2. **Human individuals**: What hospital-acquired infections are transmitted primarily by environmental surfaces?
3. **Geographic regions**[b]: What is the relationship between environmental and socioeconomic hazards and chronic disease burden?

D. Three models in three chapters

Chapter 1 of the thesis develops and evaluates a model that provides insight on a key uncertainty in the onset of environmental antibiotic resistance. Anthropogenic antibiotic resistance exhibits high impacts for human disease burden nationally and globally [40]. There is a great need for models to aid in hazard and risk assessment of antibiotic resistance development, in order to determine when and where increased resistance will occur [41,42]. Chapter 1 develops a mathematical model for the minimum selective concentration (MSC), the antibiotic concentration at which antibiotic resistant bacteria would outcompete susceptible bacteria [43,44]. In terms of the exposure ontology framework outlined above [11], the stressor here is the chemical antibiotic, the receptors are sensitive and resistant strains of bacteria in

---

[a] I.e., the scale of unicellular organisms

[b] Ecological scales, in the epidemiological sense of comparing geographically separated groups of people, in this case residing in different census tracts or zip code areas

competition, the exposure event is bacterial contact with the antibiotic, and the outcome is the competitive dominance of the resistant strain over the sensitive strain (Figure 1). Effectively, Chapter 1 develops a mathematical model for the antibiotic dose–response relationship for resistance development. This model, combined with empirical antibiotic growth curve and susceptibility data, will aid in understanding environmental antibiotic concentrations at which resistant strains will be preferentially selected.

Much of the impact antibiotic resistant infections occurs in health care settings, with both exposure mechanism and disease burden varying among different microbes [40]. Chapter 2 employs a conceptual model to classify pathogens based on their environmental persistence and transmission. This conceptual model is mathematically formulated as a mechanistic model describing the transmission of microbes implicated in hospital-acquired infections. The model is a highly simplified but parsimonious set of mass-balance equations based on previously developed models for individual pathogens [19,45]. The model is parameterized with selected literature data on microbe environmental persistence, and is evaluated in a simplified scenario describing the contact and transmission between a health-care worker and two patients who reside in separate hospital rooms.

Part of the significance of this chapter lies in connecting the traditionally separate disciplines of exposure assessment and infectious disease modeling [46] to establish a quantitative framework for how different pathogens may respond to interventions. In particular, the ability to classify pathogens in terms of the likely merit of surface decontamination versus barrier controls is considered. Applying the exposure ontology to Chapter 2, the stressor is the pathogenic microbe, the receptor is the uncolonized patient, the exposure event is contact with the health-care worker and contaminated surface environments, and the outcome is the number of microbes that colonize the skin and mucous membranes of the previously uncolonized patient (Figure 1).

Chapter 3 evaluates the use of publicly available geospatial data to describe the relative impact of different components of the Eco-Exposome, which is the summation of external exposures that can influence human health [2]. The topic is of great importance for environmental justice because the most vulnerable human populations encounter high exposure to multiple stressors, requiring priority setting for interventions. In the context of the exposure ontology, this study focuses on the integration of multiple stressors, and their relative importance for the outcome of differences in burden of disease among different human populations (Figure 1). Chapter 3 does not mechanistically describe any exposure events, but rather employs statistical models to examine the relative importance of exposure to environmental versus socioeconomic stressors. This statistical modeling evaluates CalEnviroScreen, a simple mathematical model developed by California EPA to rank hazards. Statistical associations are examined and the results used to generate future hypotheses.

Chapter 1. Modeling the emergence of antibiotic resistance in the environment: an analytical solution for the minimum selection concentration[3]

Abstract

Environmental antibiotic risk management requires an understanding of how subinhibitory antibiotic concentrations contribute to the spread of resistance. We develop a simple model of competition between sensitive and resistant bacterial strains to predict the minimum selection concentration (MSC), the lowest level of antibiotic at which resistant bacteria are selected. We present an analytical solution for the MSC based on the routinely measured minimum inhibitory concentration (MIC) and the selection coefficient (sc) that expresses fitness differences between strains. We calibrated the model by optimizing the shape of the bacterial growth dose–response curve to antibiotic or metal exposure (the Hill coefficient, κ) to fit previously published experimental growth rate difference data. The model fit varied among nine compound-taxa combinations examined, but predicted the experimentally observed MSC/MIC ratio well ($R^2 \geq 0.95$). The shape of the antibiotic response curve varied among compounds ($0.7 \leq \kappa \leq 10.5$), with the steepest curve for the aminoglycosides streptomycin and kanamycin. The model was sensitive to this antibiotic response curve shape and to the sc, indicating the importance of fitness differences between strains for determining the MSC. The MSC can be more than one order of magnitude lower than the MIC, typically by a factor sc^{κ}. This study provides an initial quantitative depiction and a framework for a research agenda to examine the growing evidence of selection for resistant bacteria communities at low environmental antibiotic concentrations.

Introduction

Effective management of antibiotic risks in the environment requires an understanding of the factors responsible for the emergence, transmission, and maintenance of antibiotic resistance [42]. The hypothesized connection between antibiotic use in food animals and human health is supported by field studies, reports of farmers exposed to antibiotic-resistant bacteria from food-animals, ecological and temporal associations, and food-borne outbreaks [47,48]. However, insights are also needed into the extent to which antibiotics in the water environment contribute to the spread of resistance, and to the long-term prevalence of resistant infections in humans [49,50]. It is particularly important to address the question of when resistant bacteria predominate as a result of environmental antibiotic pollution [42,49–52].

The mutant selection window (MSW) paradigm states that resistant mutants may develop between the lowest boundary concentration of selection for resistance, and the upper boundary concentration of growth inhibition of the most resistant potential mutant (the mutant prevention concentration, MPC) [53,54]. The paradigm further indicates that the lower boundary concentration of the MSW is the minimum concentration that inhibits colony formation (MIC, ng ml^{-1}), and the MIC has been useful to evaluate hazard of selection for resistance in natural aquatic environments [52]. Considerable research *in vitro* and *in vivo* has demonstrated that

resistant mutants develop between the MIC and the MPC [54–56], but many laboratory and theoretical studies indicate that resistant mutants can also be preferentially selected above the minimum selective concentration (MSC, ng ml^{-1}), defined as the lowest concentration at which a resistant strain outcompetes and displaces sensitive isolates [42,44,57–63]. Because the MSC can be lower than the MIC, and to minimize the hazard of resistance occurring in the natural environment (e.g., aquatic systems), further characterization and understanding of the MIC versus MSC relationship would be beneficial [64].

Recent laboratory experiments [44,57,58,61] have elegantly demonstrated MSCs ranging from 1/4 to below 1/200 of the MIC for antibiotics of several classes (e.g., macrolide, aminoglycoside, fluoroquinolone, and antifolate) and for two metals in *Escherichia coli* (*E. coli*) or *Salmonella enterica* serovar Typhimurium LT2 (*S.* Typhimurium). This finding may help explain the high levels of resistance found in the environment, particularly at subinhibitory antibiotic concentrations [44,49,50]. These studies further indicate that the fitness cost of the resistance-conferring mutations is more important than differences in MIC between strains for discerning how much below the MIC the resistant bacteria will predominate [57]. There is, however, a need to mathematically describe the competition between strains in order to better understand and generalize strain- and antibiotic-specific results to a wider range of situations.

Mechanistic mathematical models, including experimentally validated pharmacodynamic/ pharmacokinetic models, describe antibiotic effects better than simple MIC measurements [65–67]. For example, the shape of the antibiotic dose–response curve is very important to the microbiological efficacy of antibiotic treatment regimens at high (treatment) levels [65]. The implications of this understanding of dose–response curve shape for low (subinhibitory) antibiotic levels and for calculation of the MSC, while relevant for selection of resistance, have not been considered in as much depth. To complement the recent empirical research [42,44,57,58], there remains a need for a quantitative model describing the MSC, i.e., the minimum environmental antibiotic concentration that allows resistant bacterial strain to dominate. Such a model can generate testable predictions, identify the factors that determine water or soil antibiotic concentrations that select for resistance, and be incorporated into risk assessments of antibiotic resistance development [42].

An analytical solution for the MSC has two potential uses. First, model sensitivity analysis and examination of parameter structure may provide insight on the relationship between commonly considered bacterial growth and antibiotic dose–response parameters and the MSC itself. Second, current methodology to accurately measure the MSC requires direct measurement of competition between bacterial strains and specialized methods such as fluorescent cell tagging and flow cytometry [57,61]. An analytical solution provides a potential alternative to these labor-intensive methods, instead estimating the MSC based on bacterial growth rate and antibiotic dose–response parameters that are routinely obtained within microbiology laboratories. To that end this paper answers three questions: 1. How do we quantitatively define the minimum selective concentration (MSC) via a simple mathematical model in combination with readily available measurements? 2. How well does such a mathematical model of MSC fit to published empirical data? 3. What model parameters, representing biological characteristics, are most important to describe the MSC?

The model we propose in this paper describes the MSC based on the competition between a wild-type and a resistant strain of bacteria, and the key factors that favor the growth of resistant strains at subinhibitory antibiotic concentrations. The model focuses on conspecific gram negative bacteria (GNB), and is calibrated to the recent experimental results of Gullberg et al. [57,61] for *E. coli* and *S.* Typhimurium. The model illustrates the shape of the antibiotic dose–response curve as a measurable and influential driver on the ratio of the MSC and MIC, and presents a hypothesized dose–response relationship for use in risk assessment of resistance development in environmental settings. Finally, we discuss the implications of the MSC results for increased risk of antibiotic resistance selection at antibiotic concentrations observed in antibiotic-contaminated waste streams and natural waters.

Theory

We develop a simple analytical expression of the ratio between the MSC and the MIC for a sensitive strain (i.e., MSC/MIC), which mathematically describes the factors that determine risks of subclinical antibiotic concentrations [44,57]. The model is based on the competition between two bacterial strains: a wild type sensitive strain, and mutant strain that is more resistant.

Model derivation for net growth rate

At a given antibiotic concentration a [ng ml^{-1}], the net growth rate ($N(a)$, [h^{-1}]) for each strain is given by

$$N_s(a) = N_{int,s} - D_{ab,s}(a) = R_{int} - D_{int} - D_{ab,s}(a) \tag{1}$$

$$N_r(a) = N_{int,r} - D_{ab,r}(a) = R_{int} + \sigma - D_{int} - D_{ab,r}(a) \tag{2}$$

where subscript s = sensitive bacteria, subscript r = resistant bacteria, $N_{int,s} = (R_{int} - D_{int}) =$ intrinsic net growth rate in the absence of antibiotic [h^{-1}], $R_{int} =$ intrinsic growth rate [h^{-1}], D_{int} = loss due to mortality (or, in continuous cultures, dilution) [h^{-1}], $D_{ab}(a)$ = loss in net growth [h^{-1}] due to a given antibiotic concentration, a, and σ = selection coefficient [h^{-1}].

The selection coefficient (σ) represents the fitness cost of resistance-conferring genes as the absolute difference in net growth rate between bacteria strains (e.g., sensitive vs. resistant) in the absence of antibiotics (i.e., $N_{int,r} = N_{int,s} + \sigma$). Resistance-conferring mutations exhibit variable fitness costs in comparison to sensitive strains, ranging from no detectable difference to half the wild-type growth rate in competition assays [68–70], with compensatory mutations often reducing or reversing the fitness cost of resistance mechanisms [69,71,72]. Accurate measurement of selection (σ) is difficult, requiring competition experiments employing labeled strains and flow cytometry [57,61]. For the purposes of this model, we run simulations on the assumption that resistance-conferring mutations engender a fitness cost, resulting in lower growth rates relative to less resistant strains, i.e., σ < 0 in Eq. 2.

The loss in net growth due to antibiotics can be described by a generalized Hill Equation [65,73–75]:

9

$$D_{ab}(a) = k_{max} \frac{a^{\kappa}}{a^{\kappa} + (EC_{50})^{\kappa}} \tag{3}$$

in which k_{max} [h^{-1}] is the maximum death rate due to antibiotic, EC_{50} [ng ml^{-1}] is the antibiotic concentration that achieves half of this maximum rate, and will thus increase with increased resistance, and κ is the Hill coefficient, which gives an indication of how steeply D_{ab} increases near the MIC [74]. For $\kappa = 1$ in the range of antibiotic concentrations below the MIC, the death rate increases roughly linearly. For a given strain, antibiotics with a high κ value (> 1) will have lower efficacy at sub-therapeutic levels, but higher efficacy at therapeutic levels above the MIC. The opposite relation is true for antibiotics with low κ values [65] as illustrated in Figure 2A.

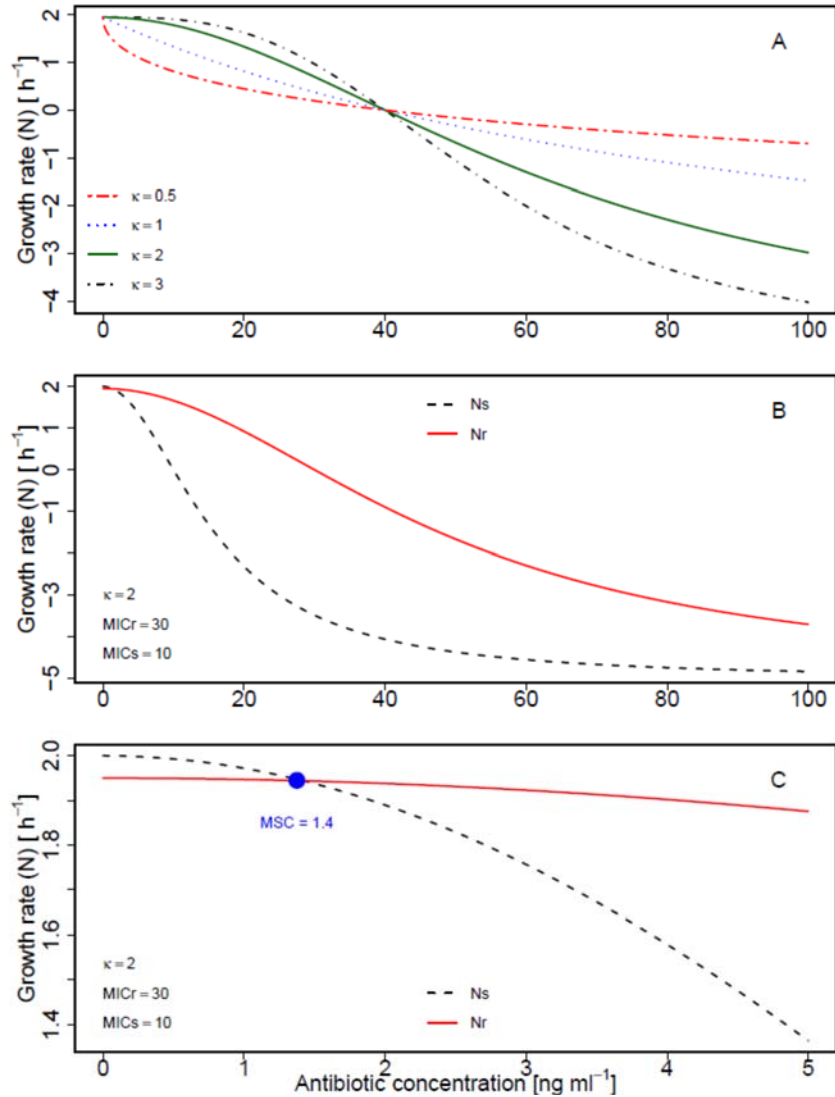**Figure 2.** Growth rate versus antibiotic concentration. (A) Single strain (MIC$_r$ = 40) with different kappa (κ) values. (B, C) Sensitive and resistant strains (MIC$_s$ = 10, MIC$_r$ = 30, κ = 2). (C) enlarged view around the MSC (●), the concentration where growth curves cross ($\Delta N = 0$). Other parameter values: sc = 0.05, $N_{int,s}$ = 2, N_{min} = −5.

To determine k_{max} from growth and death rates, we note that k_{max} should correspond to the difference between the maximum possible net growth rate (not limited by resource availability or antibiotics; i.e., N_{int}), and the minimum possible growth rate, after accounting for the growth limiting activity of antibiotic (N_{min}):

$$k_{max} = N_{int} - N_{min} = R_{int} - D_{int} - N_{min} \tag{4}$$

Generally, $N_{min} < 0$, indicating population decline at maximum antibiotic exposure level. The EC_{50} can be directly related to the MIC value [ng ml^{-1}]; as a result, the following formulation of D_{ab} applies for our formalism (full derivation in Appendix 1):

$$D_{ab,s}(a) = (N_{int,s} - N_{min}) \frac{a^\kappa}{a^\kappa + \frac{(-N_{min})}{N_{int,s}}(MIC_s)^\kappa} \tag{5}$$

$$D_{ab,r}(a) = (N_{int,r} - N_{min}) \frac{a^\kappa}{a^\kappa + \frac{(-N_{min})}{N_{int,r}}(MIC_r)^\kappa} \tag{6}$$

Equations 5 and 6 assume identical κ and N_{min} for sensitive versus resistant strains, which may not be accurate. Later in the text, we revisit the impact of this assumption for estimation of the MSC.

Difference in net growth rate and derivation of MSC as function of MIC

Competition between different bacterial strains is expressed by the difference in net growth rates. According to the conceptual model described by Andersson and Hughes [44] and Gullberg et al. [57], $N_s > N_r$ at low antibiotic concentrations, but the greater sensitivity causes more antibiotic-dependent growth inhibition for the sensitive strain. As a result, at high antibiotic concentrations, $N_r > N_s$, and the MSC is the point of intersection of the two growth curves ($N_s = N_r$) for which the difference in net growth rate is zero (Figure 2B-C).

Analytically, this difference in net growth rates between the resistant and the sensitive strain ($\Delta N(a)$ [h^{-1}]) is determined by subtracting Eq. 1 from Eq. 2, giving:

$$\Delta N(a) = N_r(a) - N_s(a) = N_{int,r} - N_{int,s} + \left(D_{ab,s}(a) - D_{ab,r}(a)\right) = \sigma + D_{ab,s}(a) - D_{ab,r}(a) \tag{7}$$

Thus, the MSC is the antibiotic concentration (i.e., a = MSC) at which the two net growth rates are equal and the difference (Eq. 7) is zero:

$$\Delta N(MSC) = N_{int,r} - D_{ab,r}(MSC) - N_{int,s} + D_{ab,s}(MSC) = \sigma + D_{ab,s}(MSC) - D_{ab,r}(MSC) = 0 \tag{8}$$

This is the concentration at which the additional loss in net growth due to antibiotic in the sensitive strain compared to the resistant compensates for the effect of fitness cost.

To derive the ratio of MSC/MIC we employ a dimensionless selection coefficient (sc [unitless]), obtained by reversing the sign of the reported experimental selection coefficient (σ)

[57,61], and then dividing by the net growth rate of the sensitive strain (derivation in Appendix 1):

$$\text{sc} = -\frac{\sigma}{N_{\text{int,s}}} = \frac{N_{\text{int,s}} - N_{\text{int,r}}}{N_{\text{int,s}}} = 1 - \frac{N_{\text{int,r}}}{N_{\text{int,s}}} \tag{9}$$

Based on the above equations, and further assuming that κ and N_{min} are the same for sensitive and resistant strains, the analytical solution for MSC/MIC$_s$ is obtained (derivation in Appendix 1):

$$\text{MSC/MIC}_s = \left(\frac{\text{sc}}{1 + \frac{N_{\text{int,r}}}{-N_{\text{min}}} - \frac{(1-\text{sc})\left(1 + \frac{N_{\text{int,s}}}{-N_{\text{min}}}\right)}{\left(\frac{\text{MIC}_r}{\text{MIC}_s}\right)^{\kappa}}} \right)^{\frac{1}{\kappa}} \tag{10}$$

In the case of a large difference in resistant versus sensitive MIC, the right-hand term in the denominator approaches zero, and the equation simplifies to:

$$\text{MSC/MIC}_s = \left(\frac{\text{sc}}{1 - \frac{N_{\text{int,r}}}{N_{\text{min}}}} \right)^{\frac{1}{\kappa}} \tag{11}$$

This simplification does not apply to small increases in MIC, such as the Δ*marR* and Δ*acrR* mutants which double the MIC for ciprofloxacin [57]. Eq. 11 becomes appropriate once MIC$_r$ > 5 x MIC$_s$, at which point results from Eqs. 10 and 11 become approximately equal (Appendix 1 Figure A1).

To summarize, Eq. 10 presents a quantitative hypothesis regarding the relationship between the antibiotic dose–response and the resulting MSC. The form of the equation aids in determining which aspects of the growth rate are most important for competition at low antibiotic doses. Eqs. 10 and 11 also provide a potential alternative to direct measurement of MSC. κ, $N_{\text{int,s}}$, and N_{min} could be measured in the laboratory [e.g., 65,76] and other parameters obtained from literature. Eq. 4 would be fit to experimental measures of the sensitive strain's antibiotic versus growth dose–response to obtain κ and k_{max}, and N_{min} would then be based on the strain intrinsic growth rate (N_{int}) minus k_{max} (Eq. 5).

As mentioned above, Eq. 10 rests on the assumption of identical κ and N_{min} for sensitive versus resistant strains. An analytical solution analogous to Eq. 10 could not be obtained assuming separate κ and N_{min} (i.e., κ_s, κ_r, $N_{\text{min,s}}$, $N_{\text{min,r}}$). In the Results section below and Appendix 1, we employ a Monte Carlo Simulation sensitivity analysis to evaluate this assumption of identical κ and N_{min}.

Model evaluation against experimental results

The analytical solution was evaluated by comparison to the experimental results of Gullberg et al. [57,61]. This evaluation was performed to determine whether the model fit to actual

competition data was reasonable, and to identify representative parameter sets for N_{min} and κ, parameters that are very system-specific, given the published values of MIC_s, MIC_r, σ and $N_{int,s}$ [57,61], parameters which have been characterized for a wide range of strains, conditions, and resistance mechanisms [e.g., 52,68,69,71,77–79]. Model fitting was achieved by fitting ΔN values calculated from Eq. 7 (based on Eqs. 1, 2, and 4 - 6) to the ΔN values observed in Gullberg et al. [57,61]. The function NonLinearModel.fit in MATLAB (Statistics Toolbox, R2013a, MathWorks, Natick, MA, USA) was used to estimate N_{min} and κ. Fitting was performed separately for seven individual antibiotic-bacteria combinations across the published range of experimental concentrations, as well as for arsenite- and copper-exposed *E. coli* [57,61]. These metals were included based on co-resistance and cross-resistance with antibiotics, as well as similar mechanisms of genetic transmission among bacteria [49,50,80,81]. For each compound, resistance was compared between a sensitive (wild-type) strain and one to four resistant strains in *S.* Typhimurium or *E. coli*. From the published experiments [57,61], only the chromosomal mechanism of trimethoprim resistance was excluded because it exhibited an average selection coefficient $\sigma > 0$, indicating no selective disadvantage of resistance [57].

To evaluate Eq. 7 and the underlying model assumptions, model-predicted vs. observed ΔN were compared. To evaluate robustness to individual observations, cross-validation (CV) was also employed. For CV, the optimization was performed with each single data point removed in series, and the average and range of N_{min} and κ results were examined, as well as the calculated vs. observed ΔN for the out-of-sample observations. The PRESS statistic (predictive residual sum of squares) was calculated, and PRESS/SSY and PRESS/SSE examined to indicate model prediction error and robustness to individual observations, respectively [82]. All analyses were performed on both the experimental average results for each strain and antibiotic concentration examined, reported by Gullberg et al, as well as the raw data for each experimental observation [Supplemental Information in 57,61], in order to consider the impact of experimental variation on results.

Results

The model depicts the change in growth rate versus antibiotic concentration (Eq. 7), the crossover point between growth rate of sensitive and resistant strains (i.e., the MSC, Eq. 8; Figure 2), and an analytical solution for the MSC/MIC ratio (Eqs. 10, 11). The MSC can be observed as the antibiotic concentration at which N_s (Eq. 1) and N_r (Eq. 2) cross, indicating identical growth of sensitive and resistant strains (Figure 2B-C, Eq. 8). The MSC/MIC ratio [57] is of interest because it indicates how much lower the MSC is relative to the MIC; this enables estimation of the environmental antibiotic concentration at which resistance selection could occur among competing bacteria populations [42,60]. The MSC can be estimated employing this ratio, in combination with routinely available MIC data [e.g., 52,77, and the EUCAST database: http://www.srga.org/eucastwt/wt_eucast.htm].

We first evaluate the model by examining a key assumption and then comparing predicted growth rates and MSC/MIC ratios to published data. We then examine model behavior and implications for MSC/MIC ratio prediction. Finally, we perform a sensitivity analysis to identify the most important parameters for predicting this ratio.

Model evaluation

Effect of varying κ and N_{min} for sensitive versus resistant strains: The analytical solution for Eq. 10 requires identical κ and N_{min} for sensitive versus resistant strains. We evaluated the impact of this assumption on model predictions by determining which strain-specific parameter values (i.e., $κ_s$, $κ_r$, $N_{min,s}$, or $N_{min,r}$) were most important for predicting MSC. To achieve this, we performed a Monte Carlo Simulation sensitivity analysis, detailed in the text and Appendix 1 Table A1. In two simulations, the predicted value of MSC was obtained in Eq. 10, assuming separate $κ_s$, $κ_r$, $N_{min,s}$, and $N_{min,r}$ in Eqs. 5 and 6. To be robust to MIC ratio variations, the first simulation had $MIC_r = 1.5$ x MIC_s whereas the second had $MIC_r = 10$ x MIC_s. In both simulations, MSC was highly sensitive to $κ_s$, (Spearman rank correlation coefficient, $ρ > 0.8$) but was insensitive to $N_{min, s}$ or $N_{min, r}$ (|Spearman $ρ$| ≤ 0.11). This much stronger influence of κ than N_{min} is expected based on the fact that κ is an exponential term (Eqs. 10, 11). MSC was also more sensitive to $κ_s$ than $κ_r$, and |Spearman $ρ$| between $κ_s$ and MSC was more than twice |$ρ$| between $κ_r$ and MSC. When $MIC_r = 10$ x MIC_s, almost all variation in MSC was explained by $κ_s$ ($ρ = 0.97$), with $ρ = −0.09$ for $κ_r$. These results indicate that MSC will strongly depend on $κ_s$, the shape of the antibiotic dose–response for the sensitive strain. As a result, for indirect estimation of MSC using Eq. 10, $κ_s$ should be well characterized experimentally.

Model performance for difference in net growth rates: Figures 3 and 4 displays the difference in net growth rates for sensitive versus resistant strains (ΔN) for previously published experimental data in comparison to the model (Eq. 7). Figure 3 illustrates how ΔN increases with increasing antibiotic concentration for the ciprofloxacin experiments in Gullberg et al. [57], with variability due to examination of four bacterial strains. Figure 4 directly compares the experimentally observed versus model predicted ΔN for all antibiotics and metals examined. The model predicted results overlapped with the range of experimental observations for most conditions. Much of the variability was attributable to experimental variation at specific antibiotic concentrations, as evident in the horizontal spread of the colored points. However, the model underpredicted experimental results for the aminoglycosides, KAN and STR (open circles in Figure 4) at ΔN < 0.05. Consequently, linear regression indicated $ΔN_{modeled} = 0.93(ΔN_{observed}) − 0.002$, a slight underprediction. Examining results for individual compounds, model performance (R^2, Q^2, and PRESS/SSY) was generally similar for either one parameter (κ) or two parameter (κ, N_{min}) fitted, and for either the raw or averaged experimental data (Appendix 1 Table A2). For CIP, ERY, KAN, and STR, the model fit was insensitive to N_{min}, exhibiting a wide range of possible values, and a limited impact on model fit. Therefore, N_{min} was fixed at a representative literature value of $N_{min} = −2$ [65,76,83], and κ was fitted to experimental observations. The fitted model was generally consistent with raw observations ($R^2 > 0.8$) and the model exhibited high predictive value in cross validation ($Q^2 > 0.8$) for TET, TMP, ERY, and As in *E. coli,* and for TET in Salmonella (Table 1, Appendix 1 Figure A2). Model fit was moderate for CIP ($R^2 = 0.78$, Figure 3), Cu ($R^2 = 0.73$), and STR ($R^2 = 0.67$). For KAN, the model fit was poor, worse than a simple average of the data, i.e., slope = 0 ($R^2 < 0$), indicating that it was not possible to fit the model to the KAN data (Appendix 1 Figure A3). Model fit to KAN was also poor for alternative statistical models, including Weibull, logit, logistic, and probit formulations.

Table 1. Results of model (Eq. 7) optimization to published [57,61] empirical growth rate differences (ΔN) between sensitive and resistant strains, based on strain-specific σ, MIC_s, and MIC_r. κ was fitted and raw experimental ΔN data was employed. Other model parameters: $N_{int,s}$ = 1.8, N_{min} = −2 [65,76,83]. MSC/MIC_s calculated using Eq. 10 with selection coefficients published for resistant strains [57,61].

Compound	Organism	Strains	n	κ	R^2	$Q^{2\,a}$	PRESS/SSY	PRESS/SSE	MSC/MIC_s
Tetracycline (TET)	*E. coli*	3	60	1.6	0.89	0.89	0.09	1.03	0.014
									0.063
Trimethoprim (TMP)	*E. coli*	2	118	2.5	0.88	0.87	0.07	1.03	0.18
Erythromycin (ERY)	*E. coli*	3	64	3.5	0.94	0.93	0.07	1.06	0.074
									0.27
Kanamycin (KAN)	*E. coli*	2	72	10.5	−0.47	−0.48	0.43	1.01	0.66
Arsenite (As)	*E. coli*	2	20	0.7	0.84	0.81	0.30	1.14	0.0064
Copper sulfate (Cu)	*E. coli*	2	8	1.9	0.73	0.43	0.80	2.13	0.035
Ciprofloxacin (CIP)	*E. coli*	5	144	2.0	0.78	0.77	0.31	1.03	0.024
									0.088
Streptomycin (STR)	*Salmonella*	2	87	5.0	0.67	0.66	0.25	1.02	0.38
Tetracycline (TET)	*Salmonella*	2	154	1.2	0.93	0.93	0.04	1.02	0.0077

a. Q^2 = cross validated R^2 = 1 (PRESS/TSS)

As shown in Table 1, the fitted κ ranged widely across the nine compounds examined (0.7 to 10.5). CV results generally produced a very narrow range, with κ varying by < 0.1 within individual compounds, except KAN and Cu (Appendix 1 Table A2). Similarly, CV PRESS/SSY results were < 0.4 for all compounds except KAN and Cu; values < 0.4 are considered to indicate reasonably low model prediction error [82]. The PRESS/SSE were below 1.15 for all compounds except Cu; these values of PRESS/SSE close to 1 indicate limited dependence of model prediction accuracy on individual observations.

Because it included four resistance genotypes, ciprofloxacin was examined more closely. Overall, fit and predictive ability were generally reasonable (R^2 = 0.81, Q^2 = 0.78, PRESS/SSY = 0.29, Appendix 1 Table A2) except for downward bias in the two highest ΔN results (Figure 3). These were both *gyrA1* [*S83L*] versus sensitive wild-type above 2 ng/ml ciprofloxacin [65]. The *gyrA1* [*S83L*] comparison had a substantially different curve shape, and removing this strain from the data greatly improved the model fit (R^2 = 0.97, Q^2 = 0.97, PRESS/SSY = 0.04). However the change in predicted κ was trivial (from 2.0 to 2.1, with N_{min} fixed at −2).

Minimum Selection Concentration: MSC/MIC_s was estimated (Eq. 10) based on model fitted κ, and empirical values for sc, MIC_r, and MIC_s. For these estimates, $N_{int,s}$ was set at 1.8 h^{-1} and N_{min} was either fitted or set at −2 h^{-1}. Model predictions corresponded well to the observed MSC/MIC_s [57,61] for all experiments, with either fixed or fitted N_{min} (Figure 5), suggesting that the model is appropriate to estimate the MSC/MIC_s ratio, which ranged widely from < 0.01 to 0.66 (Table 1, Appendix 1 Table A3).
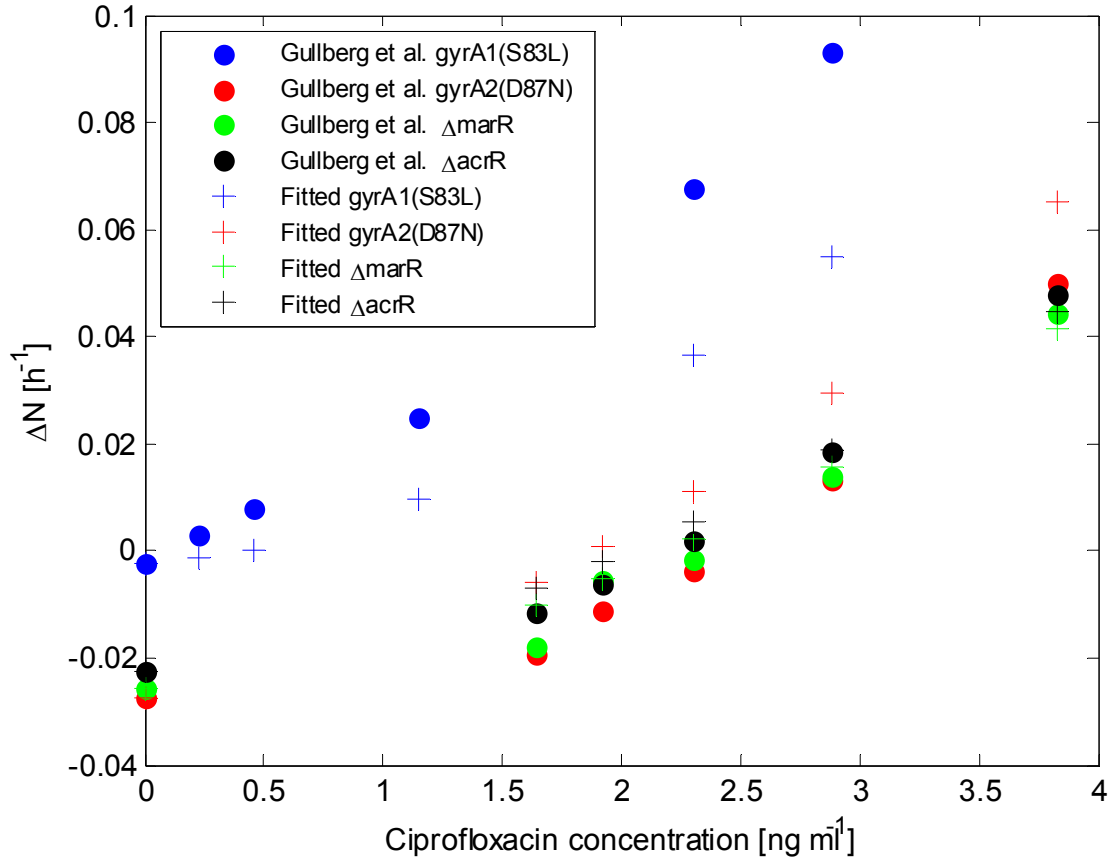
Figure 3. Comparison between experimentally observed [57] and model predicted difference in net growth rate between sensitive and resistant bacterial strains (ΔN, from Eq. 7) as a function of ciprofloxacin concentration in *E. coli*.

Sensitivity analysis

For a sensitivity analysis, behavior of Eq. 10 was examined across reasonable parameter ranges to examine sensitivity of MSC/MIC$_s$ to fitness differences (sc), antibiotic resistance differences (MIC$_r$/MIC$_s$), maximum growth rate inhibition (N$_{min}$), and intrinsic growth rate (N$_{int,s}$), respectively. Eqs. 10 and 11 indicate that MSC/MIC$_s$ is primarily a function of sc and κ, but is also modified by corrective terms that include N$_{min}$, N$_{int,s}$, N$_{int,r}$, MIC$_s$, and MIC$_r$. Figure 6 demonstrates the influences of sc and κ on MSC/MIC$_s$. Specifically, increasing sc lowers the resistant strain growth rate (Figure 6A-B), whereas increasing κ increases the curvature of the sensitive strain growth rate (Figure 6B-D), both resulting in increased MSC/MIC$_s$. As a result, modeled κ is strongly associated with model predicted MSC/MIC$_s$. For example, the Pearson correlation coefficient was very high (r = 0.94) for the κ versus MSC/MIC$_s$ results from Table 1.
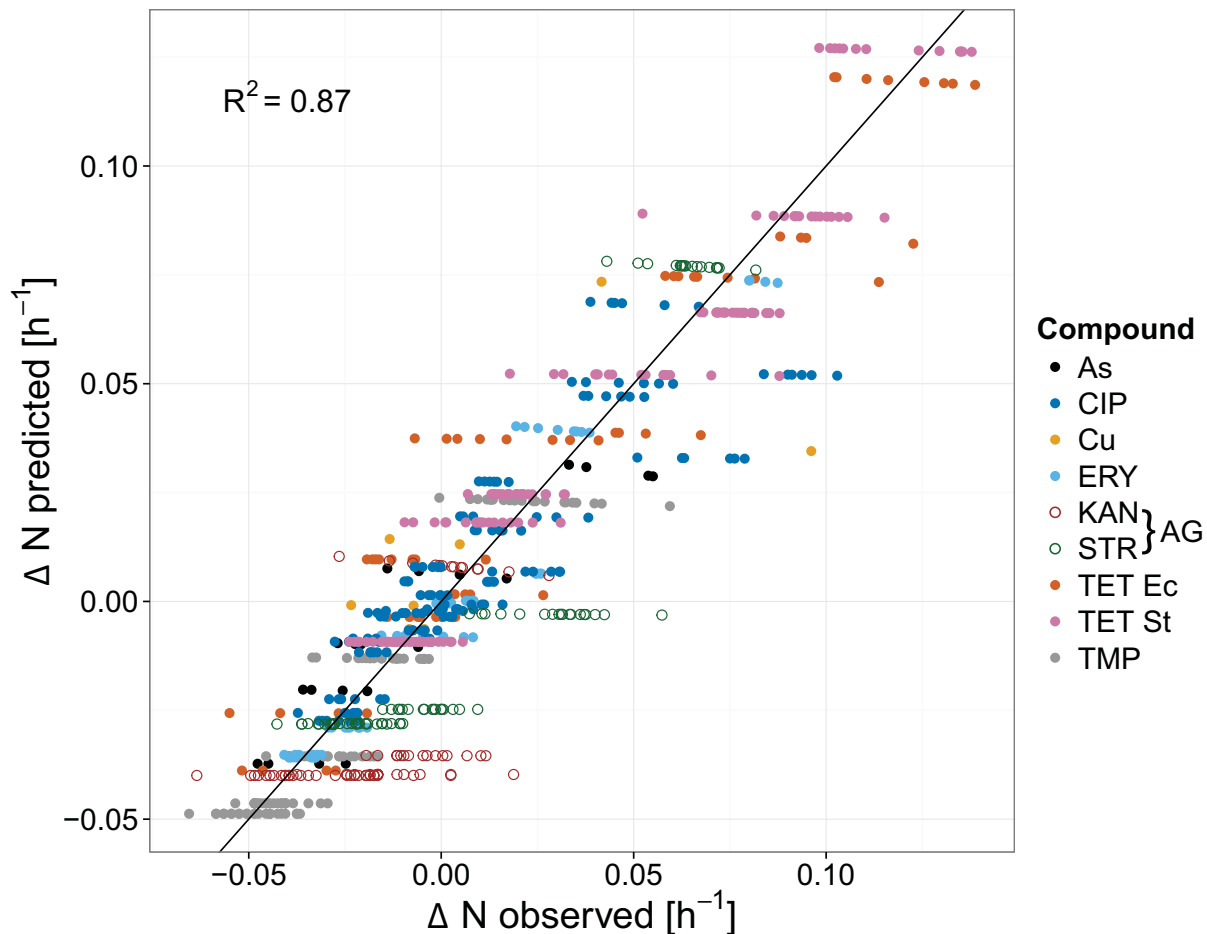
16

Figure 4. Model predicted versus ΔN observed for all experiments [57,61]. Abbreviations: As = arsenite. Cu = copper. AG = aminoglycoside antibiotic. Ec = *E. coli*. St = *S.* Typhimurium.

Figure 7 provides plots of the MSC/MIC_s ratio as the solution for Eq. 10 across different parameter values. Figure 7A confirms the dominant and interdependent influences of sc and κ on MSC/MIC_s with the largest influences at or below κ values of 1. At κ = 1, $MSC/MIC_s \approx sc$ (Figure 7A, blue dashed line). The influences of sc and κ can be combined according to Eqs. 10 and 11, which indicate that the MSC/MIC_s ratio is proportional to $sc^{1/\kappa}$. Figures 7B-D illustrate clearly that MSC/MIC_s is proportional to $sc^{1/\kappa}$ and that the slope of this relationship is modified by MIC_r, $N_{int,s}$, and N_{min}. An increase in MIC_r will decrease MSC/MIC_s, but this relationship is only sensitive when MIC_r approaches MIC_s (Figure 7B, MIC_r/MIC_s close to 1). The generally low sensitivity of MSC/MIC_s to the MIC values themselves corroborates the empirical finding of Gullberg et al. [57]. Increasing $N_{int,s}$ also decreases MSC/MIC_s but this only exhibits a minor influence in the plausible parameter range (Figure 7C). Finally, increasing N_{min} also decreases MSC/MIC_s, but this is only sensitive when N_{min} approaches zero (Figure 7D). N_{min} indirectly affects MSC/MIC_s by influencing the MIC versus EC_{50} relationship (Appendix 1, Eqs. A7 and A8).
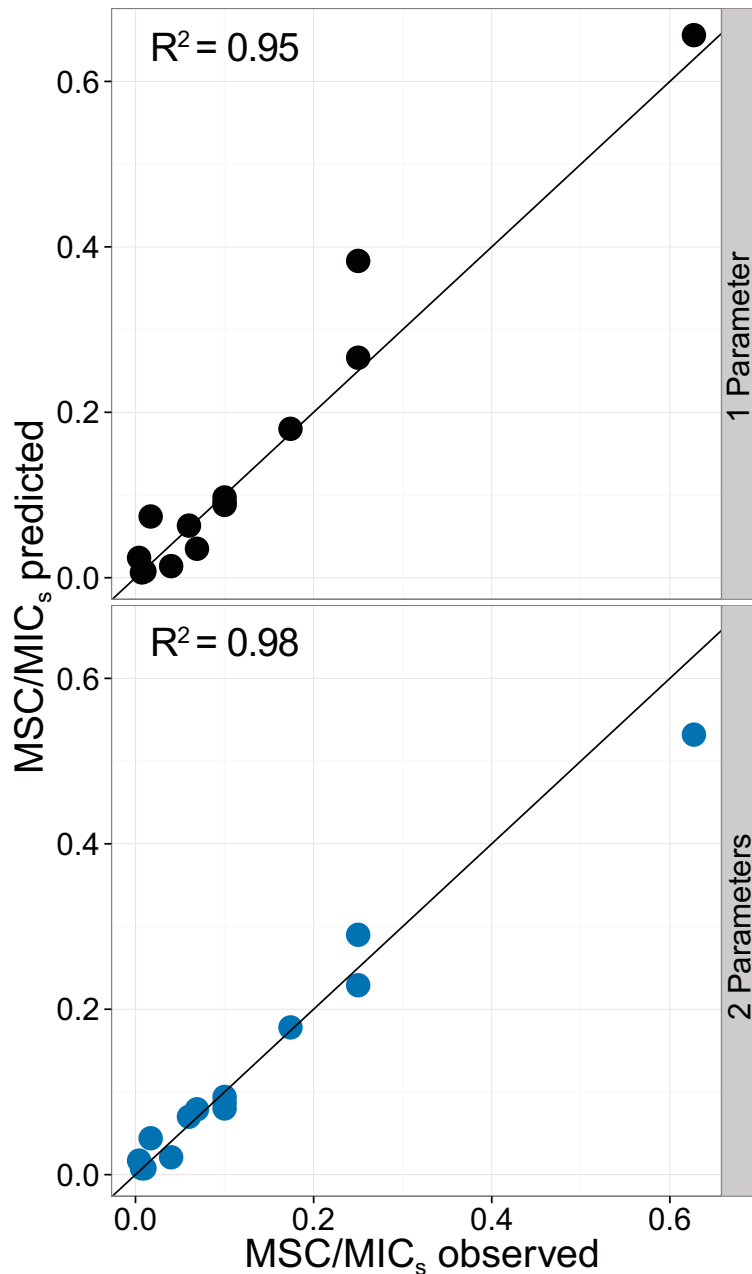
Figure 5. Comparison of observed [57,61] versus model predicted (Eq. 10) MSC/MIC$_s$. Symbols represent experimentally evaluated resistant strains (N = 14). Solid line (——) is 1:1 ratio.

Figure 7A also illustrates the expected range of the MSC/MIC$_s$ ratio across combinations of sc and κ (the most influential parameters). Over sc ranges from 0.001 to 0.1 and κ ranges from 0.5 to 5 [57,68,69,71], MSC/MIC$_s$ ranged widely from 10^{-6} to 0.5. With sc = 0.01, as κ decreased from 2 to 0.5, the MSC/MIC$_s$ ratio decreased from typically a factor of 0.1 down to less than a factor of 10^{-4}, indicating that MSC values are very sensitive around κ = 1. Especially for low sc, slight decreases in κ may correspond to steep declines in the MSC value (Figure 7A).
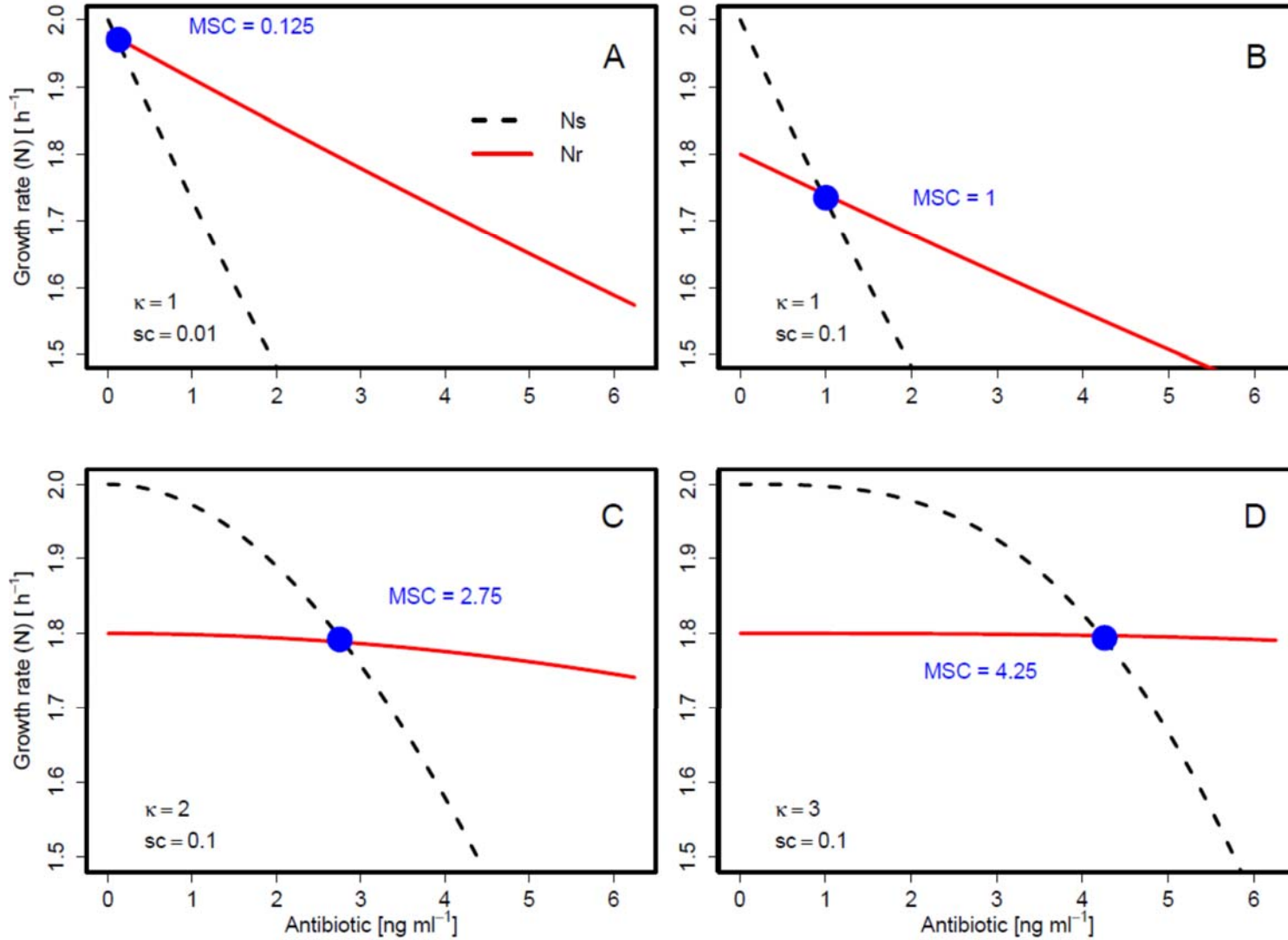
Figure 6. Growth rate versus antibiotic concentration for sensitive (N_s) and resistant (N_r) bacteria for different values of sc (0.01, 0.1) and κ (1, 2, 3). Other parameters (all scenarios): $MIC_s = 10$, $MIC_r = 40$, $N_{int,s} = 2$, $N_{min} = -5$.
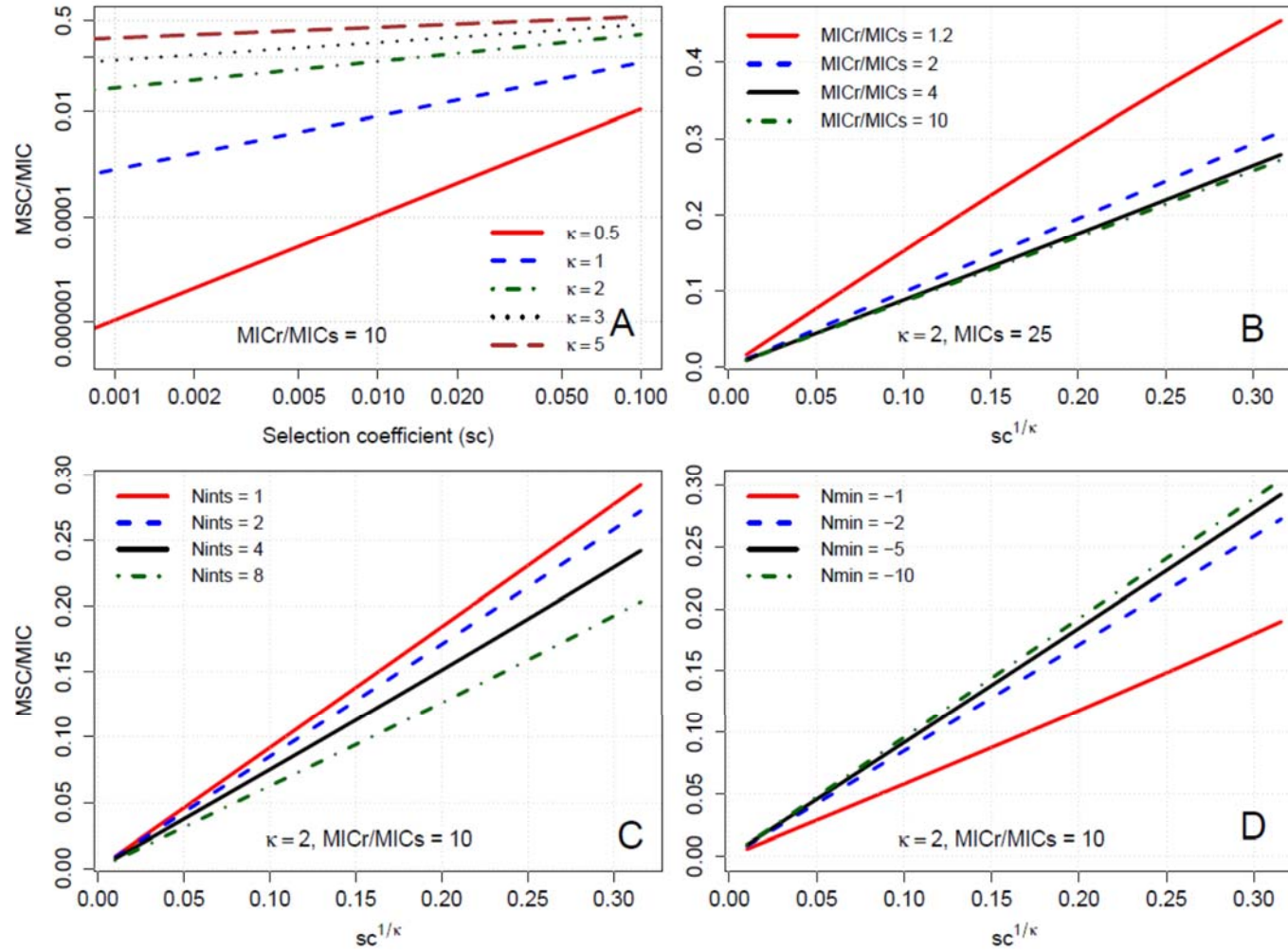
Figure 7. Sensitivity analysis of MSC/MIC$_s$ ratio (Eq. 10) to model parameters: (A) as a function of the selection coefficients (sc) for different κ (log scale); and as a function of sc$^{1/\kappa}$ for different values of (B) MIC$_s$/ MIC$_r$, (C) N$_{int,s}$, and (D) N$_{min}$. Other parameters, except as noted: MIC$_s$ = 25, MIC$_r$ = 250, N$_{int,s}$ = 2, N$_{min}$ = −5. Note: panel A is log scale and panels B-D are linear scale.

Discussion

This study model is a simple quantitative approach to describe the factors that will drive the MSC. As a relevant environmental threshold concentration for selection of resistant bacteria, the MSC helps us understand the significant issue of environmental resistance spread [42,48,49]. The model enables indirect estimation of the MSC using measurements of bacterial growth parameters that are readily obtained in the laboratory and literature (κ, N_{min}, and N_{int}), as an alternative and possible complement to direct measurement [57,58,61]. More importantly, the model mathematically illustrates the dependence of the MSC [44,60] on other more easily measured parameters and further identifies the shape of the antibiotic dose–response curve of the sensitive strain (i.e., κ) and the fitness cost of resistance (sc) as the main parameters determining the MSC/MIC ratio. These traits, combined with literature MIC ranges [e.g., 52,77, and the EUCAST database: http://www.srga.org/eucastwt/wt_eucast.htm], can be used to estimate environmental antibiotic concentrations at which resistance could spread.

The model consistently estimated the MSC/MIC ratio across the nine compound and taxa combinations examined, with overall R^2 above 0.95 (Figure 5). This finding suggests that one could estimate the MSC given: 1. the MIC; 2. intrinsic bacterial growth rate (i.e., N_{int}); 3. fitness cost (either σ or sc measurements); and 4. the shape of a dose–response curve for antibiotic concentration versus bacterial growth (i.e., κ). The first three values are readily available for a range of strains, resistance mechanisms, and conditions [52,57,65,68,69,76–79]. The antibiotic dose–response curve varies across treatment conditions but is routinely obtained, allowing experimental calculation of κ [65,67,75,76].

To illustrate use of the model, Figure 8 displays the MSC/MIC ratio from Eq. 10 across a range of selection coefficients, based on laboratory growth parameters from Regoes et al. [65] and Ankomah et al. [76]. Results vary dramatically across experiments, even for the same species-antibiotic combination (Figure 8), largely due to variations in κ. This suggests a strong impact of specific strains and growth conditions for selection, resulting in multiple orders of magnitude differences among systems, and a need to understand how the antibiotic resistance dose–response varies across antibiotic-contaminated environments [42], including water treatment systems, agricultural waste pens, and natural waters and sediments [52,84–87].

The model inconsistently predicted ΔN among compounds. ΔN was predicted least well for KAN and STR, both aminoglycosides. In these cases, the inability to fit ΔN well was due to the similarity of the study-observed MSC versus the sensitive strain MIC (i.e., high MSC/MIC$_s$ ratio). This amounted to a sudden and dramatic shift from the low experimentally determined ΔN values ($|\Delta N| < 0.04$) around the MSC versus $\Delta N > 1$ at the MIC. This steep dose–response from high to zero growth of the sensitive strain is evident in high κ values for both STR ($\kappa = 5$) and KAN ($\kappa = 10.5$). The Hill equation and other common statistical curves could not account for the similar MSC and MIC. The high κ fitted is also inconsistent with the concentration-dependent (i.e., low κ) bactericidal activity of aminoglycoside antibiotics described elsewhere [75,88]. Instead, the similar net growth rates of susceptible versus resistant strains close to the MIC$_s$ may result from adaptive resistance of the susceptible strain. Adaptive resistance for aminoglycosides has been widely observed in *Pseudomonas aeruginosa* [89,90], including at sub-MIC exposures [91], as well as in *E. coli* [83,92,93]. This temporary development of

phenotypic tolerance occurs due to elevated production of efflux pumps counteracting growth inhibition and killing at sublethal concentrations [91]. In cases of adaptive resistance, the MSC may not be much lower than the MIC. In such cases, the MIC may be a reasonable proxy for the MSC, as is often observed clinically [54].
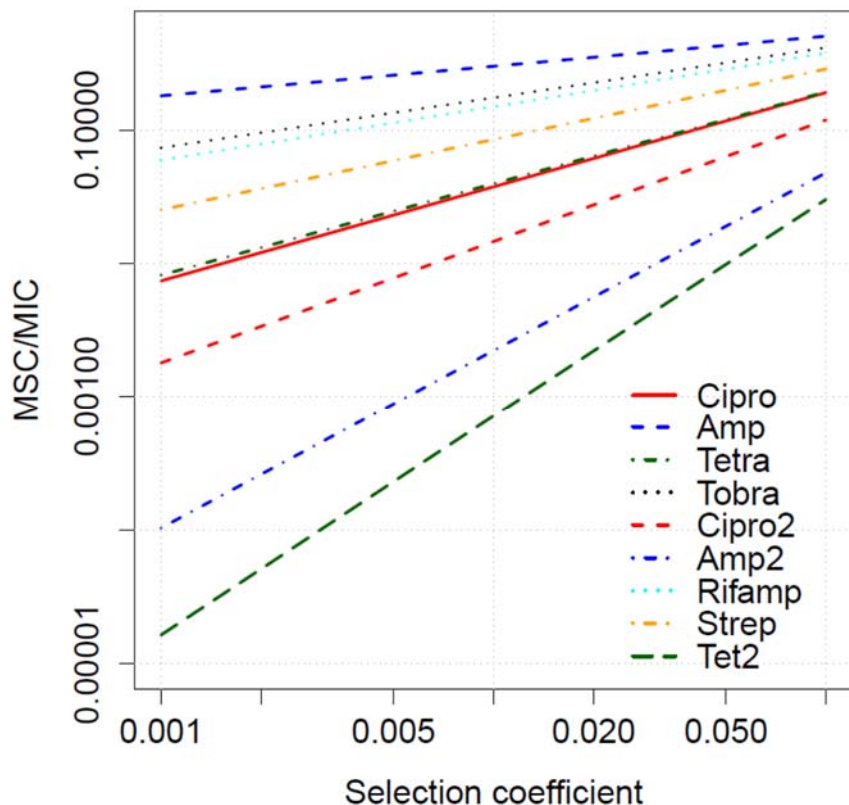
Figure 8. MSC/MIC as a function of the selection coefficient sc, calculated for parameters obtained in laboratory empirical studies [65,76]. Parameter data in Appendix 1 Table A4.

Experimental data are currently limited to a few species, strains, and antibiotics, possibly limiting the generalizability of the model performance evaluation. Thus, future experimental work is warranted to evaluate the ability to estimate MSC via Eqs. 10 and 11 across a range of subclinical conditions, species, strains, and antibiotics. This would include a comparison of MSC directly measured in competition experiments versus MSC derived from Eq. 10 based on measurement of the antibiotic dose–response of individual strains in isolation (Eqs. 4 and 5).

The shape of the antibiotic dose–response at subinhibitory concentrations

By emphasizing subinhibitory antibiotic concentrations, this study extends prior findings regarding how the behavior of the Hill equation, and κ in particular, influences the dynamics of bacterial net growth [65,75]. The model predicts that an antibiotic with a lower κ for a given set of conditions (e.g., bacterial strain, media) exerts a greater selective pressure in the subinhibitory region of concentrations found in the environment, resulting in lower MSC/MIC ratios. With $\kappa \approx 1$, there is an approximately linear decrease in growth from the intrinsic rate

with no antibiotic to zero growth when the antibiotic concentration is equal to the MIC. As a result, the intersection between the curves for the wild-type versus resistant strain can occur at a low antibiotic concentration, and the MSC is approximately equal to the MIC of the wild-type multiplied by the selection coefficient. This leads to a low MSC for low selection coefficients.

For higher κ conditions, the MSC is closer to the MIC. Thus, high κ, in addition to increasing efficacy above the MIC [65], also reduces the hazard of selection for resistance at concentrations below the MIC. Simulated and empirical dose–response measurements in the subinhibitory region are especially needed to evaluate the extent to which that 'pre-selection' of resistant strains may occur at MSC levels below the MIC of the sensitive strain, in both clinical and environmental settings.

Implications for resistance development hazard

Environmental-hazard and -risk assessments would also benefit from determining how ambient environmental concentrations in different media compare to the MSC [42]. Based on a species sensitivity distribution compared to EUCAST-published MIC results, Tello et al. found that selective pressure for resistant bacterial communities would be high in swine feces lagoon sediment but low in surface water, ground water, raw sewage, and sewage treatment plant effluent [52]. As an example of the implications of the MSC threshold (versus the MIC), we reinterpret the model of Tello et al. [52] to estimate hazard of selection for resistant bacteria. We employ a model correction factor, assuming a 100-fold lower species sensitivity distribution, to convert the study reported MIC_{50} [Figure 4 in Ref. 52] to an MSC_{50} by adjusting the reported log-logistic model location (α) parameter by -2. The 100-fold reduction follows our model results and the empirical data of Gullberg et al. [57,61], both of which indicate MSC/MIC ratios may exhibit values below 0.01. Comparing the adjusted model to the field data reported for ciprofloxacin [Table 2 in Ref. 52], the MSC_{50} model predicted a greater than 25% potentially affected fraction of bacterial taxa in at least one sample for all media reported (surface water, river sediment, raw sewage, and treatment plant effluent). For erythromycin and tetracycline, the MSC_{50} model predicted 65% and 88% potentially affected fraction in river sediment (vs. 2% and 1.6% for the MIC_{50}) [52]. Tello et al. used data from systems impacted by human and agricultural development [84,85], and our 100-fold MSC:MIC correction is more conservative than a 10-fold reduction employed in PNECs recently developed by Bengtsson-Palme and Larsson [64], thus indicating worst-case conditions. Nevertheless, these results indicate that hazard may exist for selection of resistant strains given antibiotic exposure in a wide variety of human-impacted aquatic settings.

Model scope, limitations, and future directions

The parsimonious analytical solution we developed addresses vertical gene transfer of antibiotic resistance in a well-mixed environment as a function of fitness cost, competition, and antibiotic concentration. There are many aspects of resistance dissemination that fall outside the scope of this simple exercise, including horizontal gene transfer [44,50,66], interactions among multiple strains, spatial arrangement of individual colonies, and heterogeneity in antibiotic exposure due to biofilms and other mechanisms [94,95]. Additionally, the model operates on and describes the long-term competition dynamics between bacterial strains, rather than stochastic and dynamic changes in net growth and competition over time. Thus, the

derivation assumes that the parameters governing growth (e.g., R_{int}, D_{int}, D_{ab}) will reach relatively stable values prior to the time that two strains are in direct competition. This simplified model does not incorporate inoculum effect, biphasic killing, delay functions, drug concentration changes, or other variations in parameter values that may occur in experimental settings. More sophisticated pharmacokinetic-pharmacodynamic models have been developed, incorporating these processes [96,97], but these more complex models do not lend themselves to an analytical solution similar to what we have provided. Further, investigation of varying initial ratios of resistant versus susceptible bacteria indicate no effect on selection coefficient, suggesting a limited importance of initial conditions, such as inoculum effect [57,61]. Nevertheless, theoretical and experimental investigation of how short-term growth and killing and other dynamic processes would impact the MSC/MIC ratio is warranted in future studies, as is comparison of alternative models.

The primary benefit of the present model is in illustrating the MSC paradigm and the key drivers of selection in simplified systems. As such, this paper adds to the growing scientific understanding on how to interpret laboratory data on the MIC and other parameters for predicting the emergence of resistance at subinhibitory environmental concentrations. It highlights the value of characterizing the antibiotic dose–response (i.e., the Hill Coefficient κ), particularly at antibiotic concentrations below the MIC. Ultimately, this quantification of resistance selection must be integrated into a risk assessment framework that also considers environmental antibiotic contamination, human exposure to and colonization by resistant bacteria, and the association between colonization and infection [42]. The ultimate objective is a further refined picture of the global hazard posed by antimicrobial agents.

Acknowledgements

Chapter 2. Transfer rate model for environmental surface contribution to hospital-associated infection transmission[4]

Abstract

We apply a quantitative environmental transfer rate model to evaluate which pathways contribute most to pathogen transmission. The model focuses on hospital-associated infections (HAI), the pathway of health-care worker (HCW) contact, and the relative importance of direct skin-to-skin contact versus indirect transfer via textiles and environmental surfaces. The model is formulated as a set of mass-balance equations describing the contact and transmission between the health-care worker (HCW) and an infected and an uninfected patient, residing in separate hospital rooms. The model was parameterized for a generic HAI that is transmitted by dermal contact and respiratory emissions. Elimination rate data was varied according to available literature values for six HAI: *Staphylococcus aureus*, *Streptococcus pyogenes*, *Acinetobacter baumannii*, *Bordetella pertussis*, severe acute respiratory syndrome coronavirus (SARS-CoV), and influenza A virus. Steady state results indicate that environmental surfaces are largely responsible for transmission. All pathogens except influenza exhibited high transmission to the susceptible patient skin. Excluding influenza, the range of best estimates of transmission among pathogens was similar to the variability observed within a single pathogen (*Acinetobacter baumannii*), suggesting that high parameter variability and uncertainty will impede quantitative classification of pathogens beyond existing heuristic frameworks. Our study results support the prevailing conceptual model of the importance of the non-human environment for HAI risk.

Introduction

To fully account for the transmission of pathogenic microbes in health care settings such as hospitals, the indoor environment must be considered. In military, community, and health-care settings, environmental pathways maintain infection risk even when infected individuals are isolated [98–100]. Environmental transmission pathways are diverse, including soil, water, air, textiles, indoor surfaces, or any mechanisms other than direct human-human contact. These indirect pathways have received limited attention in models of infectious disease transmission, but such models can provide insights into disease behavior and the relative merits of possible interventions [19,46,101–103]. Li et al. [46], Breban et al. [101], and Breban [102] modify traditional epidemic transmission models to incorporate environmental transmission of infectious diseases, allowing the allocation of different environmental components as reservoirs of pathogens. Based on model implications for the reproductive number[5], Li et al. [46] further classify pathogens as frequency-dependent vs. population density-dependent. Nicas and Sun [19] developed a compartmental model describing the mechanistic processes of transmission of a respiratory viral pathogen, such as influenza A. Nicas and Jones [18] incorporated that model into a transmission risk framework, enabling assessment of the relative importance of different pathways including direct contact, droplet spray, and aerosol inhalation. Mechanistic compartmental models could also be employed to evaluate the relative importance of environmental contact versus direct transmission among human individuals. Quantitative

---

[4] This study is a collaboration with Mark Nicas and Thomas E. McKone (University of California – Berkeley).

[5] Expected number of secondary cases from a single infectious case

25

estimation of the relative importance of environmental pathways can also help to evaluate the relative benefits of different interventions [18,19,45].

Infectious diseases differ widely regarding the relative importance of direct human-human pathways versus indirect environment-mediated pathways. Assessments of these differences to date have been largely qualitative rather than quantitative. Heuristically, pathogens that are exclusively foodborne, waterborne, airborne, or vector-borne (e.g., malaria) represent one end of a spectrum, whereas pathogens transmitted via direct human-to-human contact (e.g., HIV) are at the opposite end of this spectrum. However, many pathogens have more subtle differences, which can be effectively described in model-based analyses. Potentially important causes of differences among infectious diseases would include rates of transfer to and from environmental media (i.e., pick up and shedding), direct transfer rates between individuals, elimination rates, and infectious inoculum from different sources. Among these factors, elimination (inactivation) rates are especially variable among different media and pathogens [104–107], and therefore likely to be important for transmission among different pathways.

Surface-mediated transmission has been extensively studied and is an environmental pathway of great concern [106,108–110]. Factors reported to facilitate environmental surface-mediated transmission include long-term survival on surfaces, frequent contamination, ability to colonize and transfer from and to patient and health-care worker skin and hands, resistance to disinfection, and small inoculating dose [109]. Nosocomial (hospital-associated) infections reported to have these attributes, and thus an environmental transmission pathway, include methicillin-resistant *Staphylococcus aureus* (MRSA), *Pseudomonas aeruginosa*, vancomycin-resistant *Enterococcus spp.*, *Clostridium difficile*, norovirus, hepatitis B virus, *Acinetobacter spp.*, *Candida spp.* and many others [109].

We apply a simple mechanistic model of pathogen environmental behavior and transfer to describe differences among pathogens in the importance of environmental transmission between individuals. The model is applied in a standardized hospital scenario [18,19,111], depicting pathogen transmission between two patients residing in separate hospital rooms [45]. We perform simulations including or excluding different pathways (textiles and nonporous surfaces) to evaluate the relative contribution of environmental surfaces to total number of colonies transmitted. This is intended as a first step towards a quantitative modeling framework for evaluating how environmental transmission varies among pathogens.

Methods

The model

The study model combines and modifies components of the environmental transmission models of Plipat et al. [45] and Nicas and Sun [19], applying them in a steady-state formulation. The resulting model describes the transmission of pathogenic bacteria or virus particles between two patients residing in separate hospital rooms, as mediated by a health-care worker (HCW) caring for both patients [45]. We model patients in separate rooms in order to focus on the role of environmental (e.g., surface-mediated) transmission and the contribution of textiles and nonporous surfaces (i.e., the non-human environment) to this transmission. As such, airborne and droplet spray transmission that would occur among patients in the same room [18,19] are

26

not emphasized. This simplified scenario was selected to facilitate comparison among hypothetical scenarios and model parameters for: 1. total particle transmission and 2. contribution of human skin versus environmental surface (textiles and nonporous surfaces) to transmission between individuals. This model forms a conceptual basis to describe specific pathogens in terms of potential for transmission among patient rooms, and should ultimately enable assessment of the relative merits of different methods for controlling pathogen transmission (e.g., surface and textile sterilization, barrier controls, or hand washing) among rooms in a hospital setting.

In the model scenario, patient 1 is assumed to be infected and also colonized in the mucous membranes (e.g., nares and mouth), and the HCW and patient 2 are assumed to be uninfected and uncolonized. The model calculates the rate and number (i.e., count) of pathogen colony-forming units (cfu) transmitted to the HCW and patient 2 by different pathways, enabling assessment of the relative contribution of skin transmission versus transmission via porous surfaces (i.e., textiles) and nonporous surfaces. In the initial simulation, there is no assumed loss of pathogen due to hand washing; this model formulation represents a "worst case scenario" for transmission rates [45]. However, each 8 hours, the model assumes that HCW skin is cleaned to zero bacteria concentration as a result of shift turnover [45].

The model describes contributions for each pathogen transmission pathway via a separate rate transfer equation, λ [h^{-1}] [19]. The mucous membranes of patient 1 are treated as the source (G) [cfu h^{-1}]. The model does not include the shedding source of Plipat et al. [45], instead focusing on the mucus membrane sources of dermal contact and coughing [18,19], which are relevant for a range of infectious diseases. The model is also simplified from Plipat et al. in that binary time-dependent contact vs. non-contact of HCW with patients are replaced by simple reduced contact rates (e.g., HCW spending 20 min/h in a patient's room is modeled as a 1/3 multiplier to contact rate). Unlike Nicas and Sun [19], the model does not include direct airborne transmission to the respiratory tract by droplet spray or inspiration of particles immediately released via coughing. These pathways are excluded under the assumption that this direct airborne transmission will be exceedingly small, given that patients reside in separate rooms, in contrast to Nicas and Sun's [19] scenario of patients residing in the same room.

The model is formulated as a set of coupled mass-balance equations with 11 state variables (compartments) describing disease-organism number in each compartment, and first order rate transfers between compartments:

$$\begin{bmatrix} dN_1/dt \\ dN_1/dt \\ ... \\ dN_j/dt \end{bmatrix} = \begin{bmatrix} -\lambda_{1L} & \lambda_{2,1} & ... & \lambda_{j,1} \\ \lambda_{1,2} & -\lambda_{2L} & ... & \lambda_{j,2} \\ ... & ... & ... & ... \\ \lambda_{1,i} & \lambda_{2,i} & ... & -\lambda_{jL} \end{bmatrix} \times \begin{bmatrix} N_1 \\ N_2 \\ ... \\ N_j \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \\ ... \\ G_j \end{bmatrix} \qquad (1)$$

where N = count of disease organisms (i.e., particles), subscripts $_{1....i}$ and $_{1....j}$ denote the different source and recipient compartments (i = j = 11), $\lambda_{i,j}$ = rate parameter [h^{-1}] for pathogen transfer between compartment i and compartment j, and λ_{iL} = total loss rate [h^{-1}] from

compartment i to all other compartments (i.e., $\lambda_{iL} = \sum_{n=1}^{n=j} \lambda_{i,n}$), and G_i = input to compartment i from external sources. These equations in simplified matrix form are:

$$\mathbf{N`(t) = \Lambda \times N + G} \tag{2}$$

In this formulation, the mucous membranes of patient 1 are treated as the pathogen source, such that $\mathbf{G}$ = rate of production of new pathogenic microbes in the mucus membranes. Here, production could occur in the respiratory and digestive tract, as well as within the mouth and nares. Pathogenic microbes are released to textiles (G_2), nonporous surfaces (G_3), and surrounding air (G_4) by coughing. Pathogenic microbes are released to infected patient skin (G_1) by coughing, as well as hand contact with the mucous membranes. The model compartments (compartment number) are (as shown in Figure 9): skin of patient 1, HCW, and patient 2 (compartments 1, 6, 9); textiles of patient 1, HCW, and patient 2 (compartments 2, 7, 10); nonporous surfaces of patients 1 and 2 (compartments 3, 11); mucous membranes of HCW and patient 2 (compartments 5, 8); and air in the room of patient 1 (compartment 4). All compartments except the mucus membranes exhibit compartment-specific loss via elimination (organism death), which is treated as a loss from the system, $\lambda_{i,12}$ (i.e., transformation is an "absorbing state") [19]. Loss from room air ($\lambda_{4,12}$) also includes exhaust from the room [19].

The steady state abundance of pathogenic organisms for each compartment ($N_{1...i, SS}$; N_{SS}) is adequate for our purpose of comparing across pathogens and pathways. By definition, dN/dt = 0 at steady state, such that the following solution is readily obtained:

$$\mathbf{N`(t)_{SS} = 0} \tag{3}$$

$$\mathbf{0 = \Lambda \times N_{SS} + G} \tag{4}$$

$$\mathbf{- G = \Lambda \times N_{SS}} \tag{5}$$

$$\mathbf{\Lambda^{-1} \times (- G) = \Lambda^{-1} \times \Lambda \times N_{SS}} \tag{6}$$

$$\mathbf{N_{SS} = \Lambda^{-1} \times (- G)} \tag{7}$$

Thus, the steady state abundance is the cross product of the inverse of the transfer rate matrix and the negative of the source matrix.

The model includes five airborne transfer pathways: 1. release from the infected patient's nose into air via coughing; 2. transfer from air to the skin or textiles (e.g., clothing) of the patient and a health-care worker; 3. transfer from air to the nonporous surfaces surrounding the patient; 4. transfer from air to the mucous membranes of the HCW; and 5. loss from air by exhaust or environmental transformation (i.e., organism death) [18,19,111]. Modeled skin-contact mediated transfer pathways are based on those of Plipat et al. [45], who use ordinary differential equations to depict the transfer rate of bacterial cfu between 10 hospital-setting compartments. The authors developed and parameterized this model based on colonization with MRSA, which is a model nosocomial bacterial pathogen with an environmental surface-mediated transmission pathway. Appendix 2 presents all model equations and constants.

The model was coded in R Version 3.1.2 [112]. The matrix formulation allows for ready observation and comparison of rate constants (λ values) [19], as well as straightforward removal of pathways. Model code will be provided to users on request. For this exercise, the surface environment is defined to include both textile (porous) and non-textile (nonporous) surfaces (Figure 9, beige filled boxes).
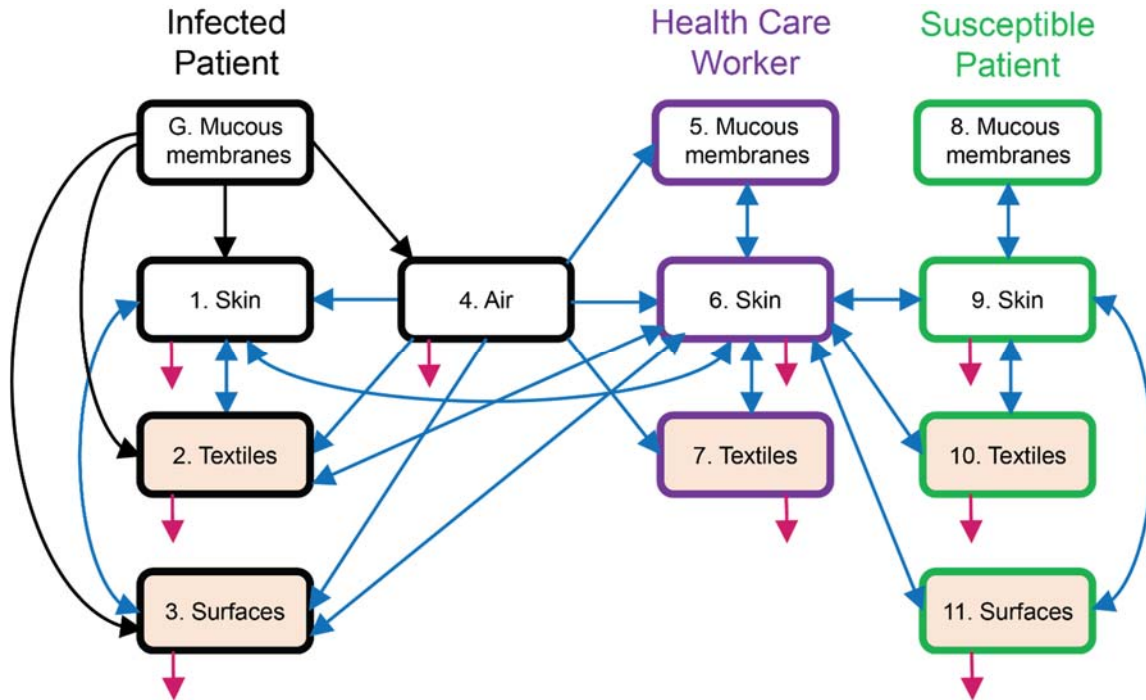
Figure 9. Conceptual depiction of the pathogen environmental transmission model. Boxes represent compartments, and arrows represent rate processes [h⁻¹]. Black arrows (→) are inputs (G), blue arrows (→) are transmission between compartments ($λ_{i,j}$), and pink arrows (→) are elimination ($λ_{iL}$). Box borders indicate the personal environments of the infected patient 1 (black □), health-care worker (violet □), and susceptible patient 2 (green □). Finally, pink filled boxes (■) indicate the non-human environment, comprising textiles and surfaces.

Parameter development focusing on elimination rates

Model parameters were obtained from published literature (Appendix 2). Parameters were largely obtained from Plipat et al. [45] and Nicas and Sun [19]. However, a literature review was performed to obtain parameter estimates for elimination rates (i.e., death of pathogenic microbes), including differences in elimination rate between skin, surfaces, and textiles (Appendix 2). Elimination was emphasized based on the expected sensitivity of model results to this parameter. Values were compiled for six pathogens known to cause hospital-associated infections: *Staphylococcus aureus*, *Streptococcus pyogenes*, *Acinetobacter baumannii*, *Bordetella pertussis*, severe acute respiratory syndrome-coronavirus (SARS-CoV), and influenza A virus.

Many studies provide experimental bacteria survival and elimination results in figure format only, without a tabular summary or elimination rate calculation. To estimate elimination rates

from these studies, study raw data were manually extracted using the open source ImageJ software platform [113], with the Figure Calibration plug-in. We estimated elimination rates [h^{-1}] from published results via one of two methods. When decay curves were approximately first-order, based on multiple measurements of concentration, we fit study data points to an exponential decay function, and estimated elimination rate as the negative exponential term [111]. Often, the experimental elimination rate was greater during the first portion of the experiment; in these cases, we used only initial data points to fit the exponential function to data, resulting in a higher elimination rate estimate (example in Figure 10). In cases where graphical results indicated zero-order loss (linear decrease) with time, or when tabular data were only reported as the time to reach zero abundance (e.g., *B. pertussis*, Appendix 2 Table A10), we estimated elimination rate as 1/t, where t = reported total time to reach zero abundance.
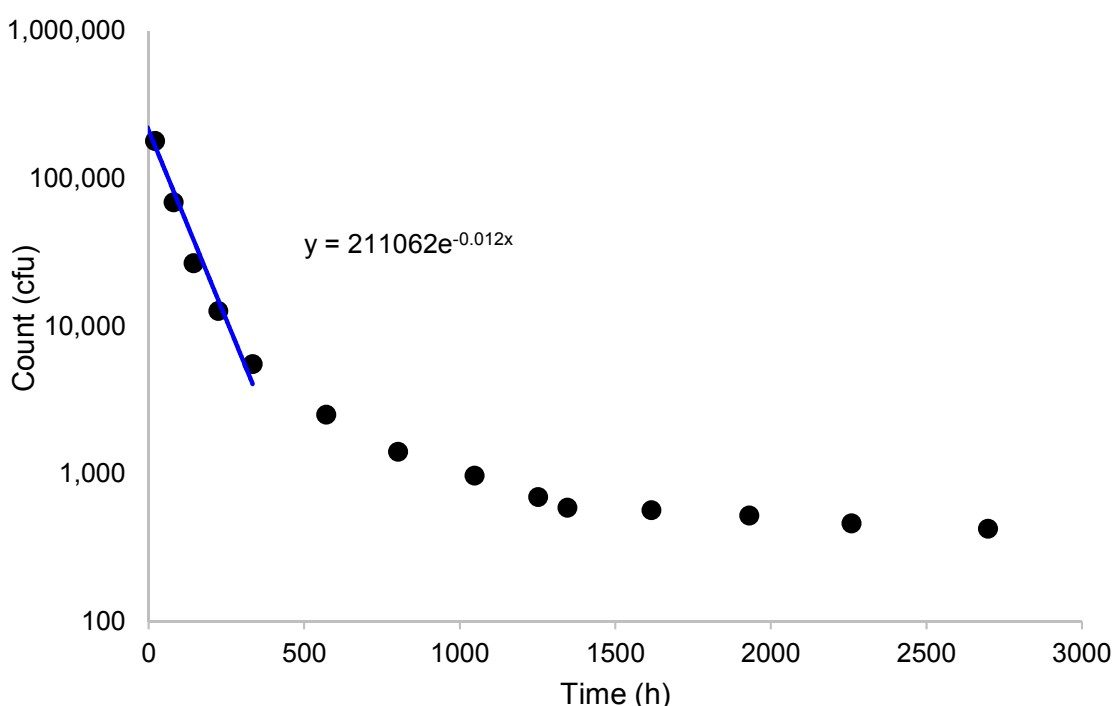
Figure 10. Example elimination rate calculation (*A. baumannii*). In this example, an exponential decay function was fitted (—) to results of the first five points, which were manually extracted from Figure 1 (Type 2 panel) of Wendt et al. [114]. Based on the fitted model, the estimated loss rate was 0.012 h^{-1}.

Most elimination studies focus on nonporous surfaces. When quantitative elimination estimates on skin or textiles were unavailable, extrapolations were made based on ratios between first-order elimination in that medium versus nonporous surface rates, extracted from the best available relevant study (Appendix 2). For SARS-CoV on textiles this ratio was based on the semiquantitative estimate of Lai et al. [115]. For *S. pyogenes* and *A. baumannii* elimination on skin, ratios were extracted from studies of the congeneric *S. pneumoniae* [116] and *A. calcoaceticus* var *anitratus* [117], respectively. Finally, to examine parameter variation within the literature for a given infection, we developed low, average, and high literature estimates for elimination of *Acinetobacter baumannii* on nonporous surfaces.

Model simulations

We first used the model to calculate steady-state results (Eq. 7), employing the different elimination rates obtained from the six pathogens, in addition to the range of plausible elimination rates in the literature for *Acinetobacter baumannii* (Table 2). In order to illustrate the contributing role of the non-human environment (textiles, surfaces and fomites) to human exposure, additional simulations were performed with some environmental compartments removed. In the first, transfer to and from nonporous surfaces was removed (boxes 3 and 11 in Figure 9). In the second, transfer to and from all environmental compartments, including surfaces and textiles, was removed (boxes 2, 3, 7, 10, and 11), such that only the skin, mucus membranes, and air were considered (boxes 1, 4, 5, 8, and 9, Figure 9).

Table 2. Elimination rate estimates [h^{-1}] for pathogens on different media.

Pathogen	Skin	Textiles	Surfaces	Air	Source [a]
Acinetobacter baumannii (low)	0.0084	0.00038	0.00038	0.00038	[114,117–120]
Acinetobacter baumannii (average)	0.072	0.0031	0.0031	0.0031	[114,117–120]
Acinetobacter baumannii (high)	0.59	0.026	0.026	0.026	[114,117–120]
Staphylococcus aureus	0.21	0.038	0.012	0.012	[45]
Streptococcus pyogenes	0.20	0.051	0.051	0.051	[116,121,122]
Influenza A	72	0.96	0.50	0.44	[18,111,123]
Bordetella pertussis	0.34	0.021	0.047	0.051	[105]
SARS-CoV	0.032	0.38	0.032	0.032	[115,124,125]

a. See Appendix 2 for details.

Results and Discussion

Elimination rates

Literature elimination rates varied by 7 orders of magnitude between 3.8×10^{-4} h^{-1} for the low estimated elimination in *Acinetobacter baumannii* to 72 h^{-1} (i.e., 1.2 min^{-1}) for influenza A on skin. Most rates were 0.003 to 0.5 h^{-1} (Table 2). We obtained 10 separate elimination rates for *A. baumannii* on nonporous surfaces from literature sources [114,117–120] (Appendix 2 Table A7), allowing an examination of the statistical properties and variability of these data. Histogram plots and normal scores plots indicate that these rates are approximately lognormally distributed (Figure 11). Low, average, and high loss rates were estimated based on appropriate summary statistics for lognormal data. The average elimination rate was taken as the geometric mean: 3.1×10^{-3} h^{-1}. The low and high estimates were taken as $10^{\text{mean(log10(values))} \pm 2 \text{ SD(log10(values))}}$, and were 3.8×10^{-4} and 2.6×10^{-2} h^{-1}, respectively. Based on elimination rates extracted from Musa et al. [117], who compared elimination rates in *Acinetobacter calcoaceticus* var *anitratus* on Formica surfaces and on fingertips, we estimated that elimination on skin was 23 times faster than that on surfaces (Appendix 2). The range of elimination rates for *A. baumannii* was lower than those for influenza A, but generally overlapped with the rates observed for other pathogens (Table 2).
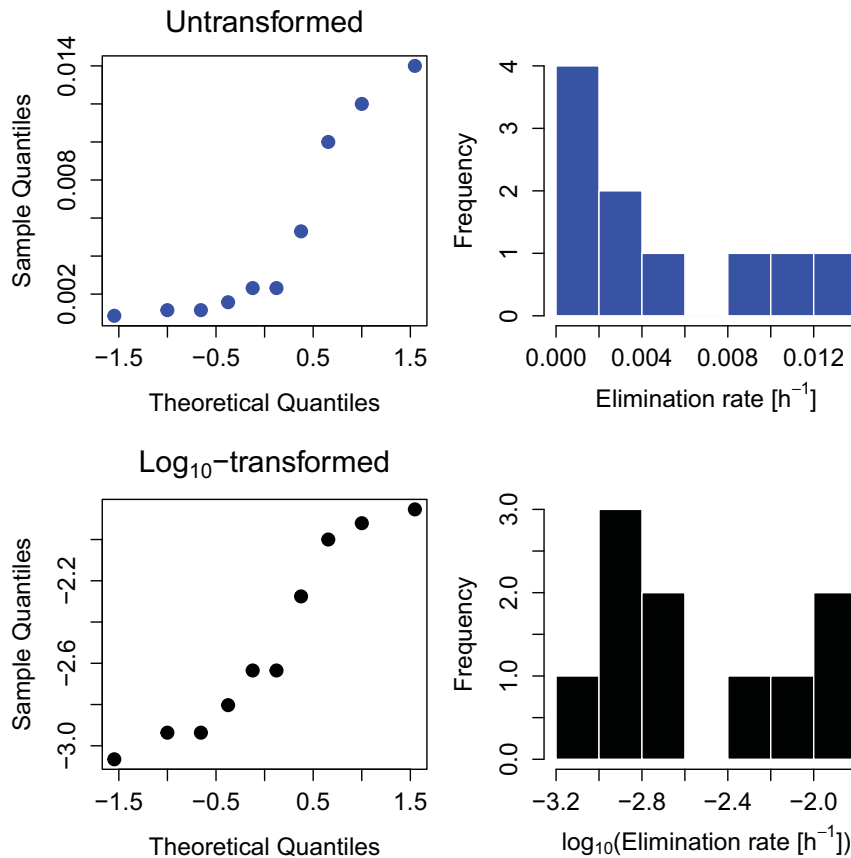
Figure 11. Normal scores plots and histograms illustrating elimination rates for *Acinetobacter baumannii* on nonporous surfaces (N = 10), with both untransformed (top) and $\log_{10}$ transformed (bottom) data.

Model simulations

Calculated rates were generally greatest for transfer between skin and surfaces, as well as between HCW skin and textiles (Table 3). Rates were relatively low between air and other compartments, indicating a high contribution of dermal pathways relative to airborne pathways in the model scenario we developed. In comparing total loss rates for each compartment (Table 3, diagonal clear cells), we observe that loss was greatest from room air, patient surfaces, and health-care worker skin. The high loss in room air resulted from room exhaust rates assumed at 6 turnovers per hour (6 h^{-1}) [19]. For surfaces and skin, the high loss resulted from the large assumed exchange between these compartments and other reservoirs, especially for health-care worker skin, which is a primary transfer pathway in the model (Figure 9). In terms of inputs from the infected patient mucous membranes (G) onto other compartments (Figure 9), estimated rates were greatest for transfer (i.e., input) onto skin and textiles (237,600 cfu h^{-1} each), intermediate for transfers onto nonporous surfaces (52,800 cfu h^{-1}), and minimal (and effectively irrelevant) for transfers to air (0.5 cfu h^{-1}). This reflects the assumed important pathways for nasal transmission of MRSA [45] and that a large portion of respiratory secretions released via coughing ultimately end up depositing on surfaces [19].

Table 3. Matrix of transfer and loss rates (λ) [h⁻¹] for all compartments, assuming model default parameters (Appendix 2 Tables A5-A6), and elimination rates for *Staphylococcus aureus* (Table 2). MM = mucous membrane. Unshaded diagonal cells are total loss rates (λ_{iL}). Remaining cells are transfer rates ($\lambda_{i,j}$) and are shaded green to red according to magnitude.

			Recipient										
			Infected patient				Health-care worker			Susceptible patient			
			1. Skin	2. Textiles	3. Surfaces	4. Air	5. MM	6. Skin	7. Textiles	8. MM	9. Skin	10. Textiles	11. Surfaces
Source	Infected	1. Skin	−0.58	0.013	0.48	0.010	0	0.07	0	0	0	0	0
		2. Textiles	0.06	−0.056	0	0.047	0	0.02	0	0	0	0	0
		3. Surfaces	0.24	0	−0.65	0.047	0	0.08	0	0	0	0	0
		4. Air	0	0	0	−6.1	0	0	0	0	0	0	0
	HCW	5. MM	0	0	0	0.014	−0.075	0.00015	0	0	0	0	0
		6. Skin	0.07	0.0044	0.16	0.0011	0.075	−0.73	0.18	0	0.07	0.0044	0.16
		7. Textiles	0	0	0	0.0011	0	0.18	−0.22	0	0	0	0
	Susceptible	8. MM	0	0	0	0	0	0	0	−0.075	0.00015	0	0
		9. Skin	0	0	0	0	0	0.07	0	0.075	−0.58	0.013	0.48
		10. Textiles	0	0	0	0	0	0.02	0	0	0.060	−0.056	0
		11. Surfaces	0	0	0	0	0	0.08	0	0	0.24	0	−0.65

All modeled pathogens also exhibited a dramatic 1 to 2 order of magnitude decline in steady state bacterial count on the susceptible patient, after the removal of nonporous surface transmission pathways, and further decline upon removal of textiles (porous surfaces; Figure 12). This dramatic change illustrates the essential role that indirect pathogen transfer via the non-human environment can play in infection transmission [46,102]; both textiles and surfaces serve as large reservoir areas effectively collecting and storing disease organisms.
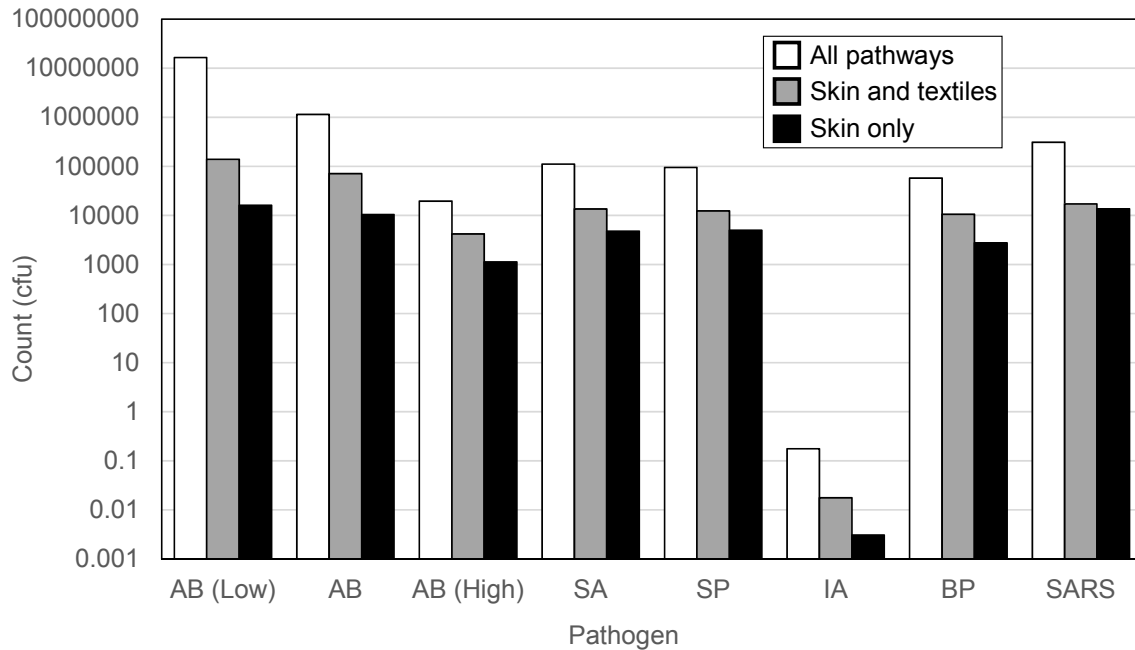
Figure 12. Steady state count of pathogens (cfu) on patient 2 skin for different modeled scenarios. AB = *Acinetobacter baumannii*, SA = *Staphylococcus aureus*, SP = *Streptococcus pyogenes*, IA = influenza A virus, BP = *Bordetella pertussis*, and SARS = severe acute respiratory syndrome coronavirus.

In our exercise, differences between pathogens resulted from differences in elimination rates only (Table 2), with all other parameters being fixed. All pathogens except influenza A exhibited a high organism count on the susceptible patient 2 (Figure 12). The low count for influenza A stems from its extremely high elimination rate, especially on skin surfaces and textiles (Table 2). The strong relationship between elimination on skin and modeled organism count on patient 2 illustrates this, with influenza A clearly separated from the other pathogens (Figure 13). This supports the prevailing conceptual model that airborne transmission, droplet spray, and other direct transmission pathways are the primary transmission risks for influenza A [104]. Similar to influenza A, literature reviews by Kramer and colleagues [106,107] indicate maximum persistence of less than 72 h on surfaces for Parainfluenza virus, Respiratory Syncytial virus, Cytomegalovirus, and bacteria including *Chlamydia pneumoniae*, *Helicobacter pylori*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, and *Proteus vulgaris*. Based on the pathways and scenario we depicted in this model, these pathogens would all be expected to exhibit a similarly low hazard of HCW-mediated transmission via the pathways we included. This suggests that in even the conservative simulation we performed (high HCW time in patient room, no hand washing), the transmission of these particular pathogens among patients in

separate hospital rooms due to dermal contact, textiles, and fomites of the HCW is likely to be exceptionally low.

Other than influenza A, the remaining pathogens we modeled exhibited high counts on the susceptible patient skin as a result of the modeled pathways. With the exceptions noted above, the majority of pathogens examined in the literature exhibit persistence on surfaces similar to, or greater than, those we evaluated [106,107]. This supports the conceptual model in which surfaces serve as reservoirs for a wide variety of nosocomial infections and surface contamination is justified as a serious concern in health care settings [106,109].

For most pathogens we considered, including textiles (as opposed to skin alone) contributed moderately to the total exposure (Figure 12, Figure 14). An exception was SARS-CoV, for which textiles played a trivial role (Figure 14). This is because the single available study comparing SARS-CoV elimination between different surfaces indicates a substantially lower persistence on textiles (cloth lab coat) than nonporous surfaces (disposable plastic gown) [115]. Data on elimination rates on textiles were limited for other pathogens, but for *B. pertussis*, persistence on textiles was comparable to that on glass, skin, or plastic (Appendix 2 Table A10).
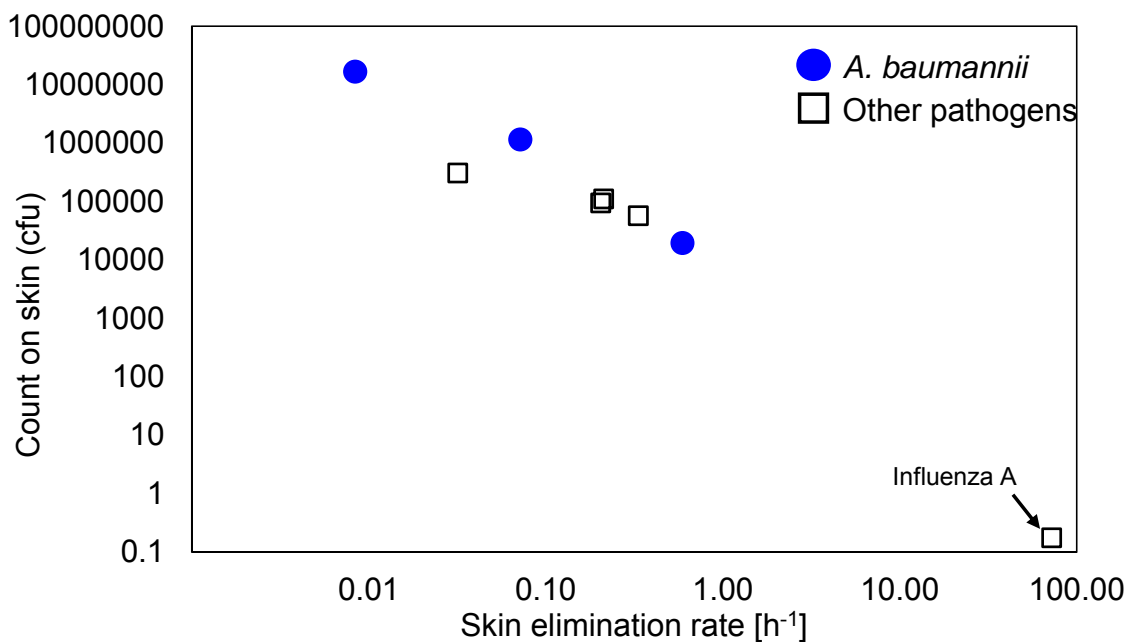
Figure 13. Steady state number of pathogens on patient 2 skin for the "all pathways" scenario versus elimination rate on skin.

Other than influenza A, model outcomes generally overlapped between the different estimates for *A. baumannii* and the remaining pathogens (Figure 12, Figure 13). This follows from the comparable range of elimination rates for the data variability in *A. baumannii* versus the average values for the other pathogens. The substantial variability among different strains and experimental conditions for *A. baumannii* is not surprising, but suggests that measurement uncertainty and strain-differences within a species will introduce substantial variability in the

expected role of environmental surfaces for pathogen transmission. Although similar model results for MRSA exhibited relatively low sensitivity to elimination rates compared to other parameters [45], the uncertainty and variability in elimination will likely impede ability to quantitatively forecast intervention outcomes without more detailed model calibration. Factors such as biofilm formation [118,122], initial composition of medium (e.g., presence of body fluids) [126], influence of environmental conditions (e.g., temperature, humidity) [106,125], size of initial inoculum [115,127], and differences among strain types [114,126], can all contribute to the variability in elimination rate within a species. As a result, differences among similar species and preliminary assessments regarding environmental persistence should be viewed with caution [119,126].
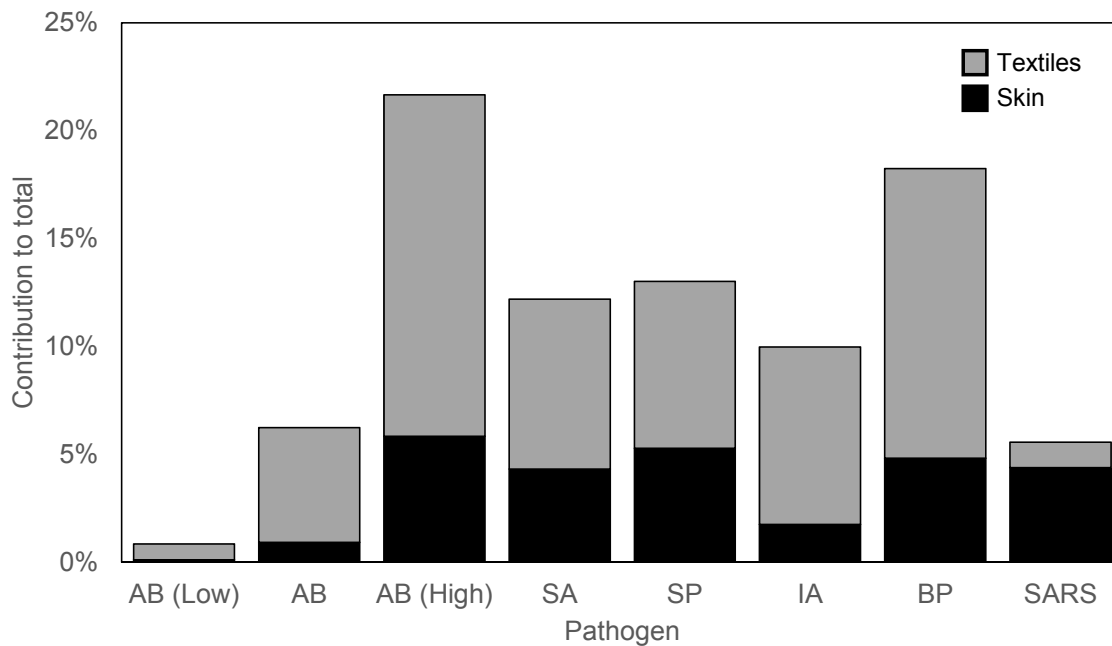
Figure 14. Contribution of different pathways to steady state pathogen count on patent 2. Skin contribution was based on the scenario where textile and nonporous surfaces were excluded. Textile contribution was the added contribution when textiles were added but nonporous surfaces were excluded.

Limitations and future directions

The current model parameterization for multiple pathogens was limited to the environmental elimination rate, with all other parameters standardized to represent a hospital-associated infection released by respiratory pathways and contact with mucous membranes. The value of elimination rates was selected for detailed examination because of its importance for ability of the pathogen to be stored on and transferred via surfaces [106,107,109]. However, the transfer efficiency, initial counts, and shedding rates also vary among pathogens, contributing to differences in exposure even for our simplified scenario. Further review and modeling are warranted to examine how the variation in these parameters will affect infection transmission. Another area for exploration is the relative sensitivity of the surface transmission pathways to different parameter formulations and combinations. Sensitivity analyses by Plipat et al. [45,

Supplemental Information] indicated a high influence of surface transfer parameters for MRSA exposure. Literature variation among pathogens in different parameters could be combined into Monte Carlo simulated model runs to establish which processes are likely to drive the differences among infections in environmental contribution.

This model describes the transfer rate of pathogens in a standardized scenario. In line with the objectives of our study, the transmission pathways in the current model are simplified and incomplete. In particular, release via shedding, vomit, and fecal matter are not included. Additionally, more sophisticated formulations of airborne transmission are available, including release of the pathogen via coughing and sneezing, as well as both near-field and far-field airborne transmission [18,19,111,128]. Further, the model is currently restricted to evaluating the quantity of pathogen (cfu) transferred to an uninfected individual. The model does not provide a basis to evaluate the likelihood of infection, because pathogen-specific and pathway-specific infectious dose are not considered. In fact, different pathogens exhibit different dose–response functions for probability of infection, and it is also likely that the dose–response will differ between transfer mechanisms (e.g., airborne versus dermal transfer). However, these considerations would require additional model complexity, and more parameters, thus introducing additional potential for model uncertainty.

A possible avenue to reduce uncertainty would be to combine this transfer rate modeling approach with epidemiological modeling of infectious disease transmission. Results from a transfer rate model essentially represent the contact rate (c) in epidemiological models of infectious disease transmission. There is an absence of work combining the environmental mechanisms of pathogen transmission with epidemiological infection models. If our model were combined with additional pathogen-specific information, it may be possible to ground-truth results by comparing them with basic reproductive number (R_0) values available in the literature. In a simplified epidemiological infectious disease transmission model, such as that illustrated for gonorrhea by Vynnycky and White [129], the contact rate may be used in combination with other parameters to generate a basic reproductive number (R_0). In particular, $R_0 = c\beta D$, where, c = effective rate of colonies transmitted from an infected individual to a susceptible individual (t^{-1}), β = probability of infection (or colonization) per colony exposed, and D = duration of infectiousness (t). Mechanistically, for transmission among individuals, β represents a dose–response function for conversion to infection [128] as a function of colony density on the individual (e.g., skin or nares). Experimental data on β are limited, although data are known to be available for *Staphylococcus aureus* and group A streptococcus (R. Jones, unpublished data). If R_0 is obtained and compared to literature values, this would be a novel corroboration of mechanistic infectious disease exposure assessment models with infectious disease transmission information.

Acknowledgements

Chapter 3. Integrative statewide assessment of combined environmental and socioeconomic stressors versus chronic disease: California case study [6]

Abstract

The health-risk assessment paradigm is shifting from single stressor evaluation towards cumulative assessments of how multiple stressors affect health outcomes in combination. Recent efforts to develop broad-scale public health hazard datasets provide an opportunity to develop and evaluate multiple exposure hazards in combination. We performed a multivariate study of the spatial relationship between 12 indicators of environmental hazard, 5 indicators of socioeconomic hardship, 2 health outcomes, and a general burden of disease measure. The exposure indicators were obtained from CalEnviroScreen, a publicly available exposure hazard dataset developed by the State of California EPA. These were compared at the zip code scale to population rates of 14 disease categories selected from the International Classification of Diseases ninth edition (ICD-9) among all hospitalized persons (an estimator of disease burden). We performed principal component analysis to reduce the exposure hazard data. Two principal component (PC) axes explained 43% of variance in environmental hazard, with the first axis indicating industrial activity and air pollution, and the second associated with ground-level ozone, drinking water contamination and $PM_{2.5}$. Mass of pesticides used in agriculture was poorly or negatively correlated with all other environmental indicators, and with the CalEnviroScreen metric, suggesting a limited ability of the metric to capture agricultural exposures. Poverty, unemployment, linguistic isolation, and low educational attainment aligned with one PC (56% of variance), representing overall socioeconomic hardship. In simultaneous autoregressive models, all of these PC axes were significantly associated with the disease burden estimator. However, the majority of model variation was explained by the socioeconomic PC. The results of this ecological health study suggest a hypothesis that, compared to environmental pollutant exposure, socioeconomic status more greatly impacts overall burden of disease.

Introduction

In order to protect vulnerable individuals and communities, environmental health science has shifted in emphasis from single stressor evaluation towards integrated assessment of multiple stressors [2,9,29,33,37]. This stems from the knowledge that a wide range of factors, including demography, socioeconomic status, psychosocial stressors, and environmental exposures, all influence health outcomes [28,29,32,38,130,131]. But how to integrate across cumulative and disparate exposures presents a substantial methodological challenge [2,28,32,34]. Accordingly, environmental justice advocates and health geographers are developing a variety of maps, indices, and tools that integrate environmental health hazards from multiple stressors at the ecological scale [37,38,132–136]. These tools incorporate a range of indicators including pollutant concentration or load estimates, contaminated sites or other hazards, built environment measures (e.g., urbanization, industry, or road and traffic density), and population characteristics, such as educational attainment and socioeconomic status. The multiple-indicator assessments combine this information into screening tools to identify geographic

---

regions where vulnerable populations encounter environmental exposures, resulting in health hazard [38,132]. To increase transparency and potential societal benefits, many integrated hazard assessment programs also engage the community at large in tool development and assessment, through community-based participatory research, solicitation of public commentary, and the provision and use of publicly accessible data [137–139]. Some of these integration tools and methods also consider the potential interaction between environmental contributors to risk and the preexisting vulnerability of exposed populations to environmental stressors [29,36,37].

To complement the development of new methods, there is an ongoing need to quantitatively examine existing metrics and tools. Critical and impartial assessment will ensure that integrated assessment metrics are well characterized, technically defensible, and appropriate for intended uses. Further, by analyzing metrics and their underlying data, we can examine the geographic and statistical patterns of public health hazards [35]. For example, the correlation among and between environmental exposures, socioeconomic vulnerability, and health outcomes all warrant investigation. In particular, an understanding of which health stressors (e.g., environmental, social, economic) are most associated with adverse health outcomes can aid in resource allocation and health policy direction across regions and populations [32]. For example, the relative health impact of environmental factors (e.g., pollution) versus population attributes (e.g., socioeconomic status, stress) warrants examination. Multivariate methods (e.g., ordination, principal component analysis) and spatial statistics [140–142] are especially useful methods for these questions given the multivariate and spatial nature of these data and metrics.

An important case study for spatial health hazard evaluation is the California Communities Environmental Health Screening Tool (abbreviated as CalEnviroScreen). CalEnviroScreen is a quantitative hazard metric developed by the State of California Environmental Protection Agency (CalEPA) in an open and public process [35,134]. CalEnviroScreen determines a numeric score based on 19 indicators: 12 measures of environmental exposure, 5 of socioeconomic vulnerability, and 2 of health outcomes [36]. In addition to describing the spatial patterns of hazard, CalEnviroScreen is also intended to help guide state resource allocation. In particular, California Assembly Bill 32 and Senate Bill 535 have established a cap and trade program for greenhouse gas emissions, and require that 10% of the anticipated 1 billion dollars of annual state revenue from this program be allocated to communities identified by CalEPA as having health vulnerabilities. CalEnviroScreen was developed as a metric for identifying these communities [132]. The methodology for developing CalEnviroScreen has been detailed elsewhere, and it has further been shown by the scientific development team to indicate strong racial disparities in environmental and socioeconomic vulnerability [35,36]. The method and some of the underlying assumptions have also been subject to scrutiny as part of a public review process [138]. However, given the ongoing challenges in environmental health hazard assessment [28,37], and the novelty and policy implications of CalEnviroScreen, an independent scientific assessment is also warranted.

We examine here the data underlying CalEnviroScreen and their statistical association with disease burden. Our analysis focuses on the 17 environmental and socioeconomic variables that describe environmental exposures and population vulnerability. We compare these data to an indicator of disease burden at the zip code scale using publicly available hospital discharge

data. We address three questions about the CalEnviroScreen source data: 1. What are the correlation structures and statistical associations? 2. How well is the CalEnviroScreen metric of hazard supported by the statistical associations in the underlying data? 3. Which multivariate hazard indicators statistically predict disease burden? For this last question, we work to generate hypotheses regarding the ability of CalEnviroScreen to address health outcomes and the relative importance of environmental versus socioeconomic factors in determining disease burden.

Data and Methods

Environmental and socioeconomic data

The environmental hazard (12 variables) and socioeconomic vulnerability (5 variables) data were downloaded along with the CalEnviroScreen scores, as an MS Excel spreadsheet file CESUpdateOct2014.xlsx. These data, obtained from the CalEnviroScreen website (http://oehha.ca.gov/ej/ces2.html) on April 6, 2015, have been pre-cleaned and carefully prepared by CalEPA, as described elsewhere [35,36]. The study covers 8036 census tracts centered within 1355 California zip code areas (http://oehha.ca.gov/ej/ces2.html). The 12 environmental hazard variables include ozone levels, concentrations of particulate matter 2.5 μm in diameter or below (hereafter, PM$_{2.5}$), diesel particulate matter concentrations (diesel PM), traffic density, drinking water contamination, active pesticide mass used in agriculture (pesticides), airborne toxic chemical releases, water body impairments, sites hazardous to groundwater, sites targeted for cleanup, hazardous waste sites, and solid waste sites. The 5 socioeconomic vulnerability variables include percent of population either under 10 or over 65 (vulnerable age), low educational attainment, linguistic isolation, poverty, and unemployment. All study variables were obtained based on data collected between 2008 and 2013, except traffic density, which was based on 2004 data. Appendix 3 Table A11 summarizes all variables, providing abbreviations, years represented, original units, and data transformations for this study. Cushing et al. [35] and Faust et al. [36] provide more extensive detail.

Disease burden measure

CalEnviroScreen also includes two specific health outcome variables: asthma and low birth-weight, which are intended to indicate a combination of vulnerability to, and effects of, environmental exposures [36]. We include these two variables when analyzing the multivariate structure of the CalEnviroScreen data. However, we developed a separate disease burden measure as our outcome variable. Our intent was to use a publicly available data source that was not developed by the CalEnviroScreen team and that would broadly indicate burden of diseases having environmental etiology.

We measured disease burden using discharge diagnostic codes (which used the ICD-9-CM schema). We obtained these data for all hospitalizations for a given calendar year using publicly available, de-identified, statewide hospital discharge data from the California Office of Statewide Health Planning and Development, spanning the years 2008-2011. Using these data, we classified hospitalizations by pre-determined ICD-9 diagnostic categories. We determined the sum total number of hospitalizations for 14 disease diagnostic categories representing serious or chronic ailments known to have potential environmental etiology. The 14 categories

were pneumonia, chronic obstructive pulmonary disease (COPD), asthma, myocardial infarction (MI), cerebrovascular accident (CVA), diarrhea, pancreatic cancer, lung cancer, breast cancer, lymphoma, leukemia, depression, schizophrenia, and low birth weight.

In a preliminary analysis, we found that the occurrence of these diseases in hospitalized persons was positively correlated and also correlated with the total number of diagnoses (Appendix 3 Table A12, Figure A4). Further, because the same patient hospitalization could have multiple diagnoses spanning multiple categories, rates for different disease categories are not mutually exclusive. To address these issues, we assembled these data into a single count of total hospitalizations resulting from the 14 diagnostic categories. Thus, a hospitalization event that included more than one diagnostic category was recorded as a single event. Avoiding the separate statistical examination of multiple disease outcomes also reduces the rate of Type I errors.

To calculate rates per person, we summed the count of total hospitalizations at the zip code tabulation area (ZCTA) level, and divided by total population. For the denominator, we used ZCTA population estimates from the 2010 United States Census. We observed high variability in rates for populations below 100 individuals within a ZCTA (see Appendix 3 Figure A5). Therefore, we excluded ZCTAs having populations < 100 individuals from the final analysis to minimize the influence of statistical outliers.

To summarize, the study's "disease burden indicator" is the total rate of hospitalizations associated with at least one of the 14 diagnostic categories selected. Because the same person could in theory be admitted multiple times for the same diagnosis, the rates reported here are only approximations of population disease prevalence. They may be viewed as representing the impact of certain diseases (particularly chronic diseases), since apart from death, hospitalization is generally the most extreme result of any disease process.

Expecting chronic disease burden to be higher among the elderly, we obtained the percent of the population over 65 years old for each ZCTA from the 2010 US Census. This parameter (Over65) was used in addition to the variables derived from CalEnviroScreen as a predictor in statistical models used to explain the disease burden indicator.

Data preparation and spatial alignment

Initial graphical analyses revealed a single extreme outlier value for traffic density. This value far exceeded the range of remaining data on both a linear scale (43,600 vs. 0 to 12,200) and cube root transformed scale (2.92 versus 0 to 1.91; N = 8036 values). Examination of the source data documentation [138] indicated that this extreme value resulted from manual data adjustment to estimate increased traffic density from a 150 m stretch of the Mexico-US port of entry road at San Ysidro West (San Diego, CA). The value of this outlier traffic density variable was unremarkable in relation to other variables, and was removed to avoid model bias due to an influential point.

Hospitalization-related diagnoses were tabulated at the zip code level, census population is at the ZCTA level, and CalEnviroScreen data are available at the census-tract scale. To align these data sets, the following data preprocessing steps were performed using ArcMap v10

41

(ESRI, Redlands, CA): 1. averaging the CalEnviroScreen data at the zip code scale to achieve a consistent analysis scale; 2. linking the CalEnviroScreen and ICD-9 data together; 3. combining with a shapefile of zip code polygons; and 4. standardizing the health outcome data to total population per zip code tabulation area (ZCTA). Appendix 3 further details this spatial alignment methodology.

Statistical methods

Statistical analysis was performed in R (version 3.2.0) [112]. Principal component analysis (PCA) was performed on the correlation matrix using commands prcomp and PCA (R package factoMineR). Census tracts with missing values were removed from the analysis. To examine multivariate patterns of the entire hazard data set, PCA was performed on all 19 CalEnviroScreen variables in combination. Two separate PCAs were also performed on the 12 environmental and the 5 socioeconomic variables to reduce these data to a smaller number of variables representing these two classes of hazards.

Hazard variables were compared to the disease burden indicator using linear models and simultaneously autoregressive models (SAR), the latter employing the R package spdep. SAR is a form of spatial autoregressive model that is appropriate for describing and testing for linear relationships in the presence of spatial autocorrelation [142,143]. Appropriate treatment of spatial autocorrelation was assessed based on Moran plots illustrating no association with spatially lagged means, global Moran's I that was not significant, and for SAR models, a spatial dependence parameter (λ) that was significant via likelihood ratio test [143–145]. Models were selected based on minimizing Akaike Information Criterion (AIC), employing the rule of thumb that given ΔAIC $\geq$ 2, the model with the smallest AIC is most likely [146]. Parameter inclusion was based on ΔAIC and reported p values ($\alpha = 0.05$). Nagelkerke pseudo-R^2 was calculated as a measure of model goodness of fit for SAR models. Analogous to traditional R^2 in meaning (though not directly comparable), the Nagelkerke pseudo-R^2 estimates from 0 to 1 the improvement in proportion of variation explained by the fitted model, versus a null (intercept-only) model [147]. In order to compare the contribution of each parameter to final variation explained by the model, the psuedo-R^2 was compared between the full model and the model with that parameter removed.

Prior to statistical analysis, all variables were transformed to approximate a normal distribution and multivariate linearity required for linear model analysis [141,148]. Transformations included natural log (5 variables), cube root (7 variables), square root (4 variables), and ArcSin square root transformation (drinking water). Only PM$_{2.5}$ did not require transformation (Appendix 3 Table A11). The combined disease burden indicator (DBI) exhibited skewness and long tails (leptokurtic) and standard transformations failed to achieve normally distributed model residuals. Normal residuals were achieved employing a modulus transformation: $\text{sign}(\sqrt{DBI}) * \ln(|\sqrt{DBI}| + 1)$ [149]. The predictor variables for the linear models and SAR were then centered and scaled by subtracting the mean and dividing by SD. This converted the transformed variables (Appendix 3 Table A11) to the same unit normal distributions, such that a comparison of model coefficients would approximately indicate relative contribution of each variable to disease burden [150].

Results

We first describe the multivariate structure of the CalEnviroScreen source data, in order to understand which exposures are associated with each other. We employ principal component analysis to visualize these associations and the prevailing patterns of overall exposure encountered in California. We then present the correlation between the main principal components and the CalEnviroScreen metric, in order to gain an understanding of what this metric indicates. Finally, we examine how well the principal components predict disease burden. We employ the principal components rather than the individual parameters to focus on overall patterns of exposure and vulnerability and to reduce risk of Type I error. The relative importance of environmental versus socioeconomic parameters in the model illustrates which factors most influence disease burden.

Multivariate data structure

Pearson's pairwise correlation coefficients (r) indicate multiple associations for the underlying CalEnviroScreen data (Table 4). Positive pairwise associations are observed among variables related to particulate air pollution and traffic, with diesel PM moderately correlated with $PM_{2.5}$ (r = 0.47) and traffic density (r = 0.46), and toxic release correlated with these three variables (Table 4). Socioeconomic indicators of vulnerability are also positively associated: low educational attainment, linguistic isolation, poverty and unemployment exhibit r values ranging from 0.51 to 0.82, with the exception of linguistic isolation versus unemployment (r = 0.26) (Table 4; Appendix 3 Figure A6). The strongest correlation between socioeconomic and environmental variables is between linguistic isolation and diesel PM (r = 0.46). The strongest negative association among all variables is water body impairments versus ozone (r = −0.33). Low birth weight is weakly correlated with most variables. Asthma is more positively correlated with the socioeconomic variables education (r = 0.47), poverty (0.51), and unemployment (0.43), than with any environmental variables.

Principal component analysis and comparison to CalEnviroScreen

Principal component analyses (PCA) were performed on the entire CalEnviroScreen data set (19 variables) and on the environmental (12 variables) and socioeconomic (5 variables) data separately. For the entire data set, the first three principal components explain 46% of data variability in combination. Examination of the direction of the variables on the first principal component (Figure 15A, horizontal axis) indicates that all variables are to some extent associated with each other except for weak negative associations with pesticides, impaired water bodies, and vulnerable age. The variables with the greatest variance along this axis, based on vector length and direction, are the socioeconomic indicators linguistic isolation, low educational attainment, and poverty. $PM_{2.5}$ and the two health indicators (low birth weight and asthma) are also associated with these socioeconomic indicators. Indicators of urban and industrial pollution and associated hazardous sites score positively with both PC1All and PC2All (Figure 15). These include hazardous waste sites, cleanup sites, toxic releases, traffic, groundwater threat sites, and diesel PM, as well as a weak association with solid waste sites. These variables are negatively correlated with pesticides, which would be expected in rural areas, and with vulnerable age. Examining a biplot of PC2All and PC3All, we see a negative association between the polluted sites (solid waste, groundwater hazards, clean up, impaired

43

water bodies, cleanup sites) and measures of drinking water contamination and ozone. Additionally, motor vehicle and industrial source-associated air pollution (PM$_{2.5}$, toxic release, traffic, and diesel PM) is negatively associated with pesticides.

For the environmental PCA (Figure 15C), the first two principal components explain 42.5% of data variability in combination. Along the axis of the first principal component, explaining 23% of variance, there is an association among most environmental hazards that would occur in urban or industrially polluted settings: PM$_{2.5}$, toxic releases, diesel PM, traffic, hazardous waste sites, cleanup sites, and groundwater threats. These are negatively associated with pesticide use. Based on these associations, this principal component (hereafter referred to as PC1Env) represents general exposure to urban and industrial pollution. Along the second principal component (hereafter, PC2Env), explaining 19.5% of variance, ozone, drinking water contamination, and PM$_{2.5}$ are negatively associated with impaired water bodies and groundwater threats (Figure 15C). This indicates that elevated hazards due to ozone and drinking water contamination will tend to occur in different areas from impaired water bodies or groundwater threats.

For the socioeconomic PCA (Figure 15D), the first two principal components explain 77% of data variance. The first principal component (hereafter PC1Soc) explains 56% of variance and is positively associated with unemployment, poverty, linguistic isolation, and low educational attainment. Thus, this principal component broadly indicates socioeconomic vulnerability. The second principal component (PC2Soc) explains 21% of variance, and is only strongly associated with vulnerable age (Figure 15D). Since this is the only variable associated with PC2Soc, vulnerable age is used below directly as a possible predictor variable for the indicator of disease burden.

The CalEnviroScreen metric is a derived metric calculated as the weighted sum of environmental variables multiplied by the sum of the socioeconomic variables, vulnerable age, asthma, and low birth weight [35,36]. We propose that any valid vulnerability metric should also represent the prevailing gradients in environmental and socioeconomic vulnerability within the population studied. That is, CalEnviroScreen (or any suitable metric) must identify areas exhibiting high hazard from a combination of environmental or socioeconomic exposures. To evaluate the validity of CalEnviroScreen based on this criterion, we calculated the Pearson product-moment correlation coefficient of the CalEnviroScreen score versus the main principal components from each PCA analysis, for all available census tracts (n = 7504). In the all variables PCA, CalEnviroScreen is strongly associated with PC1All (r = 0.91; Figure 16) and not associated with PC2All (r = 0.14). When separate PCAs are performed on the environmental and social data, CalEnviroScreen is strongly associated with PC1Soc (r = 0.81), moderately associated with PC1Env (r = 0.58), weakly associated with PC2Env (r = 0.25), and not associated with PC2Soc (r = 0.10). These results indicate that changes in the predominant gradients underlying the data (PC1All, PC1Soc, PC1Env) are generally captured by the CalEnviroScreen score. Thus, this single metric effectively captures the prevailing gradients in the underlying variability in environmental and socioeconomic exposures.

Table 4. Correlation matrix of underlying data from CalEnviroScreen. Data were transformed and analyzed using pairwise Pearson correlation coefficients (r). Sample size ranged from 7641 to 8035 census tracts. Cell colors correspond to direction and strength of association, with red indicating negative association and blue indicating positive association.

|	Ozone	PM$_{2.5}$	Diesel PM	Traffic density	Drinking water	Pesticides	Toxic release	Cleanup Sites	Groundwater sites	Hazardous waste sites	Water body impairments	Solid waste sites	Age	Education	Linguistic isolation	Poverty	Unemployment	Asthma
PM$_{2.5}$	0.39																	
Diesel PM	−0.08	0.47																
Traffic density	−0.11	0.17	0.46															
Drinking water	0.54	0.38	−0.11	−0.11														
Pesticides	−0.01	−0.12	−0.31	−0.20	0.22													
Toxic release	0.02	0.61	0.52	0.32	0.03	−0.22												
Cleanup Sites	−0.11	0.12	0.24	0.13	−0.02	0.00	0.16											
Groundwater sites	−0.29	−0.02	0.20	0.13	−0.10	0.08	0.05	0.46										
Hazardous waste sites	−0.13	0.15	0.36	0.23	−0.03	−0.06	0.24	0.50	0.35									
Water body impairments	−0.33	−0.25	−0.06	0.03	−0.18	0.12	−0.13	0.12	0.20	0.09								
Solid waste sites	−0.02	0.00	−0.13	−0.03	0.17	0.19	−0.04	0.26	0.28	0.22	0.16							
Age	0.06	−0.10	−0.18	−0.08	0.06	0.09	−0.12	−0.08	−0.03	−0.13	−0.01	0.02						
Education	0.17	0.28	0.27	0.03	0.24	0.04	0.16	0.20	0.16	0.16	−0.12	0.13	−0.04					
Linguistic isolation	−0.01	0.31	0.46	0.20	0.09	−0.11	0.30	0.22	0.19	0.23	−0.09	0.02	−0.12	0.71				
Poverty	0.16	0.20	0.26	0.00	0.19	−0.01	0.05	0.20	0.18	0.16	−0.10	0.10	−0.09	0.82	0.60			
Unemployment	0.25	0.09	0.04	−0.06	0.18	0.03	−0.01	0.07	0.04	0.03	−0.07	0.07	0.02	0.51	0.26	0.58		
Asthma	0.09	0.05	0.18	−0.04	−0.05	−0.06	0.04	0.13	0.13	0.10	−0.05	0.04	0.03	0.47	0.22	0.51	0.43	
Low birth weight	0.07	0.15	0.16	0.05	0.05	−0.11	0.12	0.05	−0.01	0.08	−0.09	0.00	−0.01	0.24	0.16	0.24	0.20	0.28
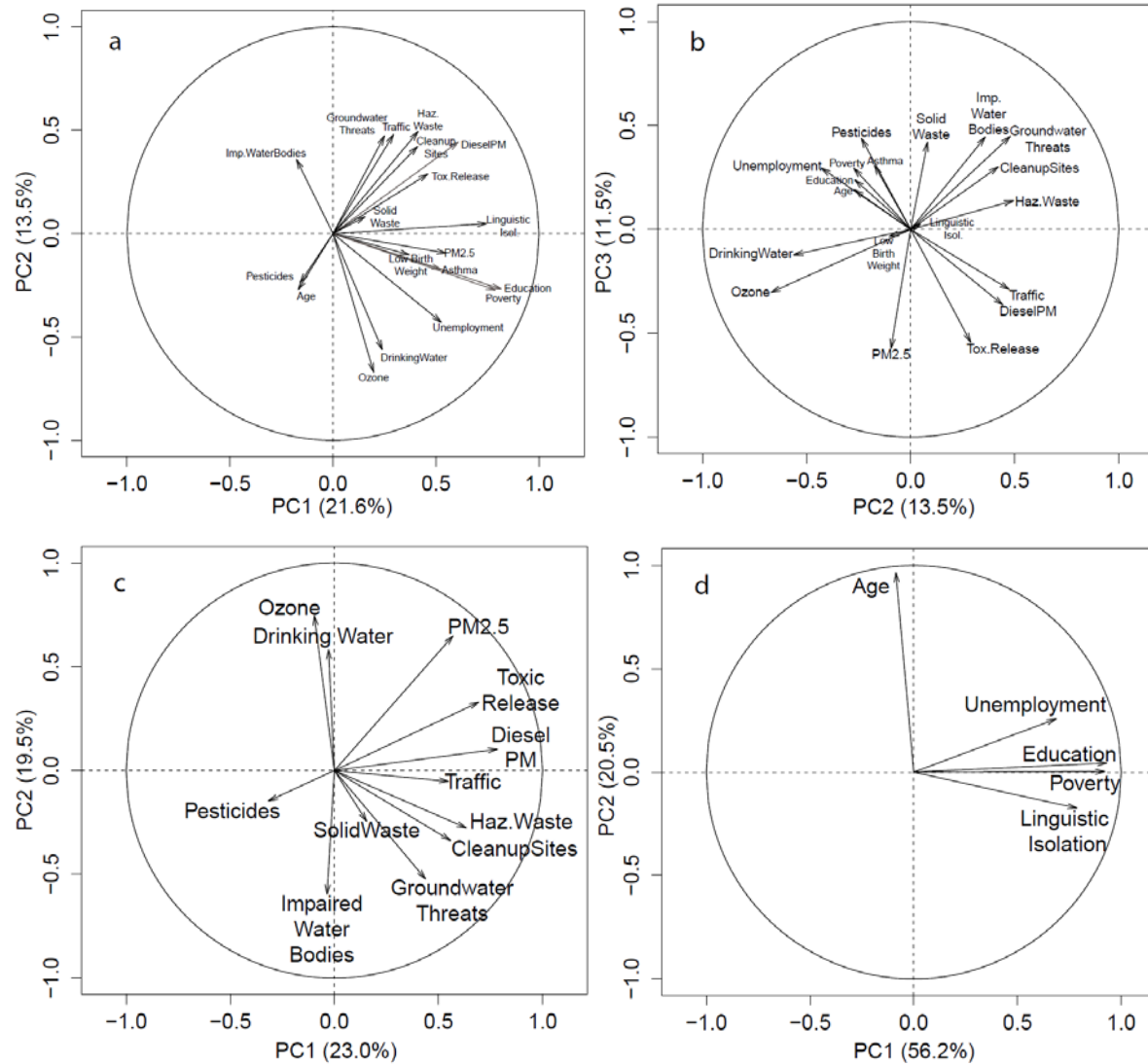
Figure 15. Results of principal component analysis of CalEnviroScreen environmental hazard and socioeconomic vulnerability variables across 7505 populated census tracts in California. Variability explained by individual principal components is in parentheses. A. All variables PC1 versus PC2. B. All variables PC2 versus PC3. C. Environmental variables only. D. Socioeconomic variables only.

Looking at the variables associated with PC2Env and PC2Soc, which were not well correlated with CalEnviroScreen, it is unsurprising to find that across census tracts, the CalEnviroScreen score exhibits very weak correlations with pesticide use (r = 0.04), impaired water bodies (r = 0.00), groundwater threats (r = 0.16), and vulnerable ages (r = 0.02). In particular, pesticide use is negatively associated with many other combined variables, such that CalEnviroScreen ranking will be generally insensitive to this measure.
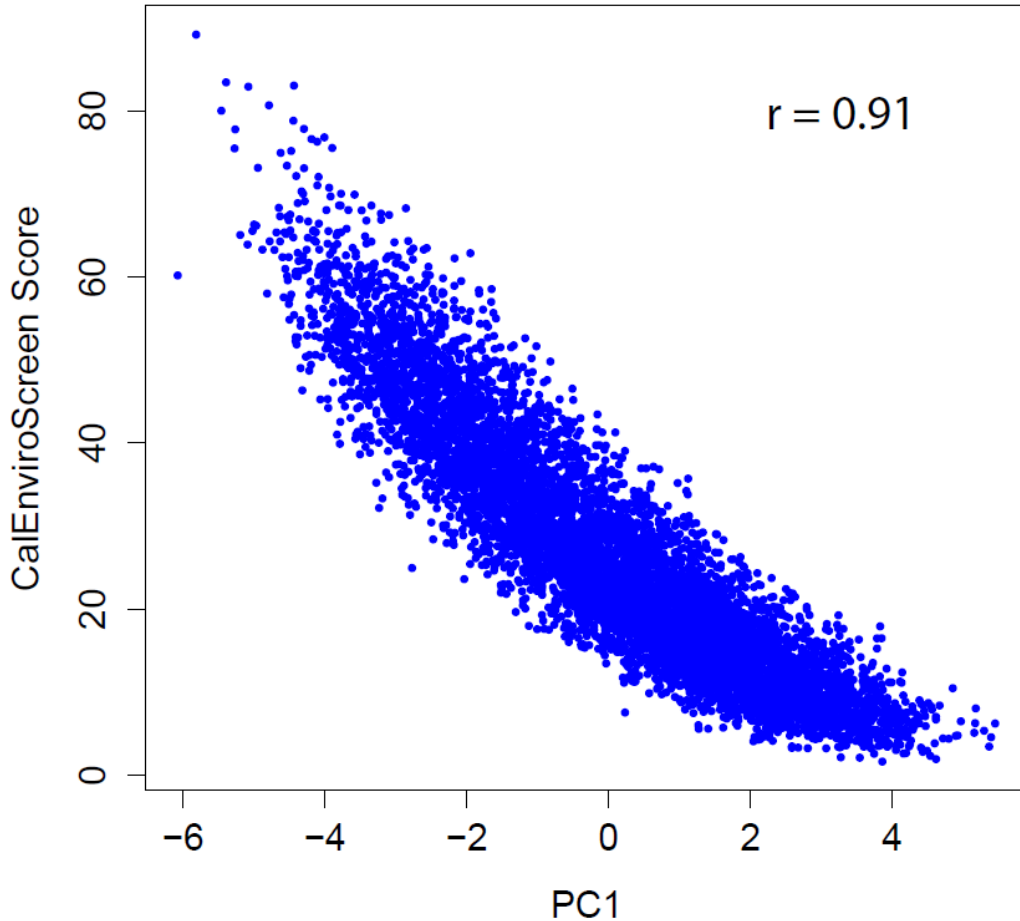
Figure 16. Association between first principal component for all variables (PC1) and CalEnviroScreen 2.0 score [36]. Each point represents a populated California census tract.

Comparison to the disease burden indicator

The results from the all data PCA (PC1All, PC2All, and PC3All) and from the separated environmental and socioeconomic PCA (PC1Env, PC2Env, PC1Soc, and vulnerable age) were evaluated as possible predictors for the disease burden indicator that we developed (hereafter, "disease burden"). The predictor variables were not correlated with each other (| r | ≤ 0.22), with the exception of a weak negative correlation between vulnerable age and PC1Env (r = −0.37). Percent population > 65 years old (hereafter, Over65) was also included as a potential predictor. Over65 was moderately correlated with PC1All (r = 0.49) and PC1Soc (r = −0.42), weakly correlated with PC1Env (r = −0.29), and uncorrelated with the other parameters.

Initial modeling using linear regression to predict disease burden indicated clear evidence of spatial autocorrelation of residuals for all models. Global Moran's I was highly significant ($p < 0.0001$) and Moran's plots exhibited clear upwards trends, indicating that model residuals were predicted by residuals in adjacent areas, even after accounting for predictor variables. Thus, simultaneous autoregressive models (SAR) were used to predict disease burden. In all SAR models, λ (lambda) was significantly greater than zero, Global Moran's I on residuals was not significant, and Moran's plots became flat, all indicating that spatial autocorrelation was accounted for in the model structure.

For the SAR model on the all data PCA results, the following model form was obtained (coefficient estimate ± SE):

Disease Burden = (0.55±0.03)PC1All + (0.15±0.03)PC2All + (0.68±0.02)Over65

For each parameter (PC1All, PC2All, and Over65), ΔAIC > 2 and $p < 0.0001$. The intercept was not significant ($p = 0.98$). PC3All was not significant ($p = 0.26$) and did not substantially improve the explanatory ability of the model (ΔAIC = 0.8). For the final model, the Nagelkerke pseudo-R^2 (hereafter, R^2) is 0.59, which is effectively unchanged when attempting to include PC3All. Comparing Pseudo-R^2 after removing individual parameters illustrates the added contribution of each parameter to explaining the variability in disease burden. Decrease is greater after removing Over65 (new $R^2 = 0.28$) than PC1All (new $R^2 = 0.49$). Decrease is trivial when removing PC2All (0.58). The association with PC1All generally describes the association of disease burden with the correlated variability in all of the CalEnviroScreen hazard indicators except for pesticides and impaired water bodies. The very weak positive association with PC2 suggests that disease burden is more associated with indicators of urban exposures (e.g., traffic, diesel PM) and contaminated sites (e.g., hazardous waste sites, groundwater threats), than with exposure to ozone contamination or drinking water contamination.

For the separate environmental and socioeconomic PCA results ($N = 1223$), all model terms contribute to describing disease burden (coefficient estimate ± SE):

Disease Burden = (0.21±0.03)PC1Env + (0.11±0.04)PC2Env + (0.44±0.03)PC1Soc + (0.69±0.02)Over65

ΔAIC > 2 and individual parameter p-values were < 0.001 for all included model terms, and the intercept was not significant ($p = 0.98$). Addition of VulnerableAge did not improve this model fit, which is unsurprising given the correlation between Over65 and VulnerableAge (Pearson's $r = 0.67$, $N = 1223$). The model coefficient is largest for Over65, which is unsurprising and simply indicates that age must be accounted for in this ecological analysis. The coefficient is greater for PC1Soc (indicating poverty, lack of education, linguistic isolation and unemployment) than for either environmental coefficient. Similarly, the Nagelkerke pseudo-R^2 indicates more variability explained by PC1Soc and especially Over65. Compared to the full model ($R^2 = 0.60$), the pseudo-R^2 declined to a much lower value when dropping Over65 ($R^2 = 0.27$), and moderately declined when removing PC1Soc ($R^2 = 0.52$), but remained essentially unchanged when dropping PC1Env ($R^2 = 0.59$) or PC2Env ($R^2 = 0.60$) from the model. These results indicate that whereas environmental hazard and socioeconomic vulnerability both contribute, socioeconomic vulnerability is considerably more important than environmental hazard for explaining disease burden.

Graphical analysis indicates weak positive associations of disease burden with PC1Soc and a stronger association with Over65 (Figure 17, top panels). The environmental principal components (PC1Env, PC2Env) exhibit very weak curvilinear relationships with disease burden (Figure 17, bottom panels). For PC2Env, a few locations exhibited relatively low exposure and disease burden. Although these results indicate a statistical effect of both environmental and socioeconomic stressors for disease burden, the magnitude of any environmental effect is extremely small.
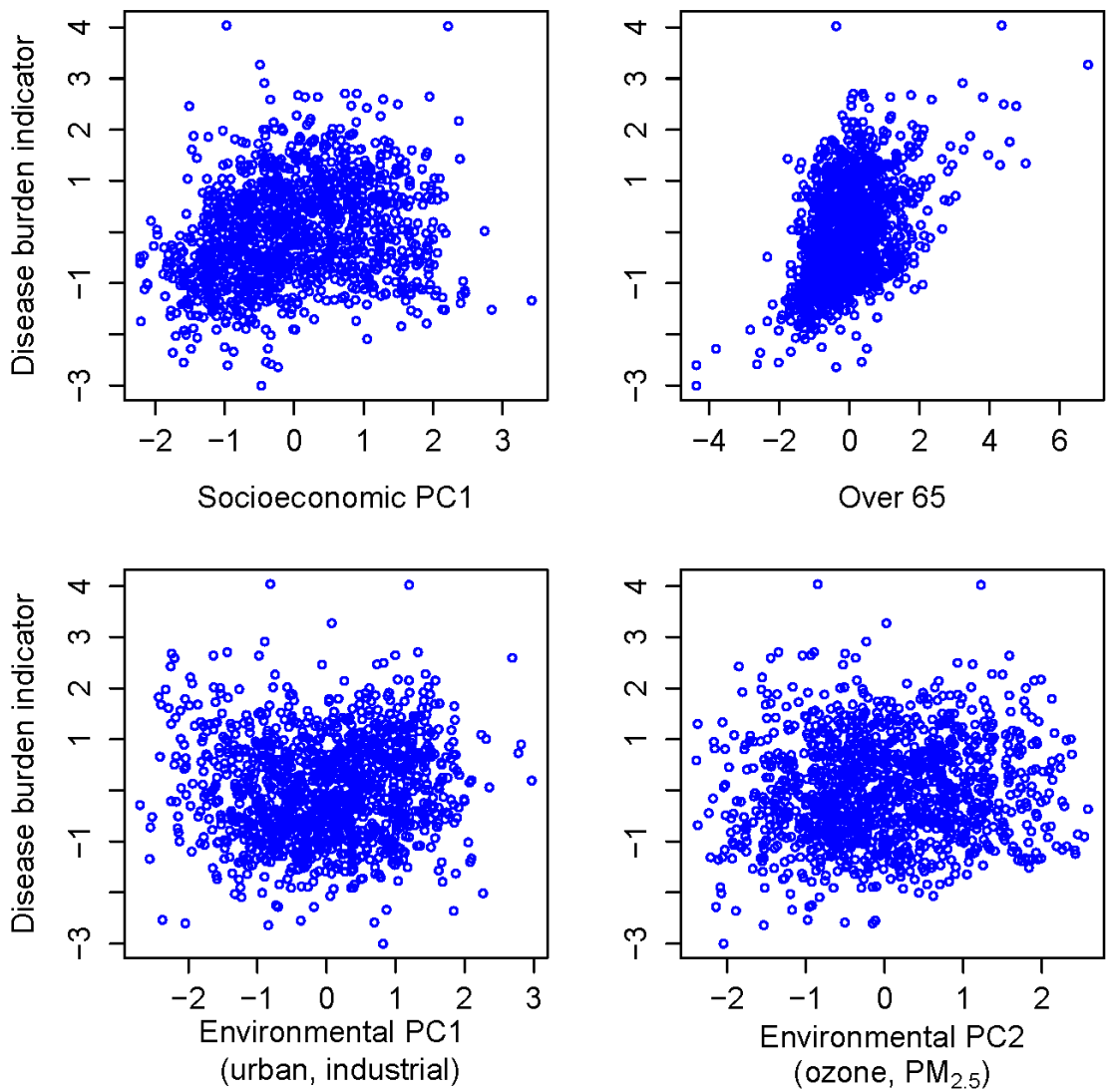
Figure 17. Association between significant predictors and disease burden. All parameters have been scaled and transformed as described in text.

Discussion

Our study results support the use of CalEnviroScreen as an ecological scale indicator of environmental health hazard. First, the CalEnviroScreen metric was strongly associated with the first principal components in all analyses, indicating that it represents the primary underlying gradients within the data set. Second, the principal components from the

CalEnviroScreen data were significantly associated with our metric of disease burden, which is a general indicator of health care burden. Models that contained these variables, and also accounted for spatial autocorrelation and proportion of population that was over 65, explained approximately 60% of the variation in the underlying data. Given all of the uncertainty and assumptions with the study data and scale, this suggests that the CalEnviroScreen data accurately represents a combination of multiple exposure hazards that combine to influence burden of disease. Our analysis, therefore, suggests that CalEnviroScreen is an appropriate tool for its intended purpose: to identify vulnerable communities for resource allocation in environmental health restoration [132].

The first few principal components explained limited variability in the underlying data set, and many of the parameters, especially environmental exposure measures, were weakly correlated. These observations indicate that for 19 hazard parameters captured in CalEnviroScreen, there will be many exposure combinations that are not fully described by shared multivariate gradients. Further examination of the statistical properties and demographic vulnerability of sites exhibiting unique exposure combinations is warranted. Additionally, some policy interventions may best be geographically targeted using additional information beyond the CalEnviroScreen metric itself. For example, pesticides were uncorrelated or negatively associated with most environmental hazard indicators, and poorly correlated to CalEnviroScreen. Most of the hazards measured were associated with urbanization and industrial activities, whereas the pesticide measure largely derived from agricultural application [36]. Given the weight of evidence linking pesticide exposure to a variety of developmental and health effects [151–153], it may be appropriate to examine the pesticide variable or other measures of agricultural exposure hazard alongside the CalEnviroScreen score. The negative association of ozone air pollution with groundwater sites and water body impairments is not readily explained but again suggests that residents of different regions encounter different exposure hazards.

When multivariate methods were employed to separate out chemical pollutant exposure hazard versus socioeconomic variation within California, there was a stronger association of disease burden with socioeconomic status than with environmental pollution exposure. This supports the paradigm that underlying population vulnerability, resulting from socioeconomic conditions, must be considered in health risk assessment [28,29,133]. This finding is implicitly acknowledged in that epidemiological investigations routinely attempt to correct for effect measure modifiers. Further study using more robust methods, including multilevel models and longitudinal analysis, may be warranted to more explicitly evaluate whether the stronger influence of socioeconomic status (compared to pollutant hazards) on disease burden is specific to the current case study or reflects a generalizable pattern.

The multivariate and exploratory approach of our study reflects objectives quite different from a traditional epidemiological comparison of two variables while accounting for confounding factors. We identified prevailing gradients of exposure and vulnerability, and observed how these patterns were associated with disease burden. We observed relatively strong associations among all of the socioeconomic indicators (education, income, unemployment, linguistic isolation), each of which may exhibit a separate impact on vulnerability [see also 134]. This could explain the stronger association between socioeconomic indicators and disease burden, in contrast to environmental hazards, which were less correlated, such that the gradients in multivariate exposures were weaker. In other words, our data suggest a hypothesis that

population disease burden will be more strongly impacted when multiple stressors occur in combination. As such, examination of the multivariate association among stressors should provide added and complementary information to bivariate analyses of exposure versus outcome. The generalizability of our finding that socioeconomic factors better explain disease burden than environmental hazards merits investigation, as it would have implications for intervention priorities, as well as for conceptualization of the primary structuring factors that influence disease.

Like all ecological-scale studies of publicly available spatial exposure and health data, this study has limitations. This study does not establish causality and we cannot extrapolate inferences to the individual level [154]. For their similar study of the San Joaquin Valley region, Huang and London [135] thoroughly discuss limitations of studies using publicly-available spatial exposure data. Preexisting data are employed, with some variation in the time range covered [36]. Data required geographic alignment, including assembly of different parameters provided at multiple and varying spatial scales. In particular, CalEnviroScreen data were available at the census-tract level, the disease burden indicator at the USPS zip code level, and spatial polygon arrangement at zip code tabulation area-level. Inaccuracies are inevitably introduced when aligning these different spatial scales [155]. In line with the protection of individual rights to anonymity in publicly accessible outcome data, individual-level demographic information was masked, and residential addresses were limited to USPS zip code.

These factors likely in part explain the limited strength of associations observed in this study. However, studies at similar spatial scales and resolutions have established relationships of environmental hazards and disease risk with race and socioeconomic status, with implications for resource allocation and policy [35,136,156]. Given that, a novel aspect of our study is that we explicitly evaluate and describe the multiple patterns of association that occur across a range of health hazards at the ecological scale. We observed that socioeconomic indicators were associated with each other and contributed to explaining disease burden, and that an environmental gradient of urban and industrial pollution also contributed to explaining disease burden. In contrast, ground-level ozone and drinking water threats were negatively associated with impaired water bodies and groundwater threats, and offered little predictive value for disease burden. Some of these findings corroborate findings from an analysis for CalEnviroScreen development using 30 zip codes in California [134]. Prior studies have also shown geographic indicators of socioeconomic status to be associated with hospitalization rates [38]. The existence of separate gradients of environmental and socioeconomic disparity, and the varying ability to predict disease burden highlight the need for continued emphasis on integrated approaches in vulnerability assessment.

Acknowledgements

Conclusions

The work in this thesis intersects along the common approach of building novel exposure models to address current and emerging threats to human health. I conclude my dissertation by comparing findings across the three chapters to examine the common attributes across the three modeling activities. After providing a summary of the main findings, I discuss potential future research directions for each chapter topic.

The models in this study share a set of qualitative attributes that are characteristic of all models. These attributes include the conceptual model, scale of analysis, type of computational model (statistical versus mechanistic), underlying mathematical framework, and the inclusion and exclusion of relevant processes (Table 1). In each study, the underlying conceptual model is geared towards answering a highly specific question. For example, the conceptual model underlying Chapter 1 depicts selection for resistance as a result of the chemical-dependent change in net growth difference between two competing bacterial strains. This answers the question of how the modeled processes (e.g., antibiotic dose–response, differences in intrinsic growth rate) affect concentration-dependent difference in growth rates. As another example of the conceptual model targeted to the question, Chapter 2 depicts a simplified scenario of microbe transmission to address the importance of differences in elimination rate for predicting transmission. Table 1 also illustrates how each model includes a wide range of processes but excludes an even wider range (Table 1). Whether this invalidates the model amounts to whether appropriate choices were made in what to include versus exclude.

Table 1. Summary and comparison of the models developed in this thesis.

Attribute	Chapter 1	Chapter 2	Chapter 3
Question and conceptual model	How do antibiotic or metal concentrations and strain attributes affect the net growth difference between two competing bacterial strains?	How does microbial pathogen elimination rate on skin, textiles, and nonporous surfaces affect pathogen transmission among patients residing in separate hospital rooms?	How well do the multivariate patterns in environmental hazards and socioeconomic stressors predict burden of disease?
Scale	Microbes	Individual humans	Human populations in geographic regions (ecological study)
Model type	Mechanistic	Mechanistic	Statistical
Mathematical framework	Analytical equations	Linear algebra matrix	Principal component analysis; spatial autoregression
Processes included	Competition of bacteria; antibiotic resistance; chemical versus growth dose–response; minimum selective concentration (MSC)	Release of microbes from mucous membranes onto air and skin; dermal contact; environmental elimination	Exposure hazard due to air pollution, water pollution, and contaminated sites; population vulnerability due to age, education, linguistic isolation, and poverty
Processes excluded (partial listing)	Horizontal gene transfer by plasmids, which can be antibiotic concentration independent; community dynamics with more than two strains or species; spatial heterogeneity, including biofilms; short-term growth dynamics	Infection, including the infectious dose–response, vulnerability of host, and effect of vehicle of transmission and exposure route; pathways other than dermal transport (e.g., droplet spray, vomit, fecal-oral)	Variation between individuals, such as stress and psychological well-being or change in residence (ecological study); residential built environment; diet and food environment; chronic diseases that don't require hospital visits

Each model was combined with external data and subjected to a sensitivity analysis (Table 2). The sensitivity analyses method varied among studies, and in the case of Chapter 3, the term "sensitivity analysis" simply refers to calculating statistical influence of the different exposures on disease burden. A combination of model limitations and uncertainties was identified, potentially warranting further investigation. For example, in Chapter 3, the CalEnviroScreen calculation of hazard was identified to be highly insensitive to agricultural pesticide use, which is arguably the only agricultural exposure hazard considered (Table 2). This raises a potential concern regarding insufficient consideration of the exposed agricultural worker population in California [151–153]. Finally, each model evaluation supported some existing or novel hypotheses (Table 2).

Table 2. Summary of the model evaluation activities and findings in this thesis.

Evaluation activity or finding	Chapter 1	Chapter 2	Chapter 3
Data used	Intrinsic growth rate, difference in strain growth rates, selection coefficient representing fitness cost for resistant strains, minimum inhibitory concentrations (MIC) of sensitive and resistant strains	Contact rate, contact area, transfer efficiency, coughing and breathing rates; disease specific elimination rates	12 environmental exposure measures, 5 socioeconomic status measures, average age, geospatial location of census tracts and zip code tabulation areas, and hospitalization frequency for 14 diseases
Forecast simulations performed	Predicted the MSC/MIC ratio for published laboratory growth data	Compared number of organisms transmitted including versus excluding surfaces for six microbial pathogens	None: forecasting not appropriate for this statistical analysis of ecological data
Sensitivity analysis method	Vary parameter individually, holding other parameters constant. Monte Carlo Simulation of selected parameters.	Compare how variability within a species versus across species influenced transmission	Compare effect size of standardized predictor variables on disease burden
Sensitivity analysis findings	MSC/MIC ratio most sensitive to selection coefficient and shape of the sensitive strain's chemical versus growth rate dose–response function	Range of predicted transmission for one species (*Acinetobacter baumannii*) overlapped with four other species	Socioeconomic stressors more important than environmental hazards for disease burden
Model limitations identified	A curve fit to growth rate differences performed poorly for two aminoglycoside antibiotics	The model evaluation did not identify any model limitations	CalEnviroScreen model was insensitive to agricultural pesticide use
Uncertainties identified for further investigation	Shape of antibiotic versus growth dose–response curve at subinhibitory concentrations	Elimination rate differences between nonporous surfaces, textiles, and skin could influence intervention strategies	None identified
Hypotheses generated or supported	For many human-impacted settings, there is a hazard of selection for resistant strains. For aminoglycosides, adaptive resistance may bring MSC close to MIC.	Environmental surfaces will be important for transfer of many pathogenic microbe species. Transfer via health-care workers between patient rooms is a hazard warranting intervention.	Socioeconomic stressors may be a greater driver of disease burden than environmental exposures. Multivariate gradients of combined stressors may explain burden of disease differently from bivariate comparisons.

Reviewing the studies individually, I observed that the model of the minimum selective concentration (Chapter 1) provided a simplified depiction of the conditions under which antibiotic resistance will develop. Model performance was variable in its ability to fit growth rate difference data, suggesting that additional formulations could be considered. Refinement and testing of this mathematical model could contribute to the goal of forecasting high-risk conditions for dissemination of antibiotic resistance, and ultimately setting priorities among natural, veterinary, and medical sources of resistance. In the future, collaborations among modelers, environmental chemists, and microbiologists may be fruitful to address several questions: 1. How robust is the model to different combinations of species, antibiotic, and environmental condition? 2. How readily, and at what concentrations, does bacterial resistance spread across multiple species within a bacterial community? and 3. How well do mathematical predictions such as this model correspond to ecosystem observations on antibiotic concentrations versus resistance? There is also a need to broaden the processes considered in mechanistic models of antibiotic resistance development. In particular, horizontal gene transfer (HGT) is an important process, and the rate and importance of HGT versus competition should be considered.

The model of pathogenic microbe transmission in indoor settings (Chapter 2) supports the paradigm that among most hospital-acquired infections, there is the potential for environmental transmission of live pathogens among patients residing in separate hospital rooms. In five of six taxa (*Streptococcus pyogenes*, *Bordetella pertussis, Acinetobacter baumannii*, MRSA, and SARS), the model predicted a high rate of transmission among patients in separate rooms due to a health-care worker in contact with surfaces. Further, the combined uncertainty and variability in published elimination rates for *Acinetobacter baumannii* was sufficient to cause the predicted colony count to overlap with four of the five remaining pathogens. Only influenza A exhibited low transfer via surfaces and dermal pathways, due to its consistently low environmental persistence. These results support continued vigilance in surface decontamination protocols, as well as barrier controls and hand washing. A collaboration among exposure scientists, epidemiologists and clinicians could be fruitful to test model predictions regarding intervention effectiveness. Nevertheless, the model-data linkage warrants expansion, including available pathogen-specific data on rates of transfer among skin and surfaces, airborne persistence (e.g., settling rates), and especially dose–response for infection (e.g., infectious inoculum). Predictions of the model should also be compared to existing literature by examining long-term health care system data sets and results of controlled interventions. Another area warranting development is expansion of the pathogen classification framework to include pathogens with differing transmission pathways (e.g., fecal-oral) and ultimately to other pathogen types and scenarios, such as foodborne and waterborne transmission.

Chapter 3 makes use of statistical models to compare the CalEnviroScreen hazard assessment dataset to public records of hospital diagnosis frequency. Overall disease burden was more strongly associated with socioeconomic indicators of hardship than with indicators of environmental pollutant exposure. The multivariate and geostatistical modeling approaches in Chapter 3 are useful to integrate different kinds of exposures within the Eco-Exposome [2,3].

This approach, employing publicly available data, demonstrates the opportunities for geospatial hazard assessment in the context of environmental justice. In the future, this approach merits applications in different regions and may be useful for comparing the spatial distribution of existing versus forecasted exposure hazards. An example would be vulnerability assessment of areas with existing and planned oil and gas development, which is an established concern in California [157] and globally. Finally, given the limitations of the ecological approach in Chapter 3, an obvious extension is to integrate multilevel exposure data, combining diverse exposures at both ecological and individual scales.

References

1.	Box GEP. Robustness in the strategy of scientific model building. Army Research Office Workshop on Robustness in Statistics April 11-12, 1978. Research Triangle Park, NC; 1979. pp. 201–236. doi:0-12-4381 50-2

2.	National Research Council. Exposure Science in the 21st Century: A Vision and a Strategy. The National Academies Press; 2012. Available: http://www.nap.edu/openbook.php?record_id=13507

3.	Lioy PJ, Smith KR. A discussion of exposure science in the 21st century: A vision and a strategy. Environ Health Perspect. 2013;121: 405–409. doi:10.1289/ehp.1206170

4.	Reis S, Seto E, Northcross A, Quinn NWT, Convertino M, Jones RL, et al. Integrating modelling and smart sensors for environmental and human health. Environ Model Softw. 2015;74: 238–246. doi:10.1016/j.envsoft.2015.06.003

5.	Lioy PJ, Rappaport SM. Exposure science and the exposome: an opportunity for coherence in the environmental health sciences. Environ Health Perspect. 2011;119: A466–7. doi:10.1289/ehp.1104387

6.	Wambaugh JF, Wang A, Dionisio KL, Frame A, Egeghy P, Judson R, et al. High throughput heuristics for prioritizing human exposure to environmental chemicals. Environ Sci Technol. 2014;48: 12760–7. doi:10.1021/es503583j

7.	Committee on Models in the Regulatory Decision Process. Models in Environmental Regulatory Decision Making. National Research Council, National Academies Press; 2007. doi:10.17226/11972

8.	U.S. EPA. Guidance on the Development, Evaluation and Application of Environmental Models. Washington, DC; 2008.

9.	Wild CP. The exposome: From concept to utility. International Journal of Epidemiology. 2012. pp. 24–32. doi:10.1093/ije/dyr236

10.	Rappaport SM. Implications of the exposome for exposure science. J Expo Sci Environ Epidemiol. 2011;21: 5–9. doi:10.1038/jes.2010.50

11.	Mattingly CJ, McKone TE, Callahan MA, Blake JA, Hubal EAC. Providing the missing link: the exposure science ontology ExO. Environ Sci Technol. 2012;46: 3046–53. doi:10.1021/es2033857

12.	Spear R. Mathematical modeling in environmental health. Environ Health Perspect. 2002;110: A382.

13.	McKone TE, Ryan PB. Human exposures to chemicals through food chains: an uncertainty analysis. Environ Sci Technol. 1989;23: 1154–1163. doi:10.1021/es00067a014

14.	Lioy PJ. Exposure science: A view of the past and milestones for the future. Environ Health Perspect. 2010;118: 1081–1090. doi:10.1289/ehp.0901634

15. Georgopoulos PG, Lioy PJ. From a theoretical framework of human exposure and dose assessment to computational system implementation: the Modeling ENvironment for TOtal Risk Studies (MENTOR). J Toxicol Environ Health B Crit Rev. 2006;9: 457–83. doi:10.1080/10937400600755929

16. Rosenbaum RK, Bachmann TM, Gold LS, Huijbregts MAJ, Jolliet O, Juraske R, et al. USEtox-the UNEP-SETAC toxicity model: recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. Int J Life Cycle Assess. 2008;13: 532–546. doi:10.1007/s11367-008-0038-4

17. McKone TE, MacLeod M. Tracking multiple pathways of human exposure to persistent multimedia pollutants: regional, continental, and global-scale models. Annu Rev Environ Resour. 2003;28: 463–492. Available: http://www.annualreviews.org/doi/abs/10.1146/annurev.energy.28.050302.105623

18. Nicas M, Jones RM. Relative contributions of four exposure pathways to influenza infection risk. Risk Anal. 2009;29: 1292–1303. doi:10.1111/j.1539-6924.2009.01253.x

19. Nicas M, Sun G. An integrated model of infection risk in a health-care environment. Risk Anal. 2006;26: 1085–1096. doi:10.1111/j.1539-6924.2006.00802.x

20. Hays SM, Becker RA, Leung HW, Aylward LL, Pyatt DW. Biomonitoring equivalents: A screening approach for interpreting biomonitoring results from a public health risk perspective. Regul Toxicol Pharmacol. 2007;47: 96–109. doi:10.1016/j.yrtph.2006.08.004

21. Oreskes N, Shrader-Frechette K, Belitz K. Verification, validation, and confirmation of numerical models in the earth sciences. Science. 1994;263: 641–646. doi:10.1126/science.263.5147.641

22. Oreskes N. Evaluation (not validation) of quantitative models. Environ Heal Perspect. 1998;106: 1453–1460.

23. Beck MB, Ravetz JR, Mulkey LA, Barnwell TO. On the problem of model validation for predictive exposure assessments. Stoch Hydrol Hydraul. 1997;11: 229–254. doi:10.1007/BF02427917

24. Walker WE, Harremoës P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, et al. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr Assess. 2003;4: 5–17. doi:10.1076/iaij.4.1.5.16466

25. Matott LS, Babendreier JE, Purucker ST. Evaluating uncertainty in integrated environmental models: A review of concepts and tools. Water Resour Res. 2009;45: 1–14. doi:10.1029/2008WR007301

26. Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. Science. 2010;330: 460–461. doi:10.1126/science.1192603

27. Krieger N. Epidemiology and the web of causation: Has anyone seen the spider? Soc Sci Med. 1994;39: 887–903. doi:10.1016/0277-9536(94)90202-X

28. Briggs D. Making a Difference: Indicators to Improve Children's Environmental Health. Geneva, Switzerland: World Health Organization Geneva; 2003.

29. DeFur PL, Evans GW, Cohen Hubal EA, Kyle AD, Morello-Frosch RA, Williams DR. Vulnerability as a function of individual and group resources in cumulative risk assessment. Environ Health Perspect. 2007;115: 817–824. doi:10.1289/ehp.9332

30. Schettler T. Toward an Ecological View of Health: An Imperative for the Twenty-First Century. The Center for Health Design, Health Care Without Harm, and the Robert Wood Johnson Foundation; 2006.

31. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PLoS One. 2010;5: e10746. doi:10.1371/journal.pone.0010746

32. National Research Council. Science and Decisions: Advancing Risk Assessment. Washington, DC; 2009. Available: http://www.nap.edu/catalog.php?record_id=12209

33. Guth JH. Cumulative impacts: Death-knell for cost-benefit analysis in environmental decisions. Barry L Rev. HeinOnline; 2008;11: 23–57.

34. Callahan MA, Sexton K. If cumulative risk assessment is the answer, what is the question? Environ Health Perspect. 2007;115: 799–806. doi:10.1289/ehp.9330

35. Cushing L, Faust J, August LM, Cendak R, Wieland W, Alexeeff G. Racial/ethnic disparities in cumulative environmental health impacts in California: evidence from a statewide environmental justice screening tool (CalEnviroScreen 1.1). Am J Public Health. 2015; e1–e8. doi:10.2105/AJPH.2015.302643

36. Faust J, August L, Alexeeff G, Bangia K, Cendak R, Cheung-Sutton E, et al. California Communities Environmental Health Screening Tool, Version 2.0 (CalEnviroScreen 2.0) Guidance and Screening Tool. Sacramento, CA; 2014. Available: http://oehha.ca.gov/ej/ces2.html

37. Scammell MK, Montague P, Raffensperger C. Tools for addressing cumulative impacts on human health and the environment. Environ Justice. 2014;7: 102–109. doi:10.1089/env.2014.0016

38. Roth R, Barsi E. The "Community Need Index." A new tool pinpoints health care disparities in communities throughout the nation. Heal Prog. 2005;86: 32–38.

39. Auchincloss AH, Gebreab SY, Mair C, Diez Roux A V. A review of spatial methods in epidemiology, 2000-2010. Annu Rev Public Health. 2012;33: 107–22. doi:10.1146/annurev-publhealth-031811-124655

40. Center for Disease Control. Antibiotic resistance threats in the United States, 2013. 2013.

41. Snary EL, Kelly LA, Davison HC, Teale CJ, Wooldridge M. Antimicrobial resistance: a microbial risk assessment perspective. J Antimicrob Chemother. 2004;53: 906–917. doi:10.1093/jac/dkh182

42. Ashbolt NJ, Amézquita A, Backhaus T, Borriello P, Brandt KK, Collignon P, et al. Human Health Risk Assessment (HHRA) for environmental development and transfer of antibiotic resistance. Environ Health Perspect. 2013;121: 993–1001. doi:10.1289/ehp.1206316

43. Sandegren L. Selection of antibiotic resistance at very low antibiotic concentrations. Ups J Med Sci. 2014; 1–5. doi:10.3109/03009734.2014.904457

44. Andersson DI, Hughes D. Evolution of antibiotic resistance at non-lethal drug concentrations. Drug Resist Updat. 2012;15: 162–172. doi:10.1016/j.drup.2012.03.005

45. Plipat N, Spicknall IH, Koopman JS, Eisenberg JN. The dynamics of methicillin-resistant Staphylococcus aureus exposure in a hospital model and the potential for environmental intervention. BMC Infect Dis. 2013;13: 595. doi:10.1186/1471-2334-13-595

46. Li S, Eisenberg JNS, Spicknall IH, Koopman JS. Dynamics and control of infections transmitted from person to person through the environment. Am J Epidemiol. 2009;170: 257–65. doi:10.1093/aje/kwp116

47. Angulo FJ, Nargund VN, Chiller TC. Evidence of an association between use of anti-microbial agents in food animals and anti-microbial resistance among bacteria isolated from humans and the human health consequences of such resistance. J Vet Med Ser B-Infectious Dis Vet Public Heal. 2004;51: 374–379.

48. Molbak K. Spread of resistant bacteria and resistance genes from animals to humans - the public health consequences. J Vet Med Ser B-Infectious Dis Vet Public Heal. 2004;51: 364–369.

49. Martinez JL. Environmental pollution by antibiotics and by antibiotic resistance determinants. Environ Pollut. 2009;157: 2893–2902. doi:10.1016/j.envpol.2009.05.051

50. Lupo A, Coyne S, Berendonk TU. Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. Front Microbiol. 2012/02/04 ed. 2012;3: 18 pp. doi:10.3389/fmicb.2012.00018

51. Sarmah AK, Meyer MT, Boxall ABA. A global perspective on the use, sales, exposure pathways, occurrence, fate and effects of veterinary antibiotics (VAs) in the environment. Chemosphere. 2006;65: 725–759. doi:http://dx.doi.org/10.1016/j.chemosphere.2006.03.026

52. Tello A, Austin B, Telfer TC. Selective pressure of antibiotic pollution on bacteria of importance to public health. Environ Health Perspect. 2012/05/11 ed. 2012;120: 1100–1106. doi:10.1289/ehp.1104650

53. Zhao X, Drlica K. Restricting the selection of antibiotic-resistant mutant bacteria: measurement and potential use of the mutant selection window. J Infect Dis. 2002;185: 561–565. doi:10.1086/338571

54. Drlica K, Zhao X. Mutant selection window hypothesis updated. Clin Infect Dis. 2007;44: 681–8. doi:10.1086/511642

55. Firsov AA, Smirnova M V., Lubenko IY, Vostrov SN, Portnoy YA, Zinner SH. Testing the mutant selection window hypothesis with *Staphylococcus aureus* exposed to daptomycin and vancomycin in an *in vitro* dynamic model. J Antimicrob Chemother. 2006;58: 1185–1192. doi:10.1093/jac/dkl387

56. Cui J, Liu Y, Wang R, Tong W, Drlica K, Zhao X. The mutant selection window in rabbits infected with *Staphylococcus aureus*. J Infect Dis. 2006;194: 1601–1608. doi:10.1086/508752

57. Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, et al. Selection of resistant bacteria at very low antibiotic concentrations. PLoS Pathog. 2011;7: e1002158. doi:10.1371/journal.ppat.1002158

58. Liu A, Fong A, Becket E, Yuan J, Tamae C, Medrano L, et al. Selective advantage of resistant strains at trace levels of antibiotics: A simple and ultrasensitive color test for detection of antibiotics and genotoxic agents. Antimicrob Agents Chemother. 2011;55: 1204–1210. doi:10.1128/AAC.01182-10

59. Jørgensen KM, Wassermann T, Jensen PØ, Hengzuang W, Molin S, Høiby N, et al. Sublethal ciprofloxacin treatment leads to rapid development of high-level ciprofloxacin resistance during long-term experimental evolution of *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 2013;57: 4215–4221. doi:10.1128/AAC.00493-13

60. Day T, Huijben S, Read AF. Is selection relevant in the evolutionary emergence of drug resistance? Trends Microbiol. 2015;23: 126–133. doi:10.1016/j.tim.2015.01.005

61. Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. MBio. 2014;5: 19–23. doi:10.1128/mBio.01918-14

62. Fleming GT, McCarthy DM, Colombet N, Patching JW. The effect of levofloxacin concentration on the development and maintenance of antibiotic-resistant clones of *Escherichia coli* in chemostat culture. J Ind Microbiol Biotechnol. 2002;29: 155–62. doi:10.1038/sj.jim.7000295

63. Tattevin P, Basuino L, Chambers HF. Subinhibitory fluoroquinolone exposure selects for reduced beta-lactam susceptibility in methicillin-resistant Staphylococcus aureus and alterations in the SOS-mediated response. Res Microbiol. 2009;160: 187–192. Available: http://www.sciencedirect.com/science/article/pii/S0923250809000254

64. Bengtsson-Palme J, Larsson DGJ. Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. Environ Int. 2016;86: 140–149. doi:10.1016/j.envint.2015.10.015

65. Regoes RR, Wiuff C, Zappala RM, Garner KN, Baquero F, Levin BR. Pharmacodynamic functions: a multiparameter approach to the design of antibiotic treatment regimens. Antimicrob Agents Chemother. 2004;48: 3670–6. doi:10.1128/AAC.48.10.3670-3676.2004

66. Opatowski L, Guillemot D, Boelle PY, Temime L. Contribution of mathematical modeling to the fight against bacterial antibiotic resistance. Curr Opin Infect Dis.

2011;24: 279–287. doi:10.1097/QCO.0b013e3283462362

67. Nielsen EI, Cars O, Friberg LE. Pharmacokinetic/pharmacodynamic (PK/PD) indices of antibiotics predicted by a semimechanistic PKPD model: a step toward model-based dose optimization. Antimicrob Agents Chemother. 2011;55: 4619–30. doi:10.1128/AAC.00182-11

68. Komp Lindgren P, Marcusson LL, Sandvang D, Frimodt-Møller N, Hughes D. Biological cost of single and multiple norfloxacin resistance mutations in *Escherichia coli* implicated in urinary tract infections. Antimicrob Agents Chemother. 2005;49: 2343–51. doi:10.1128/AAC.49.6.2343-2351.2005

69. Marcusson LL, Frimodt-Møller N, Hughes D. Interplay in the selection of fluoroquinolone resistance and bacterial fitness. PLoS Pathog. 2009;5: e1000541. doi:10.1371/journal.ppat.1000541

70. Kishii R, Takei M. Relationship between the expression of *ompF* and quinolone resistance in *Escherichia coli*. J Infect Chemother. 2009;15: 361–6. doi:10.1007/s10156-009-0716-6

71. Schrag SJ, Perrot V, Levin BR. Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. Proc Biol Sci. 1997;264: 1287–91. doi:10.1098/rspb.1997.0178

72. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? Nat Rev Microbiol. 2010;8: 260–271. doi:10.1038/nrmicro2319

73. Austin DJ, Anderson RM. Studies of antibiotic resistance within the patient, hospitals and the community using simple mathematical models. Philos Trans R Soc Lond B Biol Sci. 1999/06/12 ed. 1999;354: 721–38. doi:10.1098/rstb.1999.0425

74. Goutelle S, Maurin M, Rougier F, Barbaut X, Bourguignon L, Ducher M, et al. The Hill equation: a review of its capabilities in pharmacological modelling. Fundam Clin Pharmacol. 2008;22: 633–648. doi:10.1111/j.1472-8206.2008.00633.x

75. Mouton JW, Vinks AA. Pharmacokinetic/pharmacodynamic modelling of antibacterials in vitro and in vivo using bacterial growth and kill kinetics: the minimum inhibitory concentration versus stationary concentration. Clin Pharmacokinet. 2005;44: 201–210. doi:10.2165/00003088-200544020-00005

76. Ankomah P, Johnson PJTP, Levin BR. The pharmaco–, population and evolutionary dynamics of multi-drug therapy: experiments with *S. aureus* and *E. coli* and computer simulations. PLoS Pathog. 2013;9: e1003300. doi:10.1371/journal.ppat.1003300

77. Kahlmeter G, Brown DFJ, Goldstein FW, MacGowan AP, Mouton JW, Osterlund A, et al. European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. J Antimicrob Chemother. 2003;52: 145–8. doi:10.1093/jac/dkg312

78. Levin BR, Perrot V, Walker N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics. 2000;154: 985–997.

79. Mellefont LA, McMeekin TA, Ross T. Performance evaluation of a model describing

the effects of temperature, water activity, pH and lactic acid concentration on the growth of *Escherichia coli*. Int J Food Microbiol. 2003;82: 45–58. doi:10.1016/S0168-1605(02)00253-2

80. Baker-Austin C, Wright MS, Stepanauskas R, McArthur J V. Co-selection of antibiotic and metal resistance. Trends Microbiol. 2006;14: 176–182. doi:10.1016/j.tim.2006.02.006

81. Seiler C, Berendonk TU. Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. Front Microbiol. 2012;3: 1–10. doi:10.3389/fmicb.2012.00399

82. Wold S. Validation of QSAR's. Quant Struct Relationships. 1991;10: 191–193. doi:10.1002/qsar.19910100302

83. Wiuff C, Zappala RM, Regoes RR, Garner KN, Baquero F, Levin BR. Phenotypic tolerance: antibiotic enrichment of noninherited resistance in bacterial populations. Antimicrob Agents Chemother. 2005;49: 1483–94. doi:10.1128/AAC.49.4.1483-1494.2005

84. Luo Y, Xu L, Rysz M, Wang Y, Zhang H, Alvarez PJJ. Occurrence and transport of tetracycline, sulfonamide, quinolone, and macrolide antibiotics in the Haihe River basin, China. Environ Sci Technol. 2011;45: 1827–1833. doi:10.1021/es104009s

85. Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB, et al. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999-2000: a national reconnaissance. Environ Sci Technol. 2002;36: 1202–1211. doi:10.1021/es011055j

86. Fick J, Söderström H, Lindberg RH, Phan C, Tysklind M, Larsson DGJ. Contamination of surface, ground, and drinking water from pharmaceutical production. Environ Toxicol Chem. 2009;28: 2522–2527. doi:10.1897/09-073.1

87. Kim S, Aga DS. Potential ecological and human health impacts of antibiotics and antibiotic-resistant bacteria from wastewater treatment plants. J Toxicol Environ Heal B-Critical Rev. 2007;10: 559–573. doi:10.1080/15287390600975137

88. Vogelman B, Craig WA. Kinetics of antimicrobial activity. J Pediatr. 1986;108: 835–840.

89. Daikos GL, Jackson GG, Lolans VT, Livermore DM. Adaptive resistance to aminoglycoside antibiotics from first-exposure down-regulation. J Infect Dis. 1990;162: 414–420. doi:10.1093/infdis/162.2.414

90. Barclay ML, Begg EJ, Chambers ST. Adaptive resistance following single doses of gentamicin in a dynamic in vitro model. Antimicrob Agents Chemother. 1992;36: 1951–1957. doi:10.1128/AAC.36.9.1951

91. Hocquet D, Vogne C, Garch F El, Gotoh N, Lee A, Lomovskaya O, et al. MexXY-OprM efflux pump is necessary for adaptive resistance of *Pseudomonas aeruginosa* to aminoglycosides. Antimicrob Agents Chemother. 2003;47: 1371–1375.

doi:10.1128/AAC.47.4.1371

92. Mohamed AF, Nielsen EI, Cars O, Friberg LE. Pharmacokinetic-pharmacodynamic model for gentamicin and its adaptive resistance with predictions of dosing schedules in newborn infants. Antimicrob Agents Chemother. 2012;56: 179–188. doi:10.1128/AAC.00694-11

93. Gould IM, Milne K, Harvey G, Jason C. Ionic binding, adaptive resistance and post-antibiotic effect of netilmicin and ciprofloxacin. J Antimicrob Chemother. 1991;27: 741–748.

94. Mah T-FC, O'Toole GA. Mechanisms of biofilm resistance to antimicrobial agents. Trends Microbiol. 2001;9: 34–39. doi:10.1016/S0966-842X(00)01913-2

95. Stewart PS, Franklin MJ. Physiological heterogeneity in biofilms. Nat Rev Microbiol. 2008;6: 199–210. doi:10.1038/nrmicro1838

96. Treyaprasert W, Schmidt S, Rand KH, Suvanakoot U, Derendorf H. Pharmacokinetic/pharmacodynamic modeling of in vitro activity of azithromycin against four different bacterial strains. Int J Antimicrob Agents. 2007;29: 263–70. doi:10.1016/j.ijantimicag.2006.08.049

97. Wu B, Sy SKB, Derendorf H. Principles of applied pharmacokinetic–pharmacodynamic modeling. In: Vinks AA, Derendorf H, Mouton JW, editors. Fundamentals of Antimicrobial Pharmacokinetics and Pharmacodynamics. New York, NY: Springer; 2014. pp. 63–79.

98. Broderick MP, Hansen CJ, Russell KL. Exploration of the effectiveness of social distancing on respiratory pathogen transmission implicates environmental contributions. J Infect Dis. 2008;198: 1420–1426. doi:10.1086/592711

99. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med. 2012/08/24 ed. 2012;4: 148ra116. doi:10.1126/scitranslmed.3004129

100. Miller LG, Diep BA. Clinical practice: colonization, fomites, and virulence: rethinking the pathogenesis of community-associated methicillin-resistant Staphylococcus aureus infection. Clin Infect Dis. 2008;46: 752–60. doi:10.1086/526773

101. Breban R, Drake JM, Stallknecht DE, Rohani P. The role of environmental transmission in recurrent avian influenza epidemics. PLoS Comput Biol. 2009;5: e1000346. doi:10.1371/journal.pcbi.1000346

102. Breban R. Role of environmental persistence in pathogen transmission: a mathematical modeling approach. J Math Biol. 2013;66: 535–46. doi:10.1007/s00285-012-0520-2

103. King M-F, Noakes CJ, Sleigh PA. Modelling environmental contamination in hospital single and four-bed rooms. Indoor Air. 2015;25: 694–707. doi:10.1111/ina.12186

104. Weber TP, Stilianakis NI. Inactivation of influenza A viruses in the environment and

modes of transmission: A critical review. J Infect. 2008;57: 361–373. doi:10.1016/j.jinf.2008.08.013

105. Mitscherlich E, Marth EH. Microbial Survival in the Environment. Bacteria and Rickettsiae Important in Animal and Human Health. Springer-Verlag; 1984.

106. Kramer A, Assadian O. Survival of microorganisms on inanimate surfaces. In: Borkow G, editor. Use of Biocidal Surfaces for Reduction of Healthcare Acquired Infections. Springer; 2014. pp. 7–26. Available: http://link.springer.com/chapter/10.1007/978-3-319-08057-4_2/fulltext.html

107. Kramer A, Schwebke I, Kampf G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infect Dis. 2006;6: 130. doi:10.1186/1471-2334-6-130

108. D'Souza DH, Sair A, Williams K, Papafragkou E, Jean J, Moore C, et al. Persistence of caliciviruses on environmental surfaces and their transfer to food. Int J Food Microbiol. 2006;108: 84–91. doi:10.1016/j.ijfoodmicro.2005.10.024

109. Weber DJ, Rutala WA, Miller MB, Huslage K, Sickbert-Bennett E. Role of hospital surfaces in the transmission of emerging health care-associated pathogens: norovirus, Clostridium difficile, and Acinetobacter species. Am J Infect Control. 2010;38: S25–33. doi:10.1016/j.ajic.2010.04.196

110. Julian TR. Fomites in infectious disease transmission: a modeling, laboratory, and field study on microbial transfer between skin and surfaces. PhD Thesis. Stanford University. 2010. Available: https://searchworks.stanford.edu/view/11085857

111. Jones RM. Critical review and uncertainty analysis of factors influencing influenza transmission. Risk Anal. 2011;31: 1226–42. doi:10.1111/j.1539-6924.2011.01598.x

112. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015. Available: http://www.r-project.org/

113. Rasband WS. ImageJ Image Processing and Analysis in Java. In: U. S. National Institutes of Health, Bethesda, Maryland, USA. 1997-2016. Available: http://imagej.nih.gov/ij/

114. Wendt C, Dietze B, Dietz E, Ruden H. Survival of Acinetobacter baumannii on dry surfaces. J Clin Microbiol. 1997;35: 1394–1397. Available: http://jcm.asm.org/cgi/content/abstract/35/6/1394

115. Lai MY, Cheng PK, Lim WW. Survival of severe acute respiratory syndrome coronavirus. Clin Infect Dis. 2005;41: e67–71. doi:10.1086/433186

116. Smith-Vaughan H, Crichton F, Beissbarth J, Morris PS, Leach AJ. Survival of pneumococcus on hands and fomites. BMC Res Notes. 2008;1: 112. doi:10.1186/1756-0500-1-112

117. Musa EK, Desai N, Casewell MW. The survival of Acinetobacter calcoaceticus

inoculated on fingertips and on formica. J Hosp Infect. 1990;15: 219–227. doi:10.1016/0195-6701(90)90029-N

118. Espinal P, Martí S, Vila J. Effect of biofilm formation on the survival of Acinetobacter baumannii on dry surfaces. J Hosp Infect. 2012;80: 56–60. doi:10.1016/j.jhin.2011.08.013

119. Wagenvoort JHT, Joosten EJAJ. An outbreak Acinetobacter baumannii that mimics MRSA in its environmental longevity. J Hosp Infect. 2002;52: 226–227. doi:10.1053/jhin.2001.1294

120. Jawad A, Seifert H, Snelling AM, Heritage J, Hawkey PM. Survival of Acinetobacter baumannii on dry surfaces: Comparison of outbreak and sporadic isolates. J Clin Microbiol. 1998;36: 1938–1941.

121. Wagenvoort JHT, Penders RJR, Davies BI, Lütticken R. Similar environmental survival patterns of Streptococcus pyogenes strains of different epidemiologic backgrounds and clinical severity. Eur J Clin Microbiol Infect Dis. 2005;24: 65–7. doi:10.1007/s10096-004-1256-8

122. Marks LR, Reddinger RM, Hakansson AP. Biofilm formation enhances fomite survival of Streptococcus pneumoniae and Streptococcus pyogenes. Infect Immun. 2014;82: 1141–1146. doi:10.1128/IAI.01310-13

123. Jones RM. Unpublished report. 2015.

124. Rabenau HF, Cinatl J, Morgenstern B, Bauer G, Preiser W, Doerr HW. Stability and inactivation of SARS coronavirus. Med Microbiol Immunol. 2005;194: 1–6. doi:10.1007/s00430-004-0219-0

125. Chan KH, Peiris JSM, Lam SY, Poon LLM, Yuen KY, Seto WH. The effects of temperature and relative humidity on the viability of the SARS coronavirus. Adv Virol. 2011;2011: Article ID 734690. doi:10.1155/2011/734690

126. Walsh RL, Camilli A. Streptococcus pneumoniae is desiccation tolerant and infectious upon rehydration. MBio. 2011;2: 1–6. doi:10.1128/mBio.00092-11

127. Otter JA, Donskey C, Yezli S, Douthwaite S, Goldenberg SD, Weber DJ. Transmission of SARS and MERS coronaviruses and influenza virus in healthcare settings: The possible role of dry surface contamination. J Hosp Infect. 2015;92. doi:10.1016/j.jhin.2015.08.027

128. Jones RM, Adida E. Influenza infection risk and predominate exposure route: uncertainty analysis. Risk Anal. 2011;31: 1622–31. doi:10.1111/j.1539-6924.2011.01600.x

129. Vynnycky E, White R. An Introduction to Infectious Disease Modelling. Oxford, England: Oxford University Press; 2010. Available: http://anintroductiontoinfectiousdiseasemodelling.com/

130. Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD, et al. Accelerated

telomere shortening in response to life stress. Proc Natl Acad Sci U S A. 2004;101: 17312–5. doi:10.1073/pnas.0407162101

131. Zota AR, Shenassa ED, Morello-Frosch R. Allostatic load amplifies the effect of blood lead levels on elevated blood pressure among middle-aged U.S. adults: a cross-sectional study. Environ Health. 2013;12: 64. doi:10.1186/1476-069X-12-64

132. Alexeeff G, Mataka AY. CalEnviroScreen: A pathway to address environmental justice issues in California. EM Mag. 2014; 26–31.

133. Sadd JL, Pastor M, Morello-Frosch R, Scoggins J, Jesdale B. Playing it safe: assessing cumulative impact and social vulnerability through an environmental justice screening method in the South Coast Air Basin, California. Int J Environ Res Public Health. 2011;8: 1441–1459. doi:10.3390/ijerph8051441

134. Meehan August L, Faust JB, Cushing L, Zeise L, Alexeeff G V. Methodological considerations in screening for cumulative environmental health impacts: lessons learned from a pilot study in California. Int J Environ Res Public Health. 2012;9: 3069–3084. doi:10.3390/ijerph9093069

135. Huang G, London J. Cumulative environmental vulnerability and environmental justice in California's San Joaquin Valley. Int J Environ Res Public Health. 2012;9: 1593–1608. doi:10.3390/ijerph9051593

136. Wilson S, Burwell-Naney K, Jiang C, Zhang H, Samantapudi A, Murray R, et al. Assessment of sociodemographic and geographic disparities in cancer risk from air toxics in South Carolina. Environ Res. 2015;140: 562–8. doi:10.1016/j.envres.2015.05.016

137. Los Angeles Collaborative for Environmental Health and Justice. Hidden Hazards. A Call to Action for Healthy, Livable Communities. Los Angeles, CA; 2010. Available: https://www.libertyhill.org/news/reports/hidden-hazards-call-action-healthy-livable-communities

138. OEHHA. Responses to public comments on the method to identify disadvantaged communities under Senate Bill 535. 2014. Available: http://oehha.ca.gov/ej/ces2.html

139. Morello-Frosch R, Pastor Jr. M, Sadd J, Porras C, Prichard M. Citizens, science, and data judo. Leveraging secondary data analysis to build a community-academic collaborative for environmental justice in southern California. In: Israel BA, Eng E, Schulz AJ, Parker EA, editors. Methods in Community Based Participatory Research for Health. First Edit. San Francisco, CA: John Wiley & Sons; 2005. pp. 371–392. Available: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0787980064.html

140. Bierman P, Lewis M, Ostendorf B, Tanner J. A review of methods for analysing spatial and temporal patterns in coastal water quality. Ecol Indic. 2011;11: 103–114. doi:10.1016/j.ecolind.2009.11.001

141. James FC, McCulloch CE. Multivariate analysis in ecology and systematics: panacea or Pandora's box? Annu Rev Ecol Syst. 1990; 129–166.

142. Bivand RS, Pebesma EJ, Gomez-Rubio V. Applied Spatial Data Analysis with R. Use R! New York, NY: Springer; 2008.

143. Brunsdon C, Comber L. An Introduction to R for Spatial Analysis & Mapping. London, UK: SAGE; 2015.

144. Anselin L. The Moran Scatterplot as an ESDA tool to assess local instability in spatial association. GISDATA Specialist Meeting on GIS and Spatial Analysis. Amsterdam, The Netherlands: West Virginia University Regional Research Institute; 1993. pp. 111–125.

145. Krivoruchko K. Spatial Statistical Data Analysis for GIS Users. Redlands, CA: ESRI Press; 2011.

146. Burnham KP, Anderson DR. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd ed. Berlin, Germany: Springer; 2002.

147. Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika. 1991;78: 691–692. doi:10.1093/biomet/78.3.691

148. McCune B, Grace JB. Analysis of ecological communities. Gleneden Beach, OR: MJM Software Design; 2002. Available: www.pcord.com

149. John JA, Draper NR. An alternative family of transformations. J R Stat Soc Ser C Appl Stat. 1980;29: 190–197. doi:10.2307/2986305

150. Schielzeth H. Simple means to improve the interpretability of regression coefficients. Methods Ecol Evol. 2010;1: 103–113. doi:10.1111/j.2041-210X.2010.00012.x

151. Roberts EM, English PB, Grether JK, Windham GC, Somberg L, Wolff C. Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. Environ Health Perspect. 2007;115: 1482–1489. doi:10.1289/ehp.10168

152. Alavanja MCR, Hoppin JA, Kamel F. Health effects of chronic pesticide exposure: cancer and neurotoxicity. Annu Rev Public Health. 2004;25: 155–197. doi:10.1146/annurev.publhealth.25.101802.123020

153. Marks AR, Harley K, Bradman A, Kogut K, Barr DB, Johnson C, et al. Organophosphate pesticide exposure and attention in young Mexican-American children: The CHAMACOS study. Environ Health Perspect. 2010;118: 1768–1774. doi:10.1289/ehp.1002056

154. Morgenstern H. Uses of ecological analyses in epidemiologic research. Am J Public Health. 1982;72: 1336–1344. doi:10.2105/AJPH.72.12.1336

155. Grubesic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. Int J Health Geogr. 2006;5: 58. doi:10.1186/1476-072X-5-58

156. Cushing L, Morello-Frosch R, Wander M, Pastor M. The haves, the have-nots, and the health of everyone: the relationship between social inequality and environmental quality.

Annu Rev Public Health. 2015;36: 193–209. doi:10.1146/annurev-publhealth-031914-122646

157. Shonkoff SB, Jordan P, Brandt A, Ferrar K, Maddalena R, Greenfield BK, et al. Public health risks associated with current oil and gas development in the Los Angeles Basin. An Independent Scientific Assessment of Well Stimulation in California. Sacramento, CA: California Council on Science and Technology; 2015.

158. Dykhuizen D. Experimental studies of natural selection in bacteria. Annu Rev Ecol Syst. 1990;21: 373–398. Available: http://www.jstor.org/stable/2097030

159. Hermsen R, Deris JB, Hwa T. On the rapidity of antibiotic resistance evolution facilitated by a concentration gradient. Proc Natl Acad Sci U S A. 2012;109: 10775–80. doi:10.1073/pnas.1117716109

160. Warnes SL, Little ZR, Keevil CW. Human coronavirus 229E remains infectious on common touch surface materials. MBio. 2015;6: 1–10. doi:10.1128/mBio.01697-15

161. Sizun J, Yu MW, Talbot PJ. Survival of human coronaviruses 229E and OC43 in suspension and after drying onsurfaces: a possible source ofhospital-acquired infections. J Hosp Infect. 2000;46: 55–60. doi:10.1053/jhin.2000.0795

162. Guionie O, Courtillon C, Allee C, Maurel S, Queguiner M, Eterradossi N. An experimental study of the survival of turkey coronavirus at room temperature and +4°C. Avian Pathol. 2013;42: 248–52. doi:10.1080/03079457.2013.779364

163. Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD. Effects of air temperature and relative humidity on coronavirus survival on surfaces. Appl Environ Microbiol. 2010;76: 2712–2717. doi:10.1128/AEM.02291-09

164. U.S. Census Bureau. ZIP Code[TM] Tabulation Areas (ZCTAs[TM]) [Internet]. 2015 [cited 29 Sep 2015]. Available: https://www.census.gov/geo/reference/zctas.html

165. Missouri Census Data Center. MABLE/Geocorr12: Geographic Correspondence Engine [Internet]. 2012 [cited 17 Jun 2015]. Available: http://mcdc.missouri.edu/websas/geocorr12.html

Appendix 1.

Appendix to Chapter 1. Modeling the emergence of antibiotic resistance in the environment: An analytical solution for the minimum selection concentration

Selection coefficient

Relative fitness between strains is a key determining factor governing bacterial competition [158]. Relative fitness is often conceptualized as a fitness cost [49,68,159], indicating the intrinsic fitness difference between bacteria strains (e.g., sensitive vs. resistant), as a result of a specific antibiotic resistance mechanism. This fitness cost is operationally defined in the laboratory as the selection coefficient. In the literature, selection coefficient is typically abbreviated as s [44,57,60,69,78,159]. However, to differentiate the terminology from susceptible strains (s), we use σ and sc to refer to selection coefficient as a rate [h^{-1}] or dimensionless term respectively. The selection coefficient is a measurement of the impact of a particular heritable trait on intrinsic growth rate, and consequent rate of selection for or against that trait in competition experiments. In evaluating resistance mechanisms, selection can favor sensitive strains (sensitive more fit than resistant), resistant strains (resistant more fit than sensitive), or exhibit no difference between strains. We focus on the common case where selection favors sensitive strains.

Definitions and units must be considered when analyzing relative fitness [60], such as how we define and use the selection coefficient. In particular, the intrinsic net growth rate in the absence of antibiotic (N_{int}) of sensitive (s) and resistant (r) bacteria may be represented as:

$$N_{int,s} = R_{int} - D_{int} \tag{A1}$$

$$N_{int,r} = N_{int,s}(1 - sc) = (R_{int} - D_{int})(1 - sc) \tag{A2}$$

where R_{int} is intrinsic growth rate, D_{int} is loss due to mortality or, in continuous cultures, dilution. Here, sc is a unitless term describing the growth rate difference due to the resistance mechanism.

However, selection experiments calculate relative fitness based on differences in the ratio of measured biomass over time (B_t) between strains[57,158]:

$$\sigma = [\ln(B_{r,t}/B_{s,t}) - \ln(B_{r,t=0}/B_{s,t=0})]/t \tag{A3}$$

Here, the subscript $t = 0$ indicates biomass measured at the beginning of the experiment. Although σ is also referred to as the selection coefficient, the following calculations demonstrate that σ is in units of t^{-1} (e.g., h^{-1}):

$$B_{r,t} = B_{r,t=0} \times (e^{N_{int,r}t}); B_{s,t} = B_{s,t=0} \times (e^{N_{int,s}t}) \tag{A4}$$

$$\sigma = \frac{\ln\left(\frac{B_{r,t=0} \times e^{N_{int,r}t}}{(B_{s,t=0} \times e^{N_{int,s}t}}\right) - \ln\left(\frac{B_{r,t=0}}{B_{s,t=0}}\right)}{t} = \frac{\ln\left(\frac{e^{N_{int,r}t}}{e^{N_{int,s}t}}\right)}{t} = \frac{\ln\left(e^{N_{int,r}t}\right) - \ln\left(e^{N_{int,s}t}\right)}{t} = N_{int,r} - N_{int,s} \tag{A5}$$

We see that σ is therefore equal to the difference in net growth rate between the two strains and has units of [h^{-1}] as mentioned above. This is the experimental selection coefficient (σ) determined by Gullberg et al. and elsewhere [57,61,69,78].

Below, our MSC/MIC ratio calculations employ a dimensionless selection coefficient, sc. This is obtained by dividing the inverse of the selection coefficient ($-\sigma$) by the net growth rate of the sensitive strain (Eq. 9 in the text): $sc = -\frac{\sigma}{N_{int,s}} = \frac{N_{int,s} - N_{int,r}}{N_{int,s}} = 1 - \frac{N_{int,r}}{N_{int,s}}$.

Converting MIC to EC$_{50}$

To convert the MIC value of a strain to its corresponding EC$_{50}$ value, we substitute a = MIC into the equation for death due to antibiotic (Eq. 3) to obtain $D_{ab}(MIC) = k_{max}\frac{MIC^{\kappa}}{MIC^{\kappa}+(EC_{50})^{\kappa}}$. We will demonstrate this for the sensitive strain. We first recognize that:

$$D_{ab,s}(MIC) = k_{max,s}\frac{(MIC_s)^{\kappa}}{(MIC_s)^{\kappa}+(EC_{50,s})^{\kappa}} = (R_{int} - D_{int} - N_{min})\frac{(MIC_s)^{\kappa}}{(MIC_s)^{\kappa}+(EC_{50,s})^{\kappa}} = R_{int} - D_{int} \quad (A6)$$

based on $k_{max} = R_{int} - D_{int} - N_{min}$ (Eq. 4), and the observation of zero net growth rate at the MIC: i.e., $D_{ab,s}(MIC_s) = N_{int} = R_{int} - D_{int}$. Algebraically solving for the EC$_{50,s}$ term, we find that:

$$EC_{50,s} = MIC_s \left(\frac{-N_{min}}{R_{int}-D_{int}}\right)^{\frac{1}{\kappa}} \quad (A7)$$

Thus, for a given set of conditions (κ, R_{int}, D_{int}, N_{min},), EC$_{50,s}$ is simply a constant multiple of the sensitive strain MIC. For the resistant strain, the EC$_{50,r}$ is obtained by the same solution, substituting MIC$_r$ and replacing R_{int} by ($R_{int} + \sigma$) per Eq. 2, such that:

$$EC_{50,r} = MIC_r \left(\frac{-N_{min}}{R_{int} + \sigma - D_{int}}\right)^{\frac{1}{\kappa}} \quad (A8)$$

These relationships can be employed to convert the Hill equation formulation of $D_{ab}(a) = k_{max}\frac{a^{\kappa}}{a^{\kappa}+(EC_{50})^{\kappa}}$ (Eq. 3) to an MIC based formulation, by substituting the MIC equation for EC$_{50}$. Recalling the net growth rates without antibiotic (i.e., $N_{int,s} = R_{int} - D_{int}$):

$$D_{ab,s}(a) = (N_{int,s} - N_{min})\frac{a^{\kappa}}{a^{\kappa}+\frac{(-N_{min})}{N_{int,s}}(MIC_s)^{\kappa}} \text{ and } D_{ab,r}(a) = (N_{int,r} - N_{min})\frac{a^{\kappa}}{a^{\kappa}+\frac{(-N_{min})}{N_{int,r}}(MIC_r)^{\kappa}} .$$

These are Eqs. 5 and 6, respectively in the text.

Derivation of MSC/MIC$_s$ ratio

To obtain MSC as a function of MIC$_s$, we begin by noting that at the MSC, the difference in net growth rate is equal to zero as defined in Eq. 8:

$$\Delta N(a = MSC) = N_{int,r} - N_{int,s} + D_{ab,s}(a = MSC) - D_{ab,r}(a = MSC) = 0$$

Thus:

$$\Delta N(MSC) = N_{int,r} - N_{int,s} + (N_{int,s} - N_{min})\frac{(MSC)^{\kappa}}{(MSC)^{\kappa}+\frac{(-N_{min})}{N_{int,s}}(MIC_s)^{\kappa}} - (N_{int,r} - N_{min})\frac{(MSC)^{\kappa}}{(MSC)^{\kappa}+\frac{(-N_{min})}{N_{int,r}}(MIC_r)^{\kappa}} = 0 \quad (A9)$$

To obtain the ratio MSC/MIC$_s$, we divide numerator and denominator by MIC$_s^{\kappa}$:

$$\Delta N(MSC) = N_{int,r} - N_{int,s} + (N_{int,s} - N_{min})\frac{(MSC/MIC_s)^{\kappa}}{(MSC/MIC_s)^{\kappa}+\frac{(-N_{min})}{N_{int,s}}} - (N_{int,r} - N_{min})\frac{(MSC/MIC_s)^{\kappa}}{(MSC/MIC_s)^{\kappa}+\frac{(-N_{min})}{N_{int,s}}(MIC_r/MIC_s)^{\kappa}} \quad (A10)$$

$$= 0$$

From Eq. A10, the algebraic solution was obtained for $(MSC/MIC_s)^\kappa$ employing the Equations and Systems Solver ("solve" function) in MATLAB (Symbolic Math Toolbox, R2013a, MathWorks, Natick, MA, USA). The general solution satisfying this equation is:

$$(MSC/MIC_s)^\kappa = \frac{N_{min}(MIC_r/MIC_s)^\kappa(N_{int,r}-N_{int,s})}{(MIC_r/MIC_s)^\kappa N_{int,s}(N_{int,r}-N_{min})+N_{int,r}(N_{min}-N_{int,s})} \tag{A11}$$

$$(MSC/MIC_s)^\kappa = \frac{(-N_{int,r}+N_{int,s})}{N_{int,s}\left(1+\frac{N_{int,r}}{-N_{min}}\right)+\frac{N_{int,r}(N_{min}-N_{int,s})}{(MIC_r/MIC_s)^\kappa(-N_{min})}} = \frac{(N_{int,s}-N_{int,r})/N_{int,s}}{\left(1+\frac{N_{int,r}}{-N_{min}}\right)-\frac{N_{int,r}(N_{int,s}-N_{min})}{N_{int,s}(MIC_r/MIC_s)^\kappa(-N_{min})}} \tag{A12}$$

In order to make the selection coefficient explicit, Eq. A11 can be transformed employing the dimensionless selection coefficient (sc) by introducing Eq. 9; i.e., $N_{int,s} - N_{int,r} = (sc\ N_{int,s})$, and $N_{int,r} = N_{int,s}(1-sc)$, obtaining:

$$(MSC/MIC_s)^\kappa = \frac{-N_{min}N_{int,s}(MIC_r/MIC_s)^\kappa sc}{(MIC_r/MIC_s)^\kappa N_{int,s}(N_{int,r}-N_{min})+N_{int,s}(1-sc)(N_{min}-N_{int,s})} \tag{A13}$$

Dividing the numerator and denominator by $-N_{min}N_{int,s}(MIC_r/MIC_s)^\kappa$ obtains:

$$(MSC/MIC_s)^\kappa = \frac{sc}{\frac{(N_{int,r}-N_{min})}{-N_{min}}+\frac{(1-sc)(N_{min}-N_{int,s})}{-N_{min}(MIC_r/MIC_s)^\kappa}} \tag{A14}$$

and elevating both sides of the equation to power κ^{-1} and rearranging:

$$MSC/MIC_s = \left(\frac{sc}{1-\dfrac{N_{int,r}}{N_{min}}-\dfrac{(1-sc)\left(1-\dfrac{N_{int,s}}{N_{min}}\right)}{\left(\dfrac{MIC_r}{MIC_s}\right)^\kappa}}\right)^{\frac{1}{\kappa}} \tag{A15}$$

Finally, this is depicted with N_{min} as a negative term since the minimal growth rate is often negative, giving us Eqn. 10 in the text:

$$MSC/MIC_s = \left(\frac{sc}{1+\dfrac{N_{int,r}}{-N_{min}}-\dfrac{(1-sc)\left(1+\dfrac{N_{int,s}}{-N_{min}}\right)}{\left(\dfrac{MIC_r}{MIC_s}\right)^\kappa}}\right)^{\frac{1}{\kappa}}$$

To represent this in terms of the experimentally derived selection parameters (σ) in Gullberg et al. [57,61], we simply note that $sc = -\sigma / N_{int,s}$ (Eq. 9):

$$\text{MSC/MIC}_s = \left(\cfrac{-\dfrac{\sigma}{N_{int,s}}}{1 + \dfrac{N_{int,r}}{-N_{min}} - \cfrac{\left(1 + \dfrac{\sigma}{N_{int,s}}\right)\left(1 + \dfrac{N_{int,s}}{-N_{min}}\right)}{\left(\dfrac{MIC_r}{MIC_s}\right)^{\kappa}}} \right)^{\frac{1}{\kappa}}$$

(A16)

Effect of assuming same κ and N_{min} for sensitive versus resistant strains

To arrive at an analytical solution for MSC/MIC$_s$ (Eq. 10), it was necessary to assume identical N$_{min}$ and κ for sensitive versus resistant strains. A simple Monte Carlo Simulation-based sensitivity analysis was employed to evaluate the importance of sensitive versus resistant strain N$_{min}$ and κ, and consequently which growth parameter in which strain is most important, for predicting MSC. In the simulation, the MSC was directly calculated by solving for $\Delta N = 0$ in Eq. 8, while relaxing the assumptions of identical N$_{min}$ and κ in Eqs. 5 and 6. Separate parameters were established for sensitive (N$_{min,s}$, κ_s) versus resistant strains (N$_{min,r}$, κ_r), resulting in the following formulations of antibiotic dependent growth reduction to be substituted into Eq. 8:

$$D_{ab,s}(a) = (N_{int,s} - N_{min,s}) \frac{(a)^{\kappa_s}}{(a)^{\kappa_s} + \frac{(-N_{min,s})}{N_{int,s}}(MIC_s)^{\kappa_s}} \tag{A17}$$

$$D_{ab,r}(a) = (N_{int,r} - N_{min,r}) \frac{(a)^{\kappa_r}}{(a)^{\kappa_r} + \frac{(-N_{min,r})}{N_{int,r}}(MIC_r)^{\kappa_r}} \tag{A18}$$

Two scenarios were simulated, each including 20,000 parameter sets. In both scenarios, MIC$_s$ was set at 20. In order to examine the influence of varying growth rate parameters in the presence of either small or large increases in resistance, MIC$_r$ was set at 30 and 200 in the first and second scenarios, respectively. N$_{min,s}$, N$_{min,r}$, κ_s, and κ_r were separately selected from uniform distributions with fixed ranges, listed in Table A1. Intrinsic growth rates were fixed at N$_{int,s} = 2$, and N$_{int,r} = 1.8$. Sensitivity was estimated by comparing spearman rank correlation coefficients (ρ) among the four parameters.

In both scenarios, the predicted MSC was most sensitive to κ_s, and was not sensitive to either N$_{min}$ value (Table A1). In Scenario 1 (MIC$_s$ close to MIC$_r$), the MSC was moderately sensitive to κ_r, but in Scenario 2, where MIC$_r$ was 10 times MIC$_s$, the MSC was only sensitive to κ_s. These results indicate that κ_s is the most important parameter to estimate empirically in order to predict MSC, and that κ_r only contributes to understanding MSC when MIC$_r$ is quite close to MIC$_s$. As a result, the assumptions that $\kappa = \kappa_s = \kappa_r$ and N$_{min}$ = N$_{min,s}$ = N$_{min,r}$ will not impede prediction of MSC, provided that effort is made to determine κ_s, empirically.

Table A1. Monte Carlo Simulation evaluate sensitivity of MSC to differing values of κ and N_{min} for susceptible (s) versus resistant (r) strains.

Parameter	Range (min, max)	Spearman ρ	
		Scenario 1 $MIC_s = 20, MIC_r = 30$	Scenario 2 $MIC_s = 20, MIC_r = 200$
κ_s	0.5, 10	+0.83	+0.97
κ_r	0.5, 10	−0.39	−0.09
$N_{min,s}$	−10, −1	−0.09	−0.11
$N_{min,r}$	−10, −1	+0.03	+0.003

Table A2. Complete results from all model fitting. Fitted: which parameters were varied to allow fitting to observed data. Averaged: whether data were averaged by antibiotic and strain. n: sample size. Initial κ: starting κ value in nonlinear estimation. CV κ range: range of κ results in leave one out cross validation. Q^2 = cross validated R^2 = 1 − (PRESS/TSS).

Scenario: Results:

Compound	Taxa	Fitted	Averaged	Strains	n	Initial κ	κ	CV κ range	N_{min}	Q^2	R^2	PRESS/ SSY	PRESS/ SSE	Source
As	*E. coli*	κ	No	2	20	1	0.7	0.7	−2	0.81	0.84	0.30	1.14	[61]
As	*E. coli*	κ	Yes	2	5	1	0.7	0.7 – 0.8	−2	0.80	0.92	0.29	2.44	[61]
As	*E. coli*	κ, N_{min}	No	2	20	1	1.2	1.1 – 1.2	−0.2	0.88	0.91	0.12	1.30	[61]
As	*E. coli*	κ, N_{min}	Yes	2	5	1	1.2	1.1 – 1.2	−0.2	0.99	1.00	0.01	24.1	[61]
CIP	*E. coli*	κ	No	5	144	2	2.0	2.0	−2	0.77	0.78	0.31	1.03	[57]
CIP	*E. coli*	κ	Yes	5	24	2	2.0	1.9 – 2.0	−2	0.78	0.81	0.29	1.20	[57]
CIP [b]	*E. coli*	κ	Yes	4	18 [b]	2	2.1	2.1	−2	0.97	0.97	0.04	1.23	[57]
CIP	*E. coli*	κ, N_{min}	No	5	144	2	1.6	1.6 – 1.7	−5.2e+8 [a]	0.78	0.79	0.32	1.03	[57]
CIP	*E. coli*	κ, N_{min}	Yes	5	24	2	1.6	1.6 – 1.9	−3.8e+9 [a]	0.77	0.83	0.32	1.30	[57]
CIP [b]	*E. coli*	κ, N_{min}	Yes	4	18 [b]	2	2.4	2.3 – 2.4	−0.8	0.98	0.98	0.02	1.40	[57]
Cu	*E. coli*	κ	No	2	8	2	1.9	1.8 – 2.1	−2	0.43	0.73	0.80	2.13	[61]
Cu	*E. coli*	κ	Yes	2	4	2	1.9	1.8 – 3.1	−2	c	0.88	c	c	[61]
Cu	*E. coli*	κ, N_{min}	No	2	8	2	5.7	3.5 – 6.3	−0.0003	0.40	0.82	0.68	3.36	[61]
Cu	*E. coli*	κ, N_{min}	Yes	2	4	2	5.7	2.9 – 5.7	−0.0003	c	0.98	c	c	[61]
ERY	*E. coli*	κ	No	3	64	2	3.5	3.4 – 3.5	−2	0.93	0.94	0.07	1.06	[61]
ERY	*E. coli*	κ	Yes	3	11	2	3.4	3.3 – 3.5	−2	0.92	0.95	0.09	1.64	[61]
ERY	*E. coli*	κ, N_{min}	No	3	64	2	2.6	2.6	−4.0e+8 [a]	0.70	0.73	0.18	1.11	[61]
ERY	*E. coli*	κ, N_{min}	Yes	3	11	2	2.7	2.6	−3.4e+8 [a]	0.75	0.87	0.17	1.88	[61]
KAN	*E. coli*	κ	No	2	72	2	10.5	10.4 – 10.6	−2	−0.48	−0.47	0.43	1.01	[61]

KAN	*E. coli*	κ	Yes	2	5	2	10.5	7.1 – 10.8	−2	−19.5	−0.76	3.67	11.6	[61]
KAN	*E. coli*	κ, N_{min}	No	2	72	2	6.0	6.0 – 6.1	−1.2e+11 [a]	−7.31	−6.76	0.86	1.07	[61]
KAN	*E. coli*	κ, N_{min}	Yes	2	5	2	6.0	5.7 – 6.7	−1.5e+11 [a]	−16.2	−11.4	1.06	1.39	[61]
STR	Salmonella	κ	No	2	87	2	5.0	5.0	−2	0.66	0.67	0.25	1.02	[57]
STR	Salmonella	κ	Yes	2	5	2	5.0	4.1 – 5.2	−2	−1.05	0.70	1.47	6.87	[57]
STR	Salmonella	κ, N_{min}	No	2	87	2	3.4	3.4	−8.1e+11 [a]	−0.22	−0.16	0.30	1.05	[57]
STR	Salmonella	κ, N_{min}	Yes	2	5	2	3.4	2.9	−1.0e+12 [a]	−2.35	−0.26	0.77	2.66	[57]
TET	*E. coli*	κ	No	3	60	2	1.6	1.6	−2	0.89	0.89	0.09	1.03	[61]
TET	*E. coli*	κ	Yes	3	10	2	1.6	1.6	−2	0.94	0.95	0.05	1.19	[61]
TET	*E. coli*	κ, N_{min}	No	3	60	2	2.1	2.1 – 2.2	−0.4	0.90	0.91	0.07	1.08	[61]
TET	*E. coli*	κ, N_{min}	Yes	3	10	2	2.0	1.8 – 2.2	−0.5	0.93	0.96	0.05	1.70	[61]
TET	Salmonella	κ	No	2	154	1	1.2	1.2	−2	0.93	0.93	0.04	1.02	[57]
TET	Salmonella	κ	Yes	2	5	1	1.2	1.2	−2	0.97	0.99	0.01	2.14	[57]
TET	Salmonella	κ, N_{min}	No	2	154	1	1.2	1.2	−1.8	0.93	0.93	0.04	1.04	[57]
TET	Salmonella	κ, N_{min}	Yes	2	5	1	1.2	1.0 – 1.5	−2.1	0.91	0.99	0.04	7.70	[57]
TMP	*E. coli*	κ	No	2	118	2	2.5	2.5	−2	0.87	0.88	0.07	1.03	[61]
TMP	*E. coli*	κ	Yes	2	5	2	2.5	2.5	−2	0.99	0.99	0.01	1.19	[61]
TMP	*E. coli*	κ, N_{min}	No	2	118	2	2.2	2.1 – 2.2	−10.9	0.87	0.88	0.07	1.06	[61]
TMP	*E. coli*	κ, N_{min}	Yes	2	5	2	2.2	2.0 – 2.5	−10.3	0.96	0.99	0.02	4.17	[61]

a. Model fitting insensitive to N_{min}. b. Simulation with *gyrA1(S83L)* removed. c. Insufficient n for cross validation statistics (n = 4).

Table A3. Comparison of MSC/MIC$_s$ ratios experimentally observed [57,61] versus predicted by fitted model (Eq. 10) for different resistance mechanisms and compounds. Fitted: which parameters (κ, N$_{min}$) were varied to allow fitting to observed data. Data are plotted in Figure 4.

Compound	Taxa	Resistance gene	MSC observed (ng ml^{-1})	MIC$_s$ (ng ml^{-1})	Ref.	MSC/MIC$_s$ observed	MSC/MIC$_s$ fitted κ	MSC/MIC$_s$ fitted κ, N$_{min}$
Arsenite	*E. coli*	pUUH239.2	90 [a]	12,500 [a]	[61]	0.0072	0.0064	0.0077
Trimethoprim	*E. coli*	pUUH239.2	33	190	[61]	0.174	0.180	0.178
Tetracycline	*E. coli*	pUUH239.2	45	750	[61]	0.060	0.063	0.070
Tetracycline	*E. coli*	*tetRA*	30	750	[61]	0.040	0.014	0.021
Erythromycin	*E. coli*	pUUH239.2	3000	12,000	[61]	0.250	0.266	0.229
Erythromycin	*E. coli*	*mph*	< 200	12,000	[61]	< 0.017	0.074	0.044
Kanamycin	*E. coli*	pUUH239.2	470	750	[61]	0.627	0.656	0.532
Cu(II) sulfate	*E. coli*	pUUH239.2	90	1,300	[61]	0.069	0.035	0.079
Ciprofloxacin	*E. coli*	*GyrA1(S83L)*	b	23	[57]	0.0043	0.024	0.017
Ciprofloxacin	*E. coli*	*GyrA2(D87N)*	b	23	[57]	0.10	0.088	0.080
Ciprofloxacin	*E. coli*	*ΔmarR*	b	23	[57]	0.10	0.097	0.094
Ciprofloxacin	*E. coli*	*ΔacrR*	b	23	[57]	0.10	0.091	0.087
Streptomycin	Salmonella	*rpsL105(K42R)*	b	4,000	[57]	0.25	0.383	0.290
Tetracycline	Salmonella	*cobA367::Tn10dtet*	b	1,500	[57]	0.01	0.0077	0.0077

a. Units are µM. b. Not reported in study.

Table A4. Laboratory growth parameters employed to illustrate ranges in MSC/MIC$_s$ ratios. For all simulations, MICr was set at 10*MICs.

Compound	κ	MIC$_s$ [μg ml^{-1}]	N$_{int}$ [h^{-1}]	N$_{min}$ [h^{-1}]	Reference
Ciprofloxacin #1	1.42	0.7	1.59	−15.7	Ankomah et al. [76]
Ampicillin #1	4.53	3.47	1.57	−1.16	Ankomah et al. [76]
Tetracycline #1	1.46	0.92	1.30	−8.32	Ankomah et al. [76]
Tobramycin	2.67	1.2	1.08	−16.6	Ankomah et al. [76]
Ciprofloxacin #2	1.1	0.03	0.88	−6.5	Regoes et al. [65]
Ampicillin #2	0.75	8	0.75	−4.0	Regoes et al. [65]
Rifampin	2.5	8	0.70	−4.3	Regoes et al. [65]
Streptomycin	1.9	32	0.89	−8.8	Regoes et al. [65]
Tetracycline #2	0.61	1	0.81	−8.1	Regoes et al. [65]
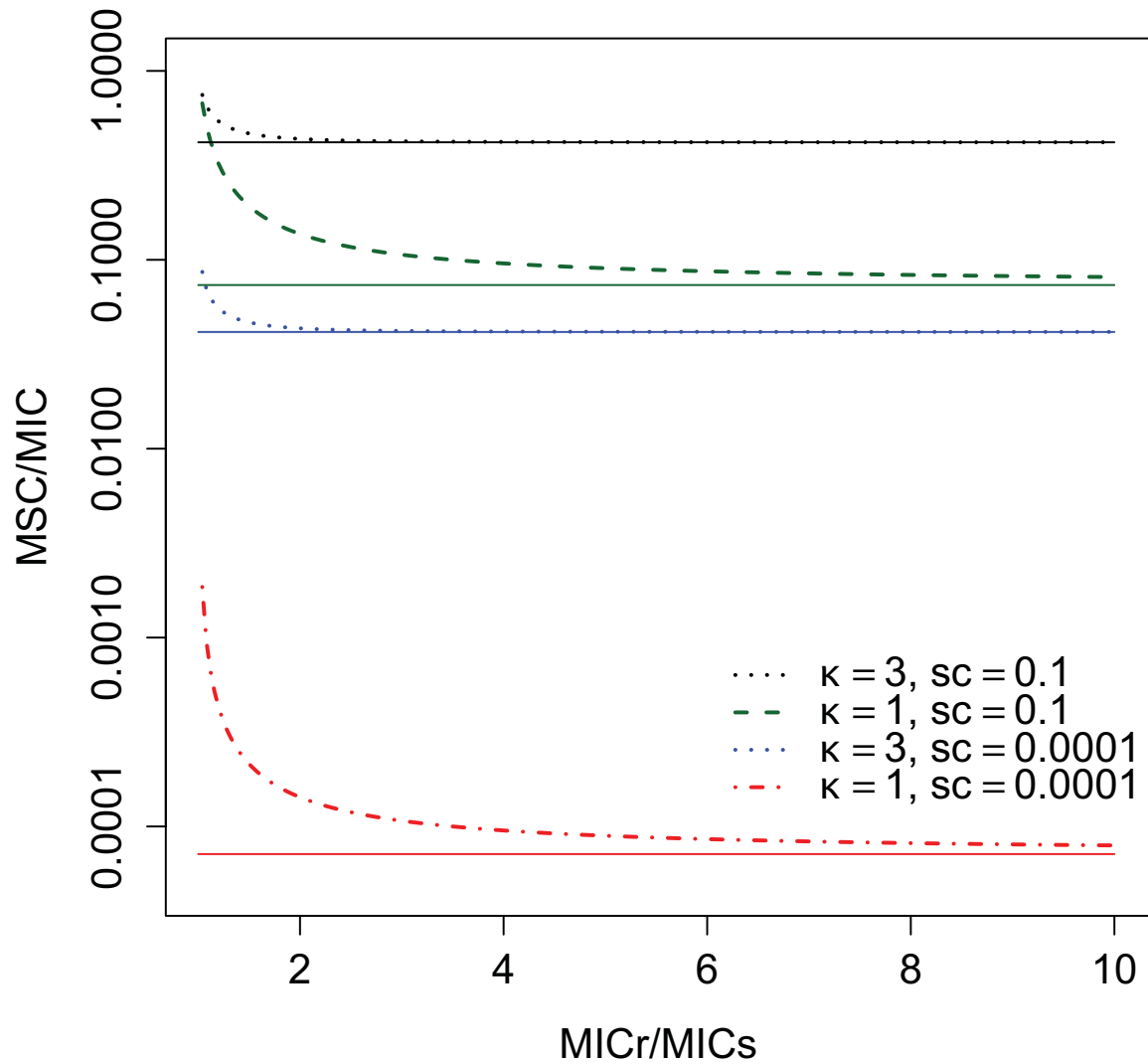
Figure A1. Calculated MSC/MIC ratio as a function of MIC_r/MIC_s for different κ and sc values. Dotted lines are from the full form of the analytical solution (Eq. 10) and solid horizontal lines are from the simplified solution, which does not include MIC_r or MIC_s (Eq. 11). MSC/MIC is most sensitive to MIC_r/MIC_s when $MIC_r < 2 \times MIC_s$ and the solutions converge for high MIC_r. Other parameter values: N_{min} = −5; $N_{int,s}$ = 2; MIC_s = 25. Note log scale y-axis.

Figure A2. Model fit to average tetracycline results, employing *Salmonella enterica* serovar Typhimurium [data from 57], varying κ only (N_{min} = −2).

Figure A3. Model fit to average kanamycin results in *E. coli* [data from 61], varying κ only (N_{min} = −2).

Appendix 2.

Appendix to Chapter 2. Transfer rate model for environmental surface contribution to hospital-associated infection transmission.
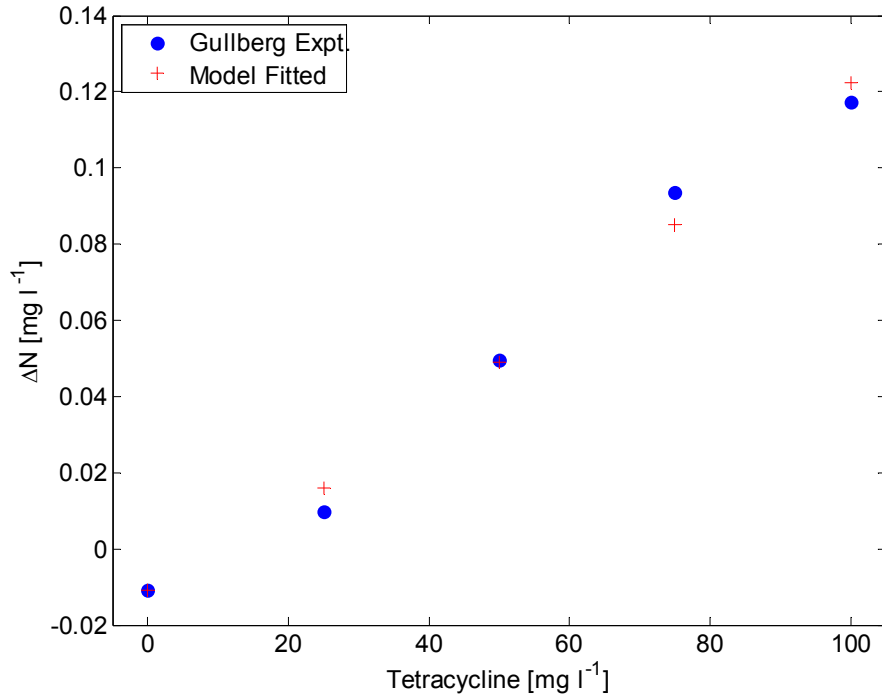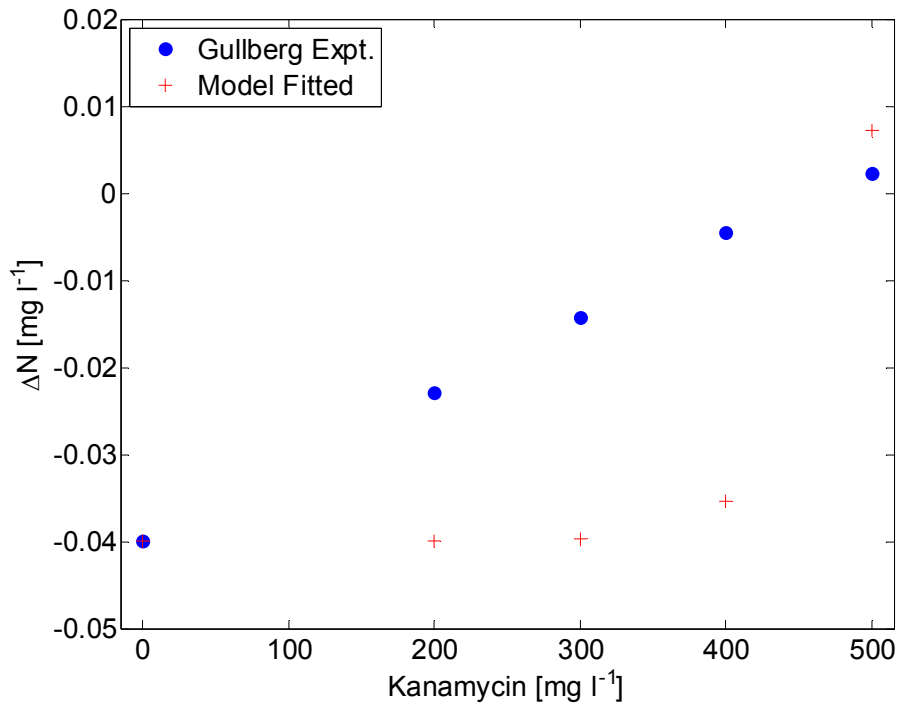
Model description and equations

Table A5. Constants for dermal contact and for elimination.

Constant	Value [units]	Definition	Data Source
Contact areas	**[cm²]**		
Af	1	Area of finger coming in contact with mucous membrane	[45]
An	4	Mucous membrane area	[45]
Apt	2000	Exposed hand and skin of patient	[45]
Ahcw	2000	Exposed hand and skin of HCW	[45]
Ap	9000	Porous surface area surrounding patient (bedding)	[18]
Anp	1000	Nonporous surface area surrounding patient (bedding)	[18]
Ac	150	Hand contact area	[45]
Ap.hcw	2000	Textiles or surfaces of HCW	[45]
Transfer efficiency	**[unitless]**		
ρ.nos	0.2	From fingertip to mucous membrane	[45]
ρ.por	0.1	From porous surface to skin	[45,103]
ρ.sur	0.4	From nonporous surface to skin	[45,103]
ρ.sk	0.35	From hand to skin	[45]
Contact (touching) rate	**[h⁻¹]**		
τ.nos	1.5	Touching nose	[45] [a]
τ.sk.por	8	Patient touching porous surfaces (textiles)	[45]
τ.sk.sur	8	Patient touching nonporous surfaces	[45]
τ.sur.pt	8	Patient touching textiles and surfaces	[45]
τ.sk.hcw	24	Health-care worker touching patient skin	[45]
τ.por.hcw	24	Health-care worker touching porous surfaces (textiles)	[45]
τ.sur.hcw	24	Health-care worker touching nonporous surfaces	[45]
Elimination rate	**[h⁻¹]**		
μ.sk	Table 2	Skin	Table 2
μ.por	Table 2	Porous surfaces	Table 2
μ.sur	Table 2	Nonporous surfaces	Table 2
μ.air	Table 2	All surfaces	Table 2
exhaust.air	6	Loss from room air due to air exchange	[19]

a. Nicas and Jones [18] report 5 h⁻¹

Table A6. Constants for respiratory transmission and other constants.

Constant	Value	Units	Definition	Data Source
visit.rate	0.111		Portion of time HCW spends in patient room	[45] [a]
CoughFluidVol	0.044	[ml]	Cough fluid volume	[18,19]
CoughRate	12	[h^{-1}]	Cough rate	[18,19]
PathConc	1,000,000	[cfu ml^{-1}]	Pathogen concentration in fluid	[18]
PathSurfConc	250	[cfu cm^{-1}]	Pathogen concentration per area on mucous membranes	[45]
CoughEmitAir	0.000001		Proportion of cough emitted that is respirable particles (remains in air)	[18,19]
CoughEmitSurf	0.10		Proportion of cough emitted that lands on nonporous surface	[19]
CoughEmitText	0.45		Proportion of cough emitted that lands on porous surfaces (textiles)	[19] [b]
CoughEmitSkin	0.45		Proportion of cough emitted that lands on patient skin	Estimate
PT1.text.PropSA	0.050		Proportion of room area on which airborne pathogen can settle that is patient-associated textile (e.g., bedding, clothing)	[19]
PT1.surf.PropSA	0.050		Proportion of room area on which airborne pathogen can settle that is patient nonporous surfaces (e.g., railing, table)	[19]
PT1.skin.PropSA	0.011		Proportion of room area on which airborne pathogen can settle that is patient nonporous surfaces (e.g., railing, table)	Estimate [c]
AirSettle	0.93	[h^{-1}]	Respirable particle settling rate from room air	[19] [d, e]
BreathingRate	1.2	[m^3 h^{-1}]	Breathing rate	[19]
RoomVolume	80	[m^3]	Room volume	[19]
PartDepositFrac	0.9		Fraction of particles in air deposited into respiratory tract	[19] [e]

a. Assuming 20 min h^{-1} for an 8 h shift per day
b. Subtracting the proportion emitted in air to achieve mass balance
c. Estimated as PT1.surf.PropSA*(Apt/Ap)
d. Assumes a representative respirable particle size of 5 μm [19]
e. Assumes a particle density of 1 g cm^{-3} [19]

Compartments
The model has 13 states (compartments) in which the infectious agent may be located. States are numbered as follows:
G. Mucous membranes of infected patient 1
1. Skin of infected patient 1
2. Porous surfaces (textiles) of infected patient 1
3. Nonporous surfaces (non-textiles) of infected patient 1
4. Infected patient 1 room air
5. Mucous membranes of health-care worker
6. Skin of health-care worker
7. Porous surfaces (textiles) of health-care worker
8. Mucous membranes of uninfected patient 2
9. Skin of uninfected patient 2
10. Porous surfaces (textiles) of uninfected patient 2
11. Nonporous surfaces of uninfected patient 2
12. Elimination (i.e., die off) – a permanent loss from system

Rate Parameters
Input rates (G, [cfu h^{-1}]) from infected patients mucous membranes to hands, air, surfaces, or textiles via coughing and skin contact (touching)

CoughEmitRate = CoughFluidVol * CoughRate *PathConc

ContactEmitSkin = An*(Af/An)*ρ.nos*τ.nos*PathSurfConc = Af*ρ.nos*τ.nos*PathSurfConc

G_1 = ContactEmitSkin + CoughEmitSkin*CoughEmitRate

G_2 = CoughEmitText*CoughEmitRate

G_3 = CoughEmitSurf*CoughEmitRate

G_4 = CoughEmitAir*CoughEmitRate

Transfer rates (λ, [h^{-1}])
Transfer from infected patient 1's skin (1), porous surfaces (2), and nonporous surfaces (3) among each other and to HCW skin (6):

$\lambda_{1,2}$ = (Ac/Apt)*ρ.por*τ.sk.por

$\lambda_{1,3}$ = (Ac/Apt)*ρ.sur*τ.sk.sur

$\lambda_{1,6}$ = (Ac/Apt)*ρ.sk*τ.sk.hcw*visit.rate

$\lambda_{2,1}$ = (Ac/Ap)*ρ.por*τ.sk.por

$\lambda_{2,6}$ = (Ac/Ap)*ρ.por*τ.por.hcw*visit.rate

$\lambda_{3,1}$ = (Ac/Anp)*ρ.sur*τ.sk.sur

$\lambda_{3,6}$ = (Ac/Anp)*ρ.sur*τ.sur.hcw*visit.rate

Transfer from air in patient 1 room to other compartments [19] including respiratory tract (5) and skin (6) and textiles (7) of HCW

HCW.text.PropSA = PT1.text.PropSA * Ap.hcw/Ap

HCW.skin.PropSA = PT1.skin.PropSA

$\lambda_{4,1}$ = PT1.skin.PropSA*AirSettle

$\lambda_{4,2}$ = PT1.text.PropSA*AirSettle

$\lambda_{4,3}$ = PT1.surf.PropSA*AirSettle

$\lambda_{4,5}$ = BreathingRate*PartDepositFrac/RoomVolume

$\lambda_{4,6}$ = HCW.skin.PropSA*AirSettle*visit.rate
$\lambda_{4,7}$ = HCW.text.PropSA*AirSettle*visit.rate

Transfer from respiratory tract (5), skin (6), and textiles (7) of HCW. From the HCW respiratory tract, there is also transfer to the skin (9) of patient 2, and the porous (10) and nonporous (11) surfaces in patient 2's room

$\lambda_{5,6}$ = (Af/An)*ρ.nos*τ.nos
$\lambda_{6,1}$ = $\lambda_{6,9}$ = (Ac/Ahcw)*ρ.sk*τ.sk.hcw*visit.rate
$\lambda_{6,2}$ = $\lambda_{6,10}$ = (Ac/Ahcw)*ρ.por*τ.por.hcw*visit.rate
$\lambda_{6,3}$ = $\lambda_{6,11}$ = (Ac/Ahcw)*ρ.sur*τ.sur.hcw*visit.rate
$\lambda_{6,5}$ = (Af/Ahcw)*ρ.nos*τ.nos
$\lambda_{6,7}$ = (Ac/Ahcw)*ρ.por*τ.por.hcw
$\lambda_{7,6}$ = (Ac/Ap.hcw)*ρ.por*τ.por.hcw

Transfer rates from patient 2 compartments. Rates from respiratory tract (mucous membranes, 8), skin (9), porous surfaces (10), and nonporous surfaces (11) largely follow the rates already calculated

$\lambda_{8,9}$ = $\lambda_{5,6}$
$\lambda_{9,6}$ = $\lambda_{1,6}$
$\lambda_{9,8}$ = (Af/Apt)*ρ.nos*τ.nos
$\lambda_{9,10}$ = $\lambda_{1,2}$
$\lambda_{9,11}$ = $\lambda_{1,3}$
$\lambda_{10,6}$ = $\lambda_{2,6}$
$\lambda_{10,9}$ = $\lambda_{2,1}$
$\lambda_{11,6}$ = $\lambda_{3,6}$
$\lambda_{11,9}$ = $\lambda_{3,1}$

Elimination rates (12) from skin (1, 6, 9), porous surfaces (2, 7, 10), nonporous surfaces (3, 11), and air (4). Elimination in air includes degradation plus exhaust.

$\lambda_{1,12}$ = $\lambda_{6,12}$ = $\lambda_{9,12}$ = μ.sk
$\lambda_{2,12}$ = $\lambda_{7,12}$ = $\lambda_{10,12}$ = μ.por
$\lambda_{3,12}$ = $\lambda_{11,12}$ = μ.sur
$\lambda_{4,12}$ = μ.air + exhaust.air

Steady state air concentration
Room.Concentration.Steady.State = EmitRate*CoughEmitAir / (exhaust.air + μ.air)

First-order transfer rate out of each compartment
$\lambda_{1,L}$ = $\lambda_{1,2}$ + $\lambda_{1,3}$ + $\lambda_{1,6}$ + $\lambda_{1,12}$
$\lambda_{2,L}$ = $\lambda_{2,1}$ + $\lambda_{2,6}$ + $\lambda_{2,12}$
$\lambda_{3,L}$ = $\lambda_{3,1}$ + $\lambda_{3,6}$ + $\lambda_{3,12}$
$\lambda_{4,L}$ = $\lambda_{4,1}$ + $\lambda_{4,2}$ + $\lambda_{4,3}$ + $\lambda_{4,5}$ + $\lambda_{4,6}$ + $\lambda_{4,7}$ + $\lambda_{4,12}$
$\lambda_{5,L}$ = $\lambda_{5,6}$
$\lambda_{6,L}$ = $\lambda_{6,1}$ + $\lambda_{6,2}$ + $\lambda_{6,3}$ + $\lambda_{6,5}$ + $\lambda_{6,7}$ + $\lambda_{6,9}$ + $\lambda_{6,10}$ + $\lambda_{6,11}$ + $\lambda_{6,12}$
$\lambda_{7,L}$ = $\lambda_{7,6}$ + $\lambda_{7,12}$
$\lambda_{8,L}$ = $\lambda_{8,9}$

$$\lambda_{9,L} = \lambda_{9,6} + \lambda_{9,8} + \lambda_{9,10} + \lambda_{9,11} + \lambda_{9,12}$$
$$\lambda_{10,L} = \lambda_{10,6} + \lambda_{10,9} + \lambda_{10,12}$$
$$\lambda_{11,L} = \lambda_{11,6} + \lambda_{11,9} + \lambda_{11,12}$$

The full mass balance matrix:

$$
\begin{bmatrix}
dN_1/dt \\
dN_2/dt \\
dN_3/dt \\
dN_4/dt \\
dN_5/dt \\
dN_6/dt \\
dN_7/dt \\
dN_8/dt \\
dN_9/dt \\
dN_{10}/dt \\
dN_{11}/dt
\end{bmatrix}
=
\begin{bmatrix}
-\lambda_{1,L} & \lambda_{2,1} & \lambda_{3,1} & \lambda_{4,1} & 0 & \lambda_{6,1} & 0 & 0 & 0 & 0 & 0 \\
\lambda_{1,2} & -\lambda_{2,L} & 0 & \lambda_{4,2} & 0 & \lambda_{6,2} & 0 & 0 & 0 & 0 & 0 \\
\lambda_{1,3} & 0 & -\lambda_{3,L} & \lambda_{4,3} & 0 & \lambda_{6,3} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\lambda_{4,L} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \lambda_{4,5} & -\lambda_{5,L} & \lambda_{6,5} & 0 & 0 & 0 & 0 & 0 \\
\lambda_{1,6} & \lambda_{2,6} & \lambda_{3,6} & \lambda_{4,6} & \lambda_{5,6} & -\lambda_{6,L} & \lambda_{7,6} & 0 & \lambda_{9,6} & \lambda_{10,6} & \lambda_{11,6} \\
0 & 0 & 0 & \lambda_{4,7} & 0 & \lambda_{6,7} & -\lambda_{7,L} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_{8,L} & \lambda_{9,8} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_{6,9} & 0 & \lambda_{8,9} & -\lambda_{9,L} & \lambda_{10,9} & \lambda_{11,9} \\
0 & 0 & 0 & 0 & 0 & \lambda_{6,10} & 0 & 0 & \lambda_{9,10} & -\lambda_{10,L} & 0 \\
0 & 0 & 0 & 0 & 0 & \lambda_{6,11} & 0 & 0 & \lambda_{9,11} & 0 & -\lambda_{11,L}
\end{bmatrix}
\times
\begin{bmatrix}
N_1 \\
N_2 \\
N_3 \\
N_4 \\
N_5 \\
N_6 \\
N_7 \\
N_8 \\
N_9 \\
N_{10} \\
N_{11}
\end{bmatrix}
+
\begin{bmatrix}
G_1 \\
G_2 \\
G_3 \\
G_4 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
$$

In shorthand:
$$\mathbf{N`(t)} = \mathbf{\Lambda} \times \mathbf{N} + \mathbf{G}$$

The steady state solution is obtained with the above matrix
$$\mathbf{N_{SS}} = \mathbf{\Lambda^{-1}} \times (-\mathbf{G})$$

To examine the role of surfaces (porous and nonporous) versus skin, we run a comparison simulation of the above equation, setting all surface transfer rates to 0
$$\lambda_{1,3} = \lambda_{1,4} = \lambda_{3,1} = \lambda_{3,5} = \lambda_{4,1} = \lambda_{4,5} = \lambda_{5,3} = \lambda_{5,4} = \lambda_{5,9} = \lambda_{5,10} = \lambda_{8,9} = \lambda_{8,10} = \lambda_{9,5} = \lambda_{9,8} = \lambda_{10,5} = \lambda_{10,8}$$
$$= \lambda_{3,12} = \lambda_{9,12} = \lambda_{4,12} = \lambda_{10,12} = \lambda_{5,13} = \lambda_{13,5} = \lambda_{4,3} = \lambda_{4,4} = \lambda_{4,13} = 0$$

Summary of literature elimination rates for selected pathogens

Acinetobacter baumannii

Acinetobacter baumannii (Table A7) are limited in only having elimination results available for nonporous surfaces [114,118–120]. There are sufficient data to generate a range of estimates, as described in the text. A study of *Acinetobacter calcoaceticus* var *anitratus* formed an adequate basis for comparing elimination on surfaces versus human skin [117]. Detailed time decay data extracted from the figures of this study indicated substantially higher first order decay coefficients on fingertips ($2.5 h^{-1}$) than on Formica surfaces ($0.11 h^{-1}$). Although these rates are one to three orders of magnitude higher than for *Acinetobacter baumannii* (Table A7), they form a basis for judging that elimination on fingertips will be approximately 23 times that on surfaces ($2.5/0.11 = 22.7$).

SARS Coronavirus

SARS coronavirus (SARS-CoV) elimination is restricted to four studies, reviewed in Otter et al. [127]. Rabenau et al. [124] and Chan et al. [125] quantitatively evaluated elimination on plastic surfaces. Though Chan et al. illustrated elimination rates are dependent on temperature and humidity, by assuming moderate humidity ($30 – 50\%$) and room temperature ($21 – 25 \,°C$), we obtain similar rates for both studies (Table A8), averaging $0.032 h^{-1}$, which we will use for elimination rates on nonporous surfaces. Other coronaviruses have also been examined on various nonporous surfaces, including human CoV, transmissible gastroenteritis virus (TGEV), and mouse hepatitis virus (MHV) [127]. The range of elimination rates observed either similar or faster than SARS-CoV on nonporous surfaces, ranging from $0.035 h^{-1}$ [160] up to $1.6 h^{-1}$ [161], also suggesting that SARS-CoV is relatively stable on dry surfaces [125,127]. Only Lai et al. [115] examine SARS-CoV stability on textiles. To judge elimination in textiles, we see from Lai et al. [115] that the time to inactivation on a plastic gown (nonporous) was longer than for a cloth lab coat (porous, textile). With a starting inoculum of 10^6, 10^5, or 10^4 $TCID_{50} ml^{-1}$, Lai et al. observed that the ratio of times to inactivation was 2, 24, and 12 between the plastic and cloth. Taking the median value (12), we extrapolate from the nonporous surface (i.e., 12 x $0.032 h^{-1}$) arriving at a rate of $0.38 h^{-1}$ for textiles. As there are no studies of SARS-CoV elimination on skin or in air, the rate for nonporous surfaces will be employed.

Streptococcus pyogenes

Ranges for elimination rates of *Streptococcus pyogenes* vary widely, depending on the data source and study method. Historic studies performed from 1897 to 1947 report similar survival time ranges on dust (3 d to 195 d) and on textiles (5 d to 120 d) [105,107], suggesting that survival on porous and nonporous surfaces would be similar. For more quantitative rates of elimination on surfaces, we use the data from Wagenvoort et al. [121] and Marks et al. [122]. Wagenvoort et al. [121] determined elimination in eight *S. pyogenes* clinical strains, including four strains in Group A, which cause more serious and invasive infections. Model fitting to the decay rates indicated a mean elimination rate of $0.023 h^{-1}$ for the eight strains ($0.020 h^{-1}$ for the four Group A strains). Marks et al. [122] compared survival of planktonic growth versus biofilms on in vitro prefixed epithelia, the latter intended to represent growth in the nasopharynx. Although the rates were two orders of magnitude lower on biofilms (Table A9),

85

the data for planktonic growth are more relevant for our purposes of depicting survival on surfaces. These results from four experiments [data extracted from Fig 1b, d in 122] indicate decay rates that are best described by a power function (i.e., [bacteria, cfu ml^{-1}] = α*[time, min]$^{\beta}$) over 120 days, with the exponential term varying between β = −2.1 and β = −3.3. In order to approximate first-order decay, an exponential decay function was fitted to the data from the first four sample points (0.5 h, 1 d, 3 d, and 7 d), resulting in an average rate of 0.078 h^{-1} (Table A9). Although the decay would be slower with longer durations, the early part of the curve represents the rate for the majority of colonies. For a representative elimination rate for textiles, surfaces, and air, we take the average of the two studies [121,122] (0.078 and 0.023), resulting in 0.051 h^{-1}.

Elimination rates for *Streptococcus* spp. are more rapid on skin, with Marks et al. reporting complete loss of *Streptococcus pneumoniae* in 1 h [122]. Although we found no quantitative data comparing *Streptococcus pyogenes* in skin versus surfaces, Smith-Vaughan et al. [116] estimate survival rate via relative cfu count swabbed from hands versus glass. Our data extraction from this study indicated that the rate on skin (2.28 h^{-1}) was 4 times that on glass surfaces (0.57 h^{-1}). Given data limitations, we estimate an elimination rate on skin that is 4 times that on surfaces (0.05 h^{-1}), resulting in 0.2 h^{-1}.

Bordetella pertussis

Mitscherlich and Marth [105] report elimination rates as survival durations in air, human skin, textiles (linen and teddy bear), and nonporous surfaces (glass and plastics) from two previously published studies. The original studies were published in German and Russian and therefore could not be consulted. Survival durations varied among media, studies, and samples, ranging from < 1 to 6 h on skin to 3 − 5 d on plastic (Table A10). Employing average values as described in Table A10, we obtain loss rates of 0.051 h^{-1} in air and 0.34 h^{-1} in skin. For textiles (porous surfaces) and nonporous surfaces, we take the average of the two values available, resulting in 0.021 h^{-1} for textiles and 0.047 h^{-1} for nonporous surfaces.

Table A7. Elimination rate data extracted for *Acinetobacter baumannii*.

Strain traits	Study method	Medium	Duration (h)	Initial Initial count (cfu)	Final Final count (cfu)	Data Extraction Method	Loss rate Average (h⁻¹)	Ref
Biofilm forming (Strain	"Survival assays were determined by viable counts of the cells inoculated on to glass cover slips and stored under controlled conditions of temperature and relative humidity"	Glass	864	~1,200	0	Zero order: 1/(Survival Time)	0.0012	[118]
Biofilm forming (Strain	See above	Glass	864	~1,200	0	Zero order: 1/(Survival Time)	0.0012	[118]
Not biofilm forming (Strain AB143)	See above	Glass	431	1,100	0	Zero order: 1/(Survival Time)	0.0023	[118]
Not biofilm forming (Strain AB001)	See above	Glass	431	1,100	0	Zero order: 1/(Survival Time)	0.0023	[118]
Long surviving strains, often dry sources (Type 1)	"Ceramic, polyvinyl chloride (PVC), rubber, and stainless steel samples (5 by 5 cm) were disinfected with 70% ethanol and contaminated with 0.1 ml of the bacterial solution. The total number of contaminated samples was 1,800 (45 samples per strain and material). All samples were stored in a dark, dust-protected climate chamber at 22 6 2°C with 50% 6 5% relative humidity."	Ceramic, PVC, Rubber, Steel	2680	17,989,000	1,413	ImageJ followed by fitting to exponential decay function	0.00086	[114]
Long surviving strains, often dry sources (Type 2)	See above	Ceramic, PVC, Rubber, Steel	339	306,196	422	ImageJ followed by fitting to exponential decay function to first 6 data points. Later data had slower decay	0.012	[114]

Short surviving strains, often wet sources (Type 3)	See above	Ceramic, PVC, Rubber, Steel	332	16,444	136	See above	0.014	[114]
Short surviving strains, often wet sources (Type 4)	See above	Ceramic, PVC, Rubber, Steel	386	265	6	See above	0.010	[114]
Hospital outbreak strain capable of surviving 329 days	Dry in darkened bottle plugged with cotton wool for 14 days. Allowed to resuspend in liquid medium, vortexed, and plated	Glass bottle	1848	25,177	1	ImageJ followed by fitting to exponential decay function to first 12 data points. Later data had no decay for long time	0.0053	[119]
Outbreak and episodic strains (N = 39 strains), all taken from patient biological cultures	Glass cover slips with 31% relative humidity	Glass	35			Extract data from table. Then zero order: 1/(survival time)	0.0016	[120]
Acinetobacter calcoaceticus var *anitratus* (3 strains)	Inoculating fingertips of healthy volunteers	Fingertips	1			ImageJ fitting to strain data from study Figure 2 followed by exponential decay function	2.5	[117]
Acinetobacter calcoaceticus var *anitratus* (3 strains)	Inoculating 2.5 cm^2 sheets of sterilized Formica	Formica surfaces	72			ImageJ fitting to strain data from study Figure 4 followed by exponential decay function	0.11	[117]

Table A8. Elimination rate data extracted for SARS-CoV and related coronavirus studies.

Species (Strain)	Study method	Medium	Duration (h)	Initial count (TCID$_{50}$ ml^{-1})	Final count (TCID$_{50}$ ml^{-1})	Data Extraction Method	Loss Rate Average (h^{-1})	Notes	Ref
SARS - CoV	Dried and then stored in polystyrene petri dish at 21–25°C. Residual infectivity based on cytopathic effect, as well as immunostaining of infected cells.	Plastic				ImageJ followed by fitting to exponential decay function	0.044	Good data quality	[124]
SARS – CoV (HKU 39849)	10 µl dried on 24 well plastic plate, incubated at multiple temperatures and relative humidities. Residual virus was determined by titration.	Plastic	672	10^7		ImageJ followed by fitting to exponential decay function	0.020	Paper looked at role of temperature and humidity. Reported loss rate is for results from 50% humidity and 22-25°C	[125]
SARS – CoV (CVU6109)	Samples serially diluted to initial concentrations, 5 ml added to material surface, absorbed at room temperature, and evaluated for survival after multiple survival times. Virus determination in a cell culture tube, after incubation at 37°C for 4 d.	Plastic (gown)	48	10^6	0	Zero order: 1/(Survival Time)	0.042	Inactivation faster with cloth than plastic. Experiment was semiquantitative (no time series; binary outcome)	[115]
SARS – CoV (CVU6109)	See above	Plastic (gown)	24	10^5	0	Zero order: 1/(Survival Time)	0.042	See above	[115]
SARS – CoV (CVU6109)	See above	Plastic (gown)	1	10^4	0	Zero order: 1/(Survival Time)	1	See above	[115]
SARS – CoV (CVU6109)	See above	Cloth (lab coat)	24	10^6	0	Zero order: 1/(Survival Time)	0.042	See above	[115]

Virus	Description	Material				Analysis	Value	Notes	Ref
SARS – CoV (CVU6109)	See above	Cloth (lab coat)	1	10^5	0	Zero order: 1/(Survival Time)	1	See above	[115]
SARS – CoV (CVU6109)	See above	Cloth (lab coat)	0.083	10^4	0	Zero order: 1/(Survival Time)	12	See above	[115]
Turkey CoV (French TCoV)	Storage in 400 µl suspension at 21.6°C in plastic vial with fetal calf serum added	Suspension				Exponential decay function to raw data	0.024	In liquid	[162]
Human CoV 229E	10 µl of suspension dried at 21°C in drying hood	Aluminum				ImageJ followed by fitting to exponential decay function	0.607		[161]
Human CoV OC43	See above	Aluminum				ImageJ followed by fitting to exponential decay function	1.6		[161]
TGEV, MHEV	Suspended in cell culture medium and placed on polished surfaces.	Stainless steel				Data provided as regression coefficients for log concentration ratios.	0.37	Surrogate mammalian strains for SARS-CoV. Median value for 18 experimental replicates at 20°C at each of 3 humidities (20, 50, and 80%) for the two strains (3 replicates per condition). The range of coefficients is wide (−0.052 to −1.083)	[163]
Human CoV 229E	Lysates were spread on coupons of material and then stored at room temperature, removed and evaluated on a plaque assay. The unit of analysis was plaque forming units	Teflon	120			ImageJ followed by fitting to exponential decay function	0.047	Objective of the study was to illustrate benefit of copper for virus mortality	[160]
Human CoV	See above	Silicone	120			ImageJ followed by fitting to exponential	0.035	See above	[160]

| 229E | | | rubber | | | | decay function | | | |
| Human CoV 229E | See above | | Stainless steel | 120 | | | ImageJ followed by fitting to exponential decay function | 0.055 | See above | [160] |

Table A9. Elimination rate data extracted relevant for *Streptococcus pyogenes*.

Species (Strain)	Study method	Medium	Duration (h)	Initial count (cfu)	Data extraction method	Loss rate (h⁻¹)	Ref
S. pyogenes (771 and JRS4)	Samples were grown to either mid-exponential or stationary phase, then transferred to 24 well plastic plates, then dried for 30 min, resuspended, and plated for viable colony counts.	Plastic	168.5	~10^8	First 4 times (median values for each time) extracted from study Figure 1B and D for planktonic results using ImageJ, exponential decay function fitted, then averaged (N=4)	0.078	[122]
S. pyogenes (771 and JRS4)	Samples grown to mid-exponential phase, and then transferred to paraformaldehyde-prefixed epithelial cells on glass plates to form biofilms. Then dried for 30 min, resuspended, and plated for viable colony counts.	Nasopharyngeal cells (in vitro)	168.5	~10^6	Median values for all times for each of two strains extracted for biofilm results (study Figure 1F) using ImageJ, exponential decay function fitted, then averaged	0.0026	[122]
S. pyogenes (8 clinical strains)	1 ml suspension sample in PBS was placed in glass bottle plugged with cotton wool and allowed to dry at room temperature. Resuspended and plated for viable colony counts. Count at d1 is initial count. Readings taken every 1 - 2 d from d 14 to d 37	Glass	888	10^8	Exponential decay function fitted for each individual strain in Table 1 [121]. ND values were set at 1/2 the lowest recorded value (5 cfu) and data after the last two ND results for each strain were deleted. Strain 7 was fitted based on first 4 data points only.	0.023	[121]
Streptococcus pneumoniae (ATCC 49619)	Known concentration in serum broth added to sterilized hands, then swabbed at given time points and evaluated by plating.	Skin	3	242–435	Nonzero data points from 3, 60, and 120 min were fitted to exponential decay curves. Initial time point (t_0) was excluded to account for loss due to swabbing. Only Serum broth results were used.	2.3	[116]
Streptococcus pneumoniae (ATCC 49619)	Known concentration in serum broth added to glass plate, then swabbed at given time points and evaluated by plating.	Glass	15	380–685	Data from times 0, 60, and 120 min were extracted from study Figure 1 with ImageJ and then fitted to exponential decay curves. Other time points excluded to match with skin study. Only Serum broth results used.	0.57	[116]

Table A10. Elimination rate data for *Bordetella pertussis*. All data were obtained from Mitscherlich and Marth [105].

Study method	Medium	Duration range	Duration assumed (h)	Initial count (cfu)	Data Extraction Method	Loss rate (h⁻¹)
Culture suspension was atomized at room temperature (RT).	Air	19 - 20 h	19.5		Zero order: 1/(Survival Time). Mean value.	0.051
Grown on saline suspension mixed with saliva, transferred to glass slides, and survival duration at RT recorded.	Glass	< 1 h - 1 d	12	10^8	Zero order: 1/(Survival Time). Using 1/2 maximum survival time (12 h)	0.083
Suspension atomized on surface and then stored at RT.	Plastic	3 - 5 d	96	10^8	Zero order: 1/(Survival Time). Using mean survival time (4 d)	0.010
10 µl suspension transferred to 2 cm² area skin of healthy volunteers.	Skin	<1 h - 6 h	3	10^8	Zero order: 1/(Survival Time). Using 1/2 maximum survival time (3 h)	0.33
10 µl suspension transferred to teddy bear hide and stored at RT.	Textile	1 - 4 d	60	10^8	Zero order: 1/(Survival Time). Using mean survival time (2.5 d)	0.017
10 µl suspension transferred to linen and stored at RT.	Textile	6 h - 3 d	39	10^8	Zero order: 1/(Survival Time). Using mean survival time (39 h)	0.026

Appendix 3.

Appendix to Chapter 3. Integrative statewide assessment of combined environmental and socioeconomic stressors versus chronic disease: California case study.

Spatial joining of CalEnviroScreen versus hospital data

The spatial levels of organization for the CalEnviroScreen (CES) versus ICD-9 datasets are different, requiring generation of a geographic correspondence. While CES is organized at the census-tract scale, the spatial reference point for ICD-9 is the zip code. Further, census tracts are precise polygon features while zip codes correspond to linear USPS route areas, post office box collection points, or large volume customer collection points, requiring further treatment to allow correspondence with polygon features [155]. Zip code tabulation areas (ZCTA) are polygons, developed and employed by the US Census to approximate the areal coverage of zip codes [164]. ZCTAs are similar but not identical to zip codes, such that treating them as the same would lead to data-comparability issues, as well as data loss. Because ZCTA is an aggregate scale that can be used to test for and address spatial autocorrelation, we incorporated all data up to the scale of ZCTA. The following procedure was employed to achieve this:

1. The ESRI 2014 zip code point data layer and the ESRI ZCTA boundary spatial polygon layer (based on 2010 Census) were downloaded from the web. Point estimates of zip codes were obtained from ESRI's zip code point file layer (http://www.arcgis.com/home/item.html?id=1eeaf4bb41314febb990e2e96f7178df) on June 14, 2015. ZCTA polygon estimates were downloaded on May 19, 2015.
2. ICD-9 zip codes that represent post office boxes or large volume customers were assigned to ZCTA in ArcMap by doing a spatial join (point to polygon) between the ESRI zip code point layer and the ESRI ZCTA polygon layer. That is, each ZCTA was defined to contain the corresponding zip code area, in addition to all PO boxes and large volume customers located within the geographic boundaries of that ZCTA. The ZCTA polygons were also automatically set to represent the spatial locations for all zip code areas (i.e., not PO boxes or large volume customers).
3. Each ZCTA has a total population estimate in the 2010 US Census. These data were downloaded from www.socialexplorer.com on May 13, 2015. This was used as the denominator term in the disease burden estimate .
4. ICD-9 diagnosis frequencies were summed from all corresponding zip codes by ZCTA. That is, diagnosis counts from all zip codes including PO Box and zip code areas were summed within the corresponding ZCTA.
5. ICD-9-based disease burden estimates were determined for each ZCTA, as described in Methods of Chapter 3
6. CalEnviroScreen (CES) results and the derived multivariate indicators (i.e., PCA results) were aggregated from the census-tract scale to the ZCTA scale using the population-weighted average of corresponding values. Data aggregation was performed on the census tracts and parameters from the original CES October, 2014 dataset after transformation, and also on the PCA results. The population-weighting scheme was

based on results from the MABLE/Geocorr12 population weighting algorithm [165], which were downloaded on September 30, 2015.

7. SAR requires all polygons to have adjacent neighbors; there can be no isolated polygons. Therefore, four geographically isolated ZCTAs having CES data were removed from the analysis: 93608 (Cantua Creek), 95257 (Wilseyville), 95925 (Challenge), and 96052 (Lewiston). All of these sites were in rural areas, with low to moderate CES scores (8.4 to 41.9), a low overall population (152 to 2196 residents in the 2010 census), and mostly Caucasian inhabitants (51% to 90%). Two ZCTAs were separated from other ZCTAs by narrow channels: 94501 (city of Alameda) and 92662 (Newport Beach). These two ZCTAs were manually joined to adjacent ZCTA polygons using the ArcMap editor.

8. Linear model and SAR model analyses were then performed on the resulting data at the ZCTA scale.

Correlation among diseases in ICD-9 database

In the study analyses, we opted to focus on hospitalization rate for a subset of specific diagnoses, rather than the total frequency of ICD-9 diagnoses (see Chapter 3 Data and Methods). The multivariate correlation structure for reported ICD-9 codes (count person^{-1} yr^{-1}) was examined across all zip codes (N = 1667) for each of the individual diagnoses, as well as the total rate of hospitalizations with at least one of the diagnoses (i.e., the "disease burden indicator" in Chapter 3). Data were square root transformed to achieve multivariate normality and analyzed via Pearson correlation coefficients. A positive correlation was observed for almost all pairwise comparisons (Table A12). Correlations were strong among total diagnosis rate, pneumonia, chronic obstructive pulmonary disease (COPD), asthma, myocardial infarction (MI; heart attack), and cerebral vascular accident (CVA; stroke) (Table A12, Figure A4). In principal component analysis of this correlation matrix, the first principal component (PC), which was negatively associated with all diseases, explained 45% of the variance in the data set. This PC was also very strongly associated with total reported codes (r = −0.90). These findings indicate that examination of the separate diagnoses would only provide limited unique information. This supported the development of a general indicator of disease burden that combines the different results. For this reason, we employed the disease burden indicator, as described in the text.

Table A11. Parameters used in analysis, description, original units, and transformation applied. All environmental and socioeconomic parameters were developed at the census-tract scale and are summarized here, following Faust et al. [36]. Hospital use was available at the zip code scale.

Name	Parameter Class	Description	Units	Years	Transformation
Ozone	Environmental	Average of daily maximum 8 hr concentrations that are above CA ambient standard of 0.07 ppm.	ppm	2009-2011	Cube root
PM$_{2.5}$	Environmental	Annual mean concentration	µg m^{-3}	2009-2011	Untransformed
Diesel PM	Environmental	Estimated daily emissions in a 4 km^2 area for a July workday from combined on-road and off-road sources	kg d^{-1}	2010	Cube root
Traffic density	Environmental	Sum of traffic volume (vehicle km) per unit time per road length within 150 m buffer of census tract	Vehicle km h^{-1} km^{-1}	2004	Cube root
Drinking water	Environmental	Drinking water contaminant metric , defined as the sum of percentiles for concentration of ten water contaminants	%	2005, 2009, 2013	ArcSin (Square root)
Pesticides	Environmental	Pesticide active ingredient used in production agricultural applications per unit area; includes 69 hazardous and volatile pesticides	lb mi^{-1}	2009-2011	Log$_{10}$[a]
Toxic release	Environmental	Modeled toxicity-weighted chemical concentrations released to air from industrial facilities; data were obtained from USEPA Toxic Release Inventory, following the Risk Screening Environmental Indicators tool.	lb yr^{-1}	2010	Log$_{10}$
Water body impairments	Environmental	Total number of pollutants listed as impairments in 303d Total Maximum Daily Load regulatory program	Unitless	2010	Cube root
Groundwater sites	Environmental	Weighted[b] sum of sites that could adversely impact groundwater, such as leaking underground chemical storage tanks, dry cleaners, dairies, airports. Sites were obtained from the GeoTracker data base maintained by the CA State Water Resources Control Board.	Unitless	2013	Log$_{10}$[a]

Cleanup sites	Environmental	Weighted[b] sum of sites with cleanup planned or underway due to hazardous chemicals. Sites were obtained from the EnviroStor data base maintained by the CA Department of Toxic Substances Control.	Unitless	2013	Cube root
Hazardous waste sites	Environmental	Weighted[b] sum of hazardous waste permitted facilities and sites that generate hazardous waste. Sites were obtained from the EnviroStor hazardous waste facilities data base maintained by the CA Department of Toxic Substances Control.	Unitless	2013	Log_{10}[a]
Solid waste sites	Environmental	Weighted[b] sum of sites or facilities containing solid waste. Site types include waste disposal and composting sites, waste tire sites, or other site types that may contain and release hazardous materials	Unitless	2013	Log_{10}[a]
Age	Socioeconomic	Percent of population that is < 10 or > 65 years old	%	2010	Square root
Education	Socioeconomic	Percent of population >25 years old without a high school degree or equivalent	%	2008-2012	Cube root
Linguistic isolation	Socioeconomic	Percent of population >14 years old that were not fluent English speakers	%	2008-2012	Cube root
Poverty	Socioeconomic	Percent of population with household income less than two times the federal poverty line	%	2008-2012	Square root
Unemployment	Socioeconomic	Percent of population >16 years old and workforce eligible that is not currently employed	%	2008-2012	Square root
Over 65	Demographic	Percent of population ≥ 65 years old (US Census)	%	2010	Square root
Disease diagnosis rate	Health outcome	Total number of instances of ICD-9 code reporting events per year per total population within zip code tabulation area	Events person^{-1} yr^{-1}	2008 - 2011	$sign(\sqrt{DDR}) * ln(|\sqrt{DDR}| + 1)$ [c]

a. To account for the presence of zeros, values were log_{10} transformed following the order-of-magnitude stabilizing procedure described in McCune and Grace [148]: $y = log_{10}(x + 10^C) – C$, where C = floor(log_{10}(min(x)))

b. Site weighting was based on expected hazard of the site type and distance from populated census block within the census tract [36]

c. Modulus transformation per John and Draper [149]

Table A12. Pearson correlation coefficients among the prevalence of ICD-9 codes for 14 specific disease diagnoses, examined in California Zip code tabulation areas. Had Dx: indicates the presence of any of the 14 diagnoses for a specific visit. The color of the boxes indicates strength and direction of the association, with red indicating negative association and blue indicating positive association.

|	Had Dx	Pneumonia	COPD	Asthma	MI	CVA	Diarrhea	Pancreatic cancer	Lung cancer	Breast cancer	Lymphoma	Leukemia	Depression	Schizophrenia
Pneumonia	0.87													
COPD	0.89	0.85												
Asthma	0.77	0.68	0.60											
MI	0.73	0.69	0.65	0.58										
CVA	0.77	0.72	0.70	0.61	0.73									
Diarrhea	0.16	0.13	0.06	0.17	0.15	0.13								
Pancreatic cancer	0.29	0.24	0.18	0.22	0.33	0.31	0.21							
Lung cancer	0.54	0.49	0.54	0.41	0.41	0.50	0.07	0.24						
Breast cancer	0.38	0.30	0.28	0.32	0.33	0.42	0.18	0.35	0.32					
Lymphoma	0.54	0.47	0.38	0.51	0.49	0.49	0.24	0.36	0.33	0.44				
Leukemia	0.31	0.28	0.22	0.27	0.31	0.33	0.19	0.28	0.24	0.27	0.31			
Depression	0.42	0.29	0.19	0.43	0.38	0.37	0.30	0.29	0.21	0.38	0.46	0.29		
Schizophrenia	0.53	0.39	0.36	0.47	0.29	0.28	0.28	0.23	0.16	0.21	0.38	0.18	0.51	
Low birth weight	0.04	0.07	−0.09	0.17	0.02	0.02	0.13	0.12	−0.02	0.08	0.14	0.10	0.11	0.21
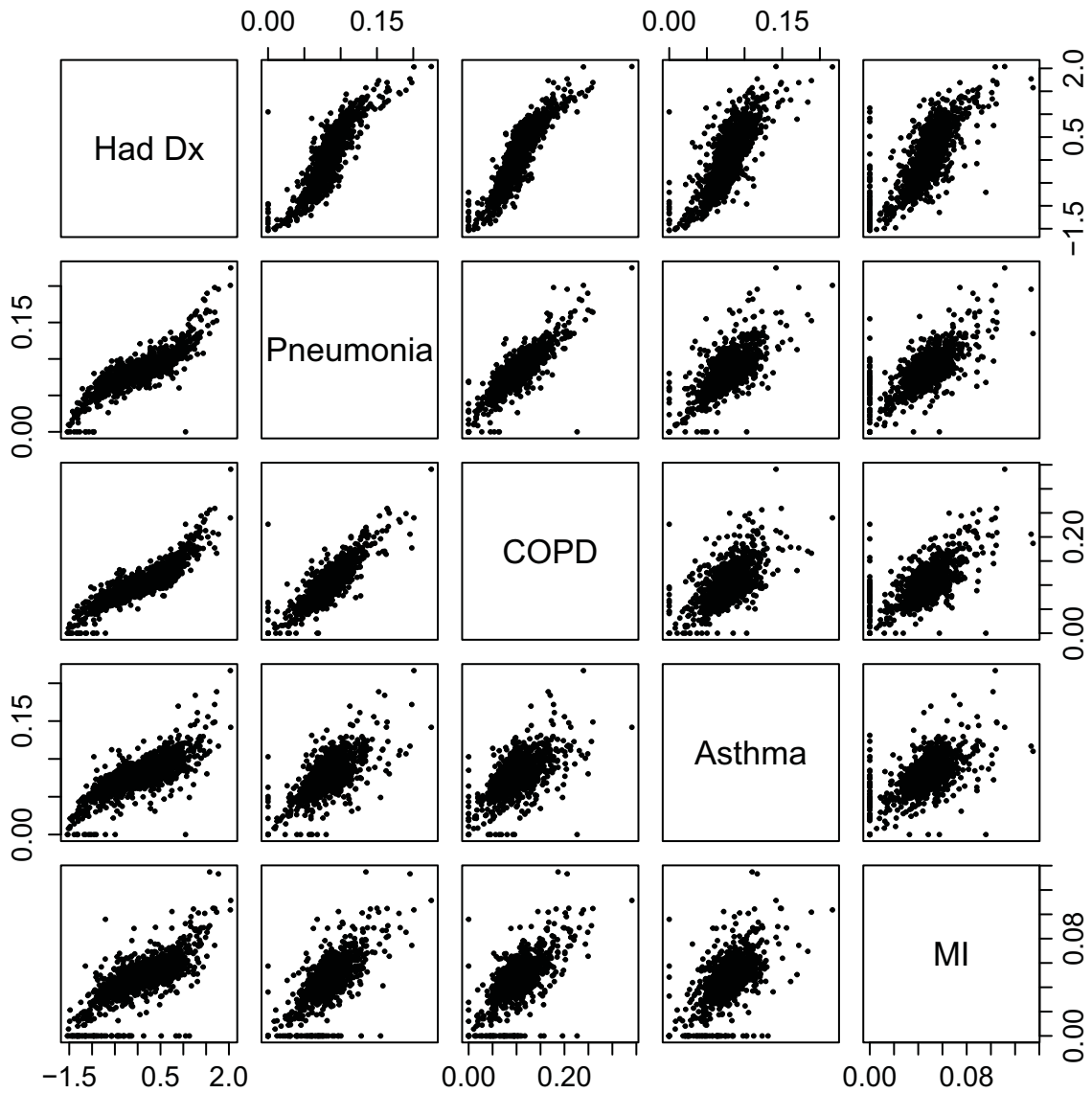
Figure A4. Scatterplot matrix of selected hospital visit rate (health outcome) parameters. All data were normalized to ZCTA total population and square root transformed. Had Dx: indicates the total rate of hospitalization for at least one of the diagnoses; COPD = chronic obstructive pulmonary disease; MI = myocardial infarction (heart attack).
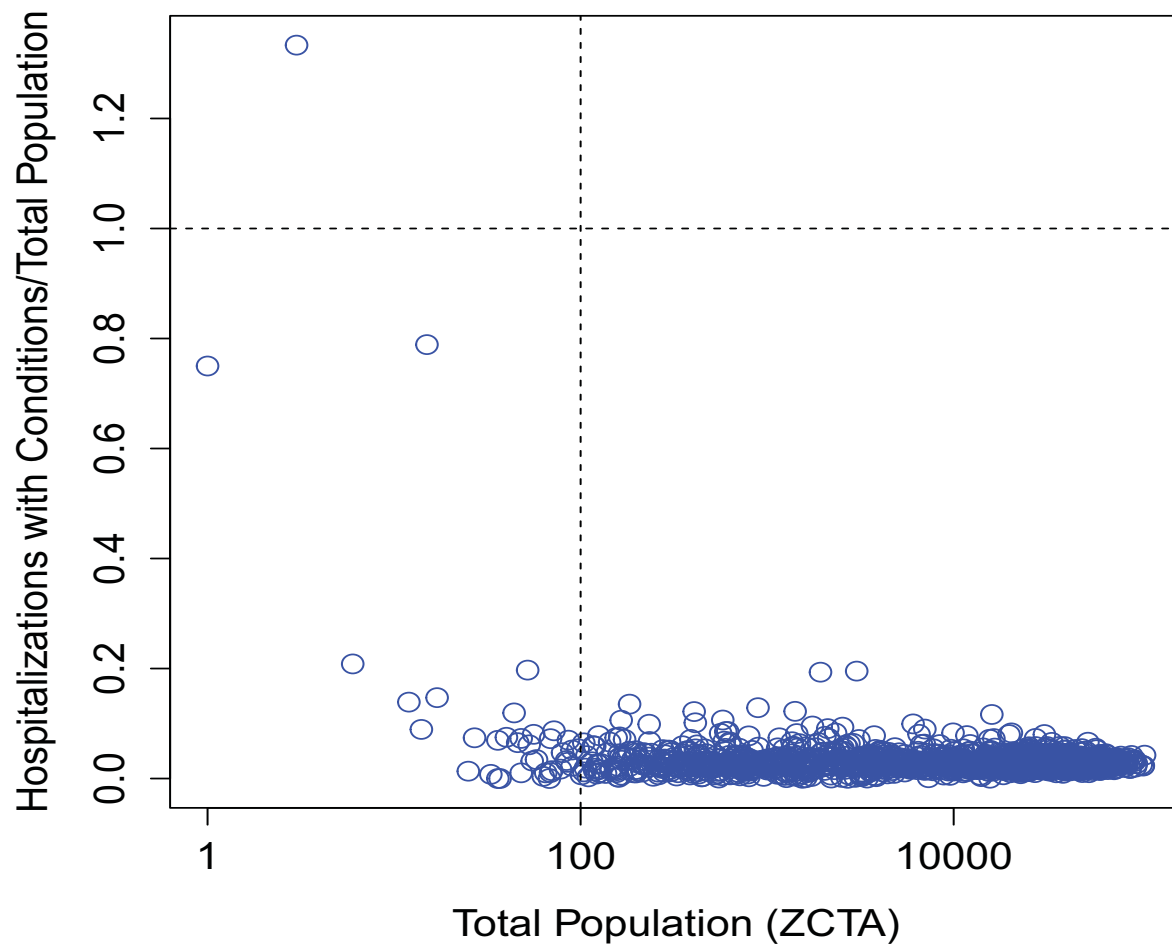
Figure A5. Total hospital ICD-9 diagnostic codes (count person^{-1} yr^{-1}) versus total population in each ZCTA (zip code tabulation area). Vertical dotted line: total population of 100; Horizontal dotted line: diagnostic code count of 1. The plot illustrates that for ZCTA with populations below 100, variability increased substantially. These ZCTA were thus removed from further analysis.
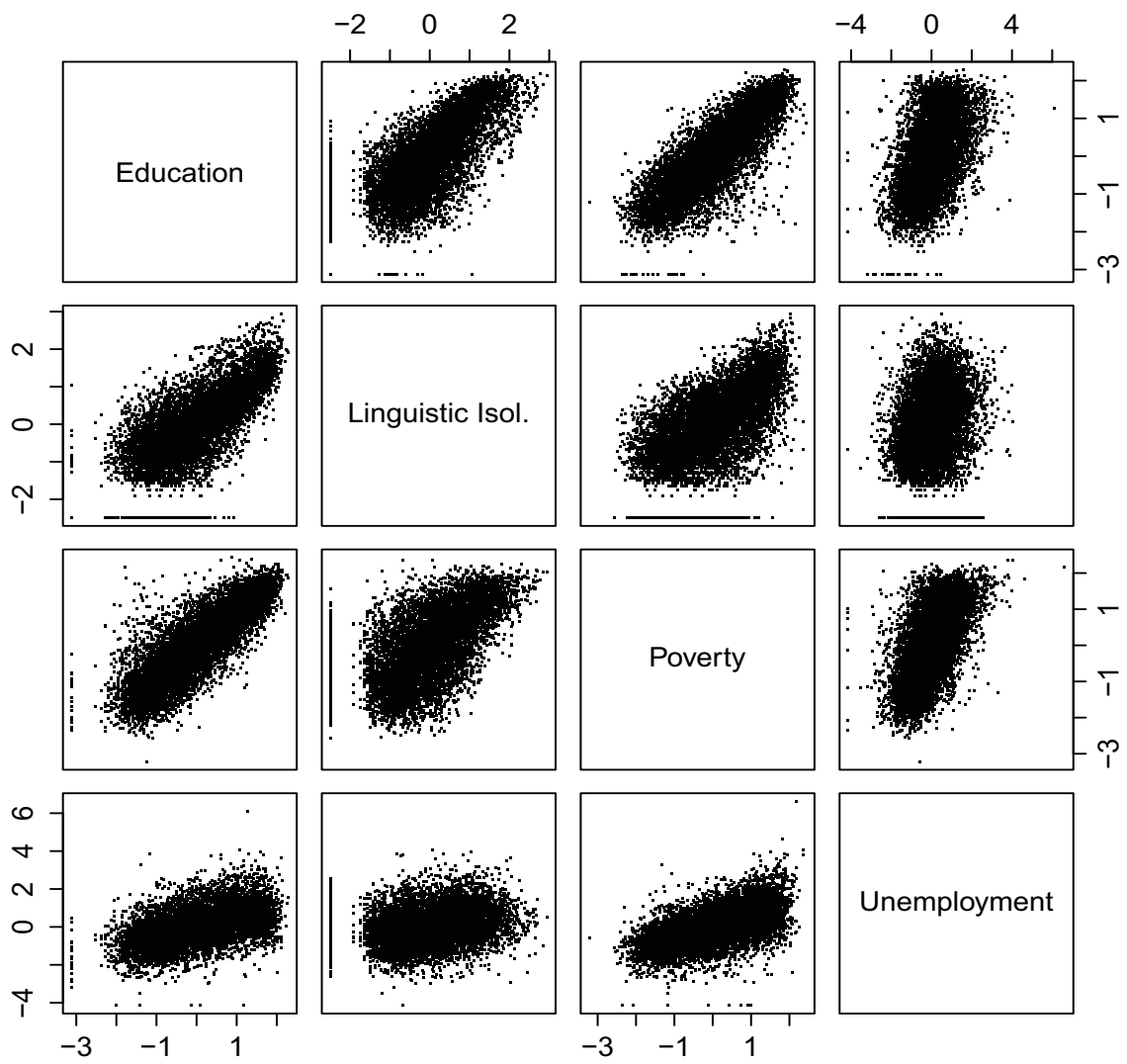
Figure A6. Scatterplot matrix indicates positive correlation among four socioeconomic vulnerability indicators.