Summer@ICERM 2024 Project Descriptions
Mathematical Models to Predict, Prepare, and Prevent

Amanda Harsy, Adam Schultze, Brittany Stephenson, and Cara Sulyok

Mathematical modeling allows one to describe dynamics happening in the real world by translating the knowledge and beliefs of interactions into the language of mathematics. For this reason, mathematical modeling is a useful technique for describing natural occurrences and addressing questions or testing hypotheses that may not be feasible to study in reality. The proposed Summer@ICERM 2024 research topics include projects centered around using mathematical modeling to make predictions and assess both associated preparations and necessary preventions within the fields of epidemiology, precision nutrition, and sports analytics.

1 Ordinary Differential Equation Models of Disease Transmission and Control in Long-Term Care Facilities

This project investigates the transmission of the gram-positive, endospore-forming bacterium *Clostridioides difficile* (*C. difficile*) in long-term care facilities (LTCFs). In particular, we will first formulate a system of ordinary differential equations (ODEs) modeling *C. difficile* spread in an LTCF that focuses on identifying the primary routes of transmission and then analyze the system in order to identify and compare control strategies for reducing healthcare-associated incidence rates.

C. difficile is one of the most common causes of healthcare-associated infections in the United States [42]. With almost half a million *C. difficile* infections (CDIs) and 29,000 deaths annually in the U.S., the corresponding attributable healthcare cost due to *C. difficile* is over four billion dollars per year [21, 22]. It has been shown that susceptibility to colonization increases with age, with persons older than 65 years of age 10 times more likely to contract *C. difficile* during an outbreak than their younger counterparts [31, 35]. Although the transmission of *C. difficile* has been widely studied in hospital settings [4, 12, 15, 48, 49, 50], age is not often a factor considered in these hospital models. According to a 2015 study by the Centers for Disease Control and Prevention, one out of every three cases of CDI occurs in patients 65 years or older, and over 100,000 cases develop among LTCF residents in the U.S. each year [20]. A review of epidemiological models of CDI from 2009-2021 [33] noted that very few models address transmission of *C. difficile* in LTCFs, such as nursing homes or skilled nursing facilities. In this work, we will more thoroughly examine the impact of an individual's age on the spread of *C. difficile* by modeling its transmission within LTCFs using multiple models, such as individual LTCFs, LTCFs in combination with hospitals, and LTCFs together with hospitals and the local community. These models will aid in understanding the transmission dynamics in healthcare settings and in determining how to control disease spread by comparing strategies that reduce incidence rates.

We expect a model of the spread of *C. difficile* in an LTCF to have notable differences from that of a hospital. Unlike a hospital that treats a variety of patients from a wide range of age groups, an LTCF focuses on the care of elderly or disabled individuals. Hospitals and LTCFs also differ in the amount of time that the patient or resident stays in the facility, the frequency of contacts with healthcare providers, and the reason for a patient's stay. Table 1 summarizes many of these key differences [40]. Due to the increased length of stay for residents in an LTCF, there will likely be a reduced risk of a CDI being introduced from an admission. While the advanced age of residents will increase the transmission rate of CDI, there will also likely be a reduction in antibiotic prescriptions in LTCFs associated with the treatment of acute illnesses

Hospitals	LTCFs
Treat patients with acute illnesses.	Maintain quality of life for residents with long-term problems/chronic illnesses.
Patients have daily contact with physician(s) and hourly monitoring by nursing staff.	Some residents may not see a medical provider for weeks if they are doing well medically.
Patients typically stay in their individual room.	Residents share a common dining hall and other spaces and interact with each other.
Patients are discharged when their illness or injury has been cured, or at least temporarily reduced.	Many residents are free to come and go.
Age varies depending on hosptial ward.	Average age is 85 years.

Table 1: Key differences between hospitals and LTCFs.

compared to that of a hospital. Finally, transmission routes will differ due to a reduction in interactions between residents and healthcare workers (HCWs) and an increase in contacts with other residents. These assumptions comprise the foundation for our model development.

Our ODE model in [50] was the first of its kind that explicitly incorporated the environmental contamination of surfaces. This model focused on the spread of *C. difficile* in a hospital ward and divided patients into four main classes: (1) resistant patients who cannot be infected with a CDI; (2) susceptible patients who have recently taken an antibiotic and can become infected with a CDI; (2) colonized, asymptomatic patients; and (4) diseased, symptomatic patients. In addition to the four patient classes, two compartments were included to consider the environmental transmission of *C. difficile*: high- and low-touch fomites, objects or surfaces that can harbor infectious agents. Results indicated that nearly three-quarters of nosocomial colonizations were from contact with a high-touch fomite, despite their additional, daily cleaning. Thus, enhanced efficacy of disinfection upon discharge and extra cleaning of high-touch fomites, reduced contact with high-touch fomites, and higher discharge rates could lead to a decrease in the incidence of colonized patients.

This project will utilize the same methodology as that of [50] to develop an ODE model of *C. difficile* transmission in an LTCF. First, the disease status classes of the residents will be updated to more accurately reflect the LTCF resident demographics. For example, susceptible residents will have a less restrictive definition as older age can cause a disruption in the gut microbiota making it easier for *C. difficile* endospores to colonize the large intestine, leading to a CDI. Moreover, we will include HCWs that interact with the residents. Another notable difference between the model in [50] and this project will be the parameter values. Parameters are the rates included in a mathematical model, derived from literature on studies completed in healthcare settings. Since we now consider disease spread in an LTCF, these values will need to be updated to ensure that our simulations accurately represent the dynamics in an LTCF rather than in a hospital.

Moreover, in previous models, we considered colonization only deriving from interactions with a contaminated surface, whether high- or low-touch, or with an infected HCW. Since residents of an LTCF are not confined to their rooms as patients in a hospital ward are, we will need to incorporate infected residents interacting with susceptible residents leading to disease transmission. This will necessitate that certain terms are changed in the mathematical model, leading to a different model structure. This new structure will likely lead to new recommendations to limit disease spread. With our model, we aim to answer the following questions:

1. Which control strategies most effectively mitigate the incidence of *C. difficile* in an LTCF?

2. What are the optimal levels of these strategies to reduce both healthcare-associated costs and total number of CDIs?

After developing the system of ODEs, we will use a combination of literature review and estimation techniques, such as the method of least squares, to determine appropriate parameter values for our model. We will then analyze the model with a global sensitivity analysis [37] using Latin Hypercube Sampling and Partial Rank Correlation Coefficients to determine which parameters lead to the greatest changes in the incidence of colonization. The results of the sensitivity analysis will inform potential control strategies to mitigate disease spread in the simulated LTCF and be the basis for applying Optimal Control Theory, a method that will allow us to determine time-varying, optimal parameters that most effectively reduce CDI incidence [36]. In our analysis, we include additional control strategies to consider such as decreased interactions between residents, representing resident lockdown and vaccination strategies. We also aim to repeat our analysis on expanded versions of the original model that will consider LTCFs in combination with hospitals and/or the local community. The expanded models will allow us to incorporate additional transmission pathways as residents can come and go freely and may be transferred to a hospital depending on their health status.

Model simulations and analysis will be completed using MATLAB. Students with knowledge of calculus and the interpretation of derivatives can succeed in this project even without explicit study of differential equations. Programming experience is a bonus but is not required for successful completion of this project.

2 Agent-Based Models to Evaluate Precision Nutrition Interventions through a Socioeconomic Lens

An individual's overall health is dependent upon many characteristics including age, demographics, physical activity, body-mass index (BMI), underlying health conditions, and socioeconomic status. To date, many guidelines to promote healthier eating have been aimed at the total population instead of at an individual level. By only considering the broader group, individuals' characteristics are often overlooked, and the precision in health guidelines is limited. More often than not, healthy eating guidelines have been based on large clinical studies that do not account for diverse individual nutrition needs and responses to diet [8, 34]. Additionally, it is widely known that there are large disparities in the quality of food environments for historically marginalized groups [19, 24, 52], but there is still much to learn about the extent to which healthy food prices and preferences can be linked to economic segregation [1]. Auchincloss et al. [1] showed that residential segregation, relative pricing of healthy foods, and dietary preferences have an impact on diet differential, and that if healthy foods were more favorably priced and encouraged, they may become better integrated into low-income areas. They also emphasize that income inequalities in diet are maintained by complex interrelated processes that are not fully understood and still need to be more completely studied.

Given the many complex, interrelated processes involved in diet and eating habits, researchers are now turning to precision nutrition research [28]. Precision nutrition focuses on targeting individual needs and differences when offering nutritional advice rather than relying on generic diet information [2, 34]. In particular, precision nutrition takes into account heterogeneity in an individual's nutrition by considering how factors such as social networks, environments, and culture impact what people eat.

To provide new insights into the field of precision nutrition, this project will investigate the role of an individual's socioeconomic status on both their overall health and the total population health through agent-based models (ABMs) of two Chicagoland villages: Broadview, IL and Clarendon Hills, IL. We will recreate the maps of the two neighborhoods, including streets, houses, restaurants, grocery stores, schools, hospitals, recreation facilities, and transport infrastructure and their occupants, using the software NetLogo [53]. These two communities are comparable in size but have very different racial breakdowns, socioeconomic statuses, and access to resources. The modeled demographic breakdowns will reflect reality, and individuals in the model will each have personal characteristics including age, sex, race and ethnicity, income, BMI, and medical information/disease status. Other features to be incorporated will include food budget parameters,

educational attainment, physical activity levels, and sleep tracking in order to incorporate decision making and daily behaviors and determine an overall health score for each agent. These individuals will interact with each other and their surroundings in their neighborhoods, such as grocery stores and restaurants (or lack thereof).

With the development of a model replicating real-world interactions and emphasizing the distinction between an individual's health and relation to quality food, we can examine how an individual's social network and resources affect their overall health score and examine impacts of socioeconomic variability on an individual's overall health. Our proposed ABMs will track individual characteristics, decision making, and daily behaviors in order to determine an overall health score for each agent. By modeling the interactions of an individual with both others and their surroundings, such as grocery stores and restaurants, we can identify how an individual's social network and resources affect their overall health score. Specifically, our work aims to answer the following questions:

1. How does the efficacy of the simulated interventions vary across different types of communities?

2. Do the simulated interventions reinforce or reduce pre-existing social disparities?

In answering these questions, the focus of our work is to identify effective interventions for improving public health through the use of individualized information. Additionally, targeted interventions to improve population health in villages with varying socioeconomic statuses can be assessed to reduce health disparities.

ABMs are based on computer simulations where the architect defines a set of rules governing how individuals interact with one another on a given spatial grid. Using an ABM allows for a variety of simulations and tracking at the agent (individual) level, which allows one to determine the impact of individual behaviors on the emerging dynamics of a population as a whole. Instead of offering generic advice, such as "eat a diet rich in fruits and vegetables," an ABM can include biological variability between individuals in response to nutrition and can also test targeted methods to tailor specific strategies for individuals or specific subgroups of communities. Moreover, with an ABM, we can identify interventions to increase the overall population health of the individuals in the communities. We have four main aims:

1. Create virtual communities patterned after the two Chicagoland villages with agents approximating the demographics of the populations inhabiting the two villages,

2. Examine the social, nutritional, and health networks and behaviors exhibited by the agents of each virtual community within and between neighborhoods and communities,

3. Classify which factors most notably influence health food availability and income differences in healthy eating, and

4. Determine which simulated interventions are most effective in the two environments in order to improve population health.

The more we understand about the factors and systems that impact individual decisions and behaviors, the more efficacious precision nutrition becomes.

We will begin by coding the two communities in question. There will be two separate models – one for each village. NetLogo works on a grid where each patch on the grid can be labeled to represent its type. The maps of these communities will be gridded and translated into the NetLogo Graphical User Interface (GUI), including streets, houses, restaurants, grocery stores, schools, hospitals, recreation facilities, and transport infrastructure. This will be important when we begin to incorporate submodels as the location of an individual's house and work will impact their access to particular grocery stores and restaurants. Stores and restaurants will be classified as healthy or unhealthy and expensive or inexpensive based on their real-world counterparts.

After developing the foundation of the ABMs, we will incorporate agents with individualized characteristics and randomly assign their locations for home and work/school. This will make it possible to identify an individual's access to food and beverage sources (supermarkets, convenience stores, sit-down restaurants,

and fast food/carry outs, or lack thereof) when they are at home and at work. There will also be assignments such as total household income, household food preferences, and proximity to a grocery store. We will include the same number of agents as the population for the two communities and will develop agents with these characteristics in both models to ensure that the agents are reflective of the true population. We will also model individuals that work outside the neighborhood where they reside. We will use data from the 2020 Census, U.S. Bureau of Labor Statistics, National Health and Nutrition Examination Survey (NHANES), Healthy Eating Index from National Health and Nutrition Examination Survey, and Reference USA to ensure that the breakdown of individuals and spatial grid accurately reflect the modeled neighborhoods. Together, these features will be used to calculate an individual health score, a continuous value between 0 (unhealthy) and 1 (healthy) that will be updated daily.

The base unit of the total health of the community is the individual's health score. Each individual's features will be combined with daily choices related to eating and exercise to arrive at a health score indicative of an individual's total health. Since the individual's diet is important in this health score, an individual's access to nutritious foods will also be indirectly incorporated. The agents' daily mobility, daily routines, and individual preferences necessarily affect the type of food that they eat. Individual health scores will be used to arrive at a global measure of community health. To assess the health of the community, we will use two calculations:

1. the sum of all of the agents' individual health scores, which we refer to as the population health, and

2. the median health score of all individuals in the community.

By running simulations on a daily time-step, each agent will have the opportunity to consume meals either at home or work/school or dine out as well as engage in physical activity. These choices will be made according to probabilities about the price, type of food, and remaining weekly food budget and will be inversely proportional to the location's geographic distance.

Every ABM consists of submodels that dictate the movement between agents. To replicate interactions within the community, we will code procedures that dictate how an individual operates daily, such as their food choices and physical activity habits. Moreover, these submodels determine the probability of individuals shopping at a particular grocery store or dining at a restaurant, based on an individual's income level. We anticipate that many of these submodels could be used in both community ABMs. Conversely, certain submodels will be community-specific, especially when incorporating probabilities. More than including the movement of individuals, it will be important to develop submodels for each individual such as meal timing and content, which will update the individual's energy level and overall health score.

After building the two community models, we will use the models to compare various health and economic interventions that increase both the overall population health and median health score. Some intervention strategies to consider will include changing consumer behavior by providing support for making healthier food choices; developing a healthy grocery store or restaurant in a food swamp, an area saturated with fast food; increasing individual/household salaries to determine impact on eating habits; and assessing how combinations of taxes, nutrition warning labels, and advertising affect purchase of certain food groups. To begin our assessment of interventions, we will first weigh the impact of each strategy implemented individually. Then, we will consider the interventions in combination to determine those overall strategies that have the largest impact on increasing the population health score and median individual health score.

Due to the stochasticity embedded in ABMs and to best assess the impact of the control intervention strategies, we will run 100 iterations over a one-year simulated time period with each combination of parameter values (representing different control strategies). The model will be run on a one-day time-step using the BehaviorSpace tool in NetLogo, which allows us to specify the various parameter combinations we want to simulate and the resulting outputs of interest. Because of the computing power required to run these simulations, we will utilize parallel processing and high performance computing. Throughout testing these interventions, we aim to understand whether these strategies reinforce or reduce pre-existing social disparities and whether optimal strategies remain consistent in both communities.

Students working on this project will be responsible for tasks such as conducting background research about demographic breakdowns of Broadview, IL and Clarendon Hills, IL; complete coding of patches

(streets, houses, restaurants, grocery stores, schools, recreation facilities, and transport infrastructure) and of full environments in NetLogo; developing submodels; determining the health score calculation based on diet, exercise, and medical conditions; and simulating and testing intervention strategies both individually and in conjunction using BehaviorSpace. Students do not need programming experience to be successful in this project, but it is a bonus.

3 Predictive Modeling and Analysis of Sports Using Linear-Algebra-Based Models

This project involves creating and testing the accuracy of different linear-algebra-based models from sports data. There are many methods for predicting and ranking sports [3, 6, 7, 10, 11, 16, 23, 25, 27, 29, 30, 46, 51], and there are many algorithmic rating sites, such as www.mratings.com, www.thepredictiontracker.com, and www.masseyratings.com, devoted to sports ranking. Among the many mathematically-inspired sports ranking systems, linear-algebra-based methods are some of the most elegantly simple. A variety of methods are used, and some are dependent on complex mathematical or statistical tools, but weighted Colley and Massey models, in addition to PageRank and Markov chain models, are very accessible for undergraduates who have taken linear algebra. Colley and Massey models essentially involve setting up and solving a system of equations to solve for a rating vector for each team.

Colley Method

The Colley method was developed by Wesley Colley. Instead of using regular win percentage $\frac{w_i}{t_i}$ to calculate the quality of a team, where w_i is the total number of wins for Team i and t_i is the total number of games played by Team i, Colley applies Laplace's Rule of Succession to calculate win percentage by $\frac{w_i + 1}{t_i + 2}$. Note that with regular win percentage, an untried team has a $0/0$ in percentage, but with this adjustment, an untried team has a win percentage of 50%. Colley also incorporates strength of schedule into his rating system by approximating the number of games played by the sum of the rating of the teams played [14].

To set up his system, Colley sets the rating for each team, r_i, equal to its Colley win percentage. That is, $r_i = \frac{w_i + 1}{t_i + 2}$. In order to incorporate strength of schedule, Colley does the following algebraic manipulation to get the following relationship:

$$(2 + t_i)r_i = 1 + \frac{w_i - l_i}{2} + \frac{w_i + l_i}{2}, \tag{1}$$

where l_i denotes the total number of losses for Team i. Note that $\frac{w_i + l_i}{2}$ is the total number of games played by team i divided by two. This is where Colley made an adjustment to his system in order to incorporate strength of schedule. He does this by replacing $\frac{w_i + l_i}{2}$ with the sum of the ratings of the teams played by Team i. Thus, our new Colley system has the following set up for each Team i:

$$(2 + t_i)r_i = 1 + \frac{w_i - l_i}{2} + S, \tag{2}$$

where S is the sum of the ratings of teams played by Team i.

These equations form a nice symmetric matrix system for n teams, $C\mathbf{r} = \mathbf{b}$, where C is an $n \times n$ matrix holding the coefficients of the r_i's in each equation and $\mathbf{b}$ is an $n \times 1$ vector containing the constant portions of each equation. After solving this system, we order the team rankings by the value of the ratings with the largest rating corresponding with the top ranked team.

Massey Method

Colley's method doesn't incorporate scores into its system. Conversely, Kenneth Massey's method of sports rating creates a system of equations which is based on the score of the games played. Massey originally created this method for ranking college football teams [32, 43]. Using the Massey method, for each game the difference in the ratings correlates to the point differential [32]. This is represented as a system: $r_i - r_j = b_k$, where r_i is the rating for team i, r_j is the rating for team j, and b_k is the score differential for game k. This method makes it easy to incorporate a cap on the score differential to potentially curb the impact of blowouts.

Note that each row of our system of equations represents a particular game from the dataset. Our columns represent the different teams being ranked. We set this system up as $X\mathbf{r} = \mathbf{b}$, where X is a matrix with m rows (one for each game) and n columns (one for each team), and $\mathbf{b}$ is an $m \times 1$ vector with the point differentials from each game. This usually means in practice that our system has more rows than columns and is often inconsistent (especially since it is rare for a team to beat another team twice by the same score). Therefore, the Massey method uses the method of least squares to solve its system [38]. To find a least squares solution, we solve the normal equation which is created by left multiplying both sides of our original equation $X\mathbf{r} = \mathbf{b}$ by the transpose of X [10]. The transpose matrix, denoted X^T, is a matrix whose rows are the columns of the original matrix X. Thus, we now solve the system $X^T X \mathbf{r} = X^T \mathbf{b}$.

Sometimes there are multiple least squares solutions [10]. To fix this, Massey needed to force his system to have a unique solution. He did this by replacing one of the rows with an equation which was not in the span of the other rows of the system (that is, not a linear combination of the other rows). He could replace any row, but for consistency, Massey replaces the last row of the system with $\sum_{i=1}^{m} r_i = 0$ [10, 23]. This ultimately means that the sum of the ratings should be zero. Like with the Colley method, after we solve our system of equations, we put the ratings in order from largest to smallest in order to create our ranking of the teams.

While at their most basic level Colley and Massey methods are useful for sports rankings, they are not particularly strong at predicting future outcomes of games [3, 10]. One way to improve these methods for ranking and predicting future outcomes is by introducing weights to these systems [6, 10, 9, 23]. For example, we can weight a win achieved on the road higher than winning a home game or weight wins at the end of the season higher than wins at the start of the season. In this research, we usually incorporate multiple weights and multiple factors and then use cross validation and sensitivity analyses to help us improve our models.

Markov Processes

Markov processes are another popular and accessible model used in sports analytics [7, 25, 27, 29, 30, 46]. This modeling technique is used to create state-based stochastic models that describe the sequences of possible events where the probability of each event depends solely on the state at which it was previously in. Similar to the Colley and Massey models, the team with the highest probability is ranked first, the team with the second highest probability is ranked second, etc. The process of finding a "ranking" of web pages can be extended to sports ranking by representing teams as "states." There are a wide range of Markov models used in sports analytics [16, 25, 29, 30, 47]. Absorbing Markov chains are used to model Major League Baseball batting and run scoring [16, 47]. Kvam and Sokol developed a Markov model which was used to rank NCAA basketball teams and can easily be extended to other sports settings [30]. This modeling technique aims to predict the probability that a team is better than its opponent, given particular features or statistics (e.g., that the home team won by a certain number of points). We can then create a matrix that encodes the probability of switching from one state (in this case, "teams") to another, called a transition or stochastic matrix. These probabilities can be approximated using past-game data. Once the transition matrix is created, we can determine the long-term behavior of the system represented by the probability vector called the steady-state vector. The teams are then ranked based on their steady-state probability value. This model can be extended in a variety of ways. For example, instead of looking at home versus

away, we can incorporate statistics such as win streaks, conference versus non-conference wins, and time during the season.

Proposed Projects

Each linear-algebra-based model has different strengths and weaknesses, and certain models may be better suited for a particular sport or dataset. Mathematicians have explored and compared these methods in sports settings [3, 6, 7, 10, 9, 11, 13, 25, 27, 29, 30, 41, 45, 46]. There are a variety of questions that can be explored using these models, and techniques from this research can translate to broader scenarios that involve ranking or predictive modeling. Some possible projects for this program include the following:

- Create and compare various linear-algebra-based ranking models on their ability to predict end-of-season outcomes in team sports such as soccer, hockey, baseball, football, and/or individual sports such as tennis, track and field, and skating. These models could incorporate features such as win streaks, conference wins, pressure, rule changes, etc. In particular, rule changes in a sport provide interesting motivation for these models. For example, since 2006, the National Hockey League implemented new overtime rules. When a game goes into overtime, each team receives one point in the league standings, even if they ultimately end up losing the game. Therefore, in addition to incorporating other factors into a Colley, Massey, or Markov model, students could explore whether or not having more ties or overtime games is more predictive for the end-of-the-season standings. Students, with the guidance of the organizers, may choose which sport(s) and features they would like to explore.

- Create and compare various linear-algebra-based ranking models on their ability to predict end-of-competition rankings or explore bias in judged sports such as gymnastics, diving, skateboarding, freestyle skiing, figure skating, or equestrian sports. Research has shown that there are known biases based on gender, competition order, difficulty of routine, the nationality of the judge, and country of origin of the judge [5, 17, 18, 26, 39, 44]. Creating weighted linear-algebra models could provide insight into some of these biases or parts of a judged score. Markov chains may also be useful in analyzing subjectively judged sports. For example, Boen found that some judges are influenced by other judge scores [5]. Students, with the guidance of the organizers, may choose which sport(s) and features they would like to explore.

This research involves linear algebra, programming, and some data analytics. Students do not need to have programming experience or to have explicitly studied linear algebra to be successful in this project, but it is a bonus.

References

[1] A.H. Auchincloss, R.L. Riolo, D.G. Brown, J. Cook, and A.V.D. Roux. An agent-based model of income inequalities in diet in the context of residential segregation. *American Journal of Preventive Medicine*, 40(3):303–311, 2011.

[2] D. Barh. Artificial intelligence in precision nutrition: From concept to applications. *Academic Press*, 2020.

[3] D. Barrow, I. Drayer, P. Elliott, G. Gaut, and B. Osting. Ranking rankings: an empirical comparison of the predictive power of sports ranking methods. *Journal of Quantitative Analysis in Sports*, 9(2):187–202, 2013.

[4] J. Bintz, S. Lenhart, and C. Lanzas. Antimicrobial stewardship and environmental decontamination for the control of *Clostridium difficile* transmission in healthcare settings. *Bulletin of Mathematical Biology*, 79(1):36–62, 2017.

[5] F. Boen, K. Van Hoye, Y. Vanden Auweele, J. Feys, and T. Smits. Open feedback in gymnastic judging causes conformity bias based on informational influencing. *Journal of sports sciences*, 26(6):621–628, 2008.

[6] A. Buente, M. Koronkiewicz, H. Shultz, and A. Harsy. Does the dropped score matter? Analyzing NCAA Division II men's golf using the Massey method. *The PME Journal*, 15(9):385–392, 2023.

[7] B. Bukiet, E.R. Harold, and J.L. Palacios. A Markov chain approach to baseball. *Operations Research*, 45(1):14–23, 1997.

[8] C.L. Bush, J.B. Blumberg, A. El-Sohemy, D.M. Minich, J.M. Ordovás, D.G. Reed, and V.A.Y. Behm. Toward the definition of personalized nutrition: a proposal by the American Nutrition Association. *Journal of the American College of Nutrition*, 39(1):5–15, 2020.

[9] T.P. Chartier. *When Life is Linear: From Computer Graphics to Bracketology*. Mathematical Association of America, 2015.

[10] T.P. Chartier, E. Kreutzer, A. Langville, and K. Pedings. Bracketology: How can math help? *Mathematics and Sports*, 43(67):55–70, 2010.

[11] T.P. Chartier, E. Kreutzer, A.N. Langville, and K.E. Pedings. Sensitivity and stability of ranking vectors. *SIAM Journal on Scientific Computing*, 33(3):1077–1102, 2011.

[12] J. Codella, N. Safdar, R. Heffernan, and O. Alagoz. An agent-based simulation model for *Clostridium difficile* infection control. *Medical Decision Making*, 35(2):211–229, 2014.

[13] P. Coletti and A. Pilkington. A comparison of the performance of the generalized google pagerank model (gem) to other ranking systems on the NCAA basketball tournament. In *Proceedings of MathSport International 2017 Conference*, page 97, 2017.

[14] W. Colley. Colley's bias free college football ranking method, 2002.

[15] E.M.C. D'Agata, P. Magal, D. Olivier, S. Ruan, and G.F. Webb. Modeling antibiotic resistance in hospitals: the impact of minimizing treatment duration. *Journal of Theoretical Biology*, 249:487–499, 2007.

[16] D.A. D'Esopo and B. Lefkowitz. The distribution of runs in the game of baseball. Technical report, SRI Internal Report, 1960.

[17] J.W. Emerson and S. Meredith. Nationalistic judging bias in the 2000 olympic diving competition. *Math Horizons*, 18(3):8–11, 2011.

[18] L.C. Findlay and D.M. Ste-Marie. A reputation bias in figure skating judging. *Journal of Sport and Exercise Psychology*, 26(1):154–166, 2004.

[19] S.E. Fleischhacker, K.R. Evenson, D.A. Rodriguez, and A.S. Ammerman. A systematic review of fast food access studies. *Obesity Reviews*, 12(5):460–471, 2011.

[20] Centers for Disease Control and Prevention. Nearly half a million americans suffered from *Clostridium difficile* infections in a single year. https://www.cdc.gov/media/releases/2015/p0225-clostridium-difficile.html. Accessed: 2022-12-15.

[21] Centers for Disease Control and Prevention. Nearly half a million americans suffered from *Clostridium difficile* infections in a single year, 2015.

[22] Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States. Technical report, U.S. Department of Health and Human Services, Atlanta, GA, 2019.

[23] J.A. Gallian. *Mathematics and Sports*, volume 43. MAA, 2010.

[24] B. Garcia-Bulle, A.L. Horn, B.M. Bell, M. Bahrami, B. Bozkaya, A. Pentland, K. de la Haye, and E. Moro. You are where you eat: effect of mobile food environments on fast food visits. *medRxiv*, Preprint, 2022.

[25] A. Harsy, H. Campos-Chavez, W. deBolt, M. Mena, J. Prince, A. Alramahi, R. Dudzinski, S. Thrawl, and A. DeLegge. Predictive modeling and analysis of hockey using Markov chains. *Mathematics and Sports*, 4(1), 2022.

[26] S. Heiniger and H. Mercier. Judging the judges: evaluating the accuracy and national bias of international gymnastics judges. *Journal of Quantitative Analysis in Sports*, 17(4):289–305, 2021.

[27] N. Hirotsu and M. Wright. A Markov chain approach to optimal pinch hitting strategies in a designated hitter rule baseball game. *Journal of the Operations Research Society of Japan*, 46(3):353–371, 2003.

[28] D. Kirk, C. Catal, and B. Tekinerdogan. Precision nutrition: a systematic literature review. *Computers in Biology and Medicine*, 133:104365, 2021.

[29] J. Kolbush and J. Sokol. A logistic regression/markov chain model for American college football. *International Journal of Computer Science in Sport*, 16(3):185–196, 2017.

[30] P. Kvam and J.S. Sokol. A logistic regression/Markov chain model for NCAA basketball. *Naval Research Logistics (NrL)*, 53(8):788–803, 2006.

[31] J.H. Kwon, M.A. Olsen, and E.R. Dubberke. The morbidity, mortality, and costs associated with *Clostridium difficile* infection. *Infectious Disease Clinics of North America*, 29(1):123–134, 2015.

[32] A.N. Langville and C.D. Meyer. *Who's #1?: the science of rating and ranking*. Princeton University Press, 2012.

[33] C. Lanzas, M. Jara, R. Tucker, S. Curtis, and Modeling Infectious Diseases in Healthcare Program. A review of epidemiological models of *Clostridioides difficile* transmission and control (2009-2021). *Anaerobe*, 74, 2022.

[34] B.Y. Lee, J.M. Ordovás, C.A.M. Parks, and et al. Research gaps and opportunities in precision nutrition: an NIH workshop report. *American Journal of Clinical Nutrition*, 2022.

[35] D. Leffler and T. Lamont. *Clostridium difficile* infection. *The New England Journal of Medicine*, 372(16):1539–1548, 2015.

[36] S. Lenhart and J.T. Workman. *Optimal control applied to biological models*. CRC Press, New York, 2007.

[37] S. Marino, I.B. Hogue, C.J. Ray, and D.E. Kirschner. A methodology for performing global uncertainty and sensitivity analysis in systems biology. *Journal of Theoretical Biology*, 254:178–196, 2008.

[38] K. Massey. Statistical models applied to the rating of sports teams. *Bluefield College*, 1997.

[39] H.N. Morgan and K.W. Rotthoff. The harder the task, the higher the score: Findings of a difficulty bias. *Economic Inquiry*, 52(3):1014–1026, 2014.

[40] Division of Vital Statistics. Post-acute and long-term care providers and services users in the United States. Technical report, National Center for Health Statistics, Hyattsville, MD, 2017-2018.

[41] R.D. Pasteur. Extending the colley method to generate predictive football rankings. *Mathematics and Sports*, pages pp–117, 2010.

[42] M.A. Rubin, M. Jones, M. Leecaster, K. Khader, W. Ray, A. Huttner, B. Huttner, D. Toth, T. Sablay, R.J. Borotkanics, D.N. Gerding, and M.H. Samore. A simulation-based assessment of strategies to control *Clostridium difficile* transmission and infection. *PLoS ONE*, 8(11):1–11, 2013.

[43] J.E. Salzman and J.B. Ruhl. Who's number one? *Environmental Forum*, 26(6):36, 2009.

[44] A. Sandberg. Competing biases: Effects of gender and nationality in sports judging. *Stockholm School of Economics, mimeo*, 2014.

[45] J. Shi and X. Tian. Learning to rank sports teams on a graph. *Applied Sciences*, 10(17):5833, 2020.

[46] Z.J. Smith. *A Markov chain model for predicting major league baseball*. PhD thesis, University of Texas at Austin, 2016.

[47] J. Sokol. A robust heuristic for batting order optimization under uncertainty. *Journal of Heuristics*, 9:353–370, 2003.

[48] B. Stephenson, C. Lanzas, S. Lenhart, and J. Day. Optimal control of vaccination in an epidemiological model of *Clostridium difficile* transmission. *Journal of Mathematical Biology*, 75(6-7):1693–1713, 2017.

[49] B. Stephenson, C. Lanzas, S. Lenhart, E. Ponce, J. Bintz, E.R. Dubberke, and J. Day. Comparing intervention strategies for reducing *Clostridioides difficile* transmission in acute healthcare settings: an agent-based modeling study. *BMC infectious diseases*, 20(1):1–17, 2020.

[50] C.J. Sulyok, L. Fox, H. Ritchie, C. Lanzas, S. Lenhart, and J. Day. Mathematically modeling the effect of touch frequency on the environmental transmission of *Clostridioides difficile* in healthcare settings. *Mathematical Biosciences*, 340:108666, 2021.

[51] B. Vaziri, S. Dabadghao, Y. Yih, and T.L. Morin. Properties of sports ranking methods. *Journal of the Operational Research Society*, 69(5):776–787, 2018.

[52] R. E. Walker, C. R. Keane, and J. G. Burke. Disparities and access to health food in the United States: a review of food deserts literature. *Health Place*, 16(5):876–884, 2010.

[53] U. Wilensky. *NetLogo*. Center for connected learning and computer-based modeling, Northwestern University, Evanston, IL, 1999.