Published in final edited form as:

Curr Opin Infect Dis. 2015 August ; 28(4): 384–391. doi:10.1097/QCO.0000000000000179.

Expanding the statistical toolbox: Analytic approaches for cohort studies with healthcare-associated infectious outcomes

Rebecca A. Pierce, MS, BSN¹, Justin Lessler¹, and Aaron M. Milstone, MD, MHS^{2,3}

¹Department of Epidemiology, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD ²Department of Pediatrics, Division of Pediatric Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, Maryland ³Department of Hospital Epidemiology and Infection Control, the Johns Hopkins Hospital, Baltimore, MD

Abstract

Purpose of review—Healthcare-associated infections (HAIs) are a leading cause of adverse patient outcomes. Further elucidation of the etiology of these infections and the pathogens that cause them has been a primary goal of research in infection control and healthcare epidemiology. Longitudinal studies, in particular, afford a range of statistical methods to better understand the process of pathogen acquisition or HAI development. This review intends to convey the scope of available statistical methodology.

Recent findings—Despite the range of methods available, logistic regression remains the dominant statistical approach in use. Poisson regression, survival methods, and mechanistic (mathematical) models remain underutilized. Recent studies that use these approaches are looking beyond associations to answer questions about the timing, duration, and mechanism of infectious risk.

Summary—Logistic regression remains an important approach to the study of healthcare associated infections, but in the context of cohort studies, it is most appropriate for short observation periods, where mechanism is not of primary interest. Additional statistical methodologies are available to build upon risk factor analysis to better inform the process of risk and infection in the hospital setting.

Keywords

Healthcare-associated infection; logistic regression; Poisson regression; survival analysis; mathematical models

Introduction

Healthcare-associated infections (HAIs) are a major threat to patient safety. A large literature exists on the etiology of HAIs, the transmission dynamics and risk factors for

Corresponding author: Rebecca A. Pierce MS, BSN, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205, Phone: 908-432-4197, rpierc13@jhu.edu.

Potential conflicts of interest: All authors report no conflicts.

acquiring specific healthcare-associated pathogens, and the impact of interventions to prevent HAIs and organism transmission. Most available data are from observational cohort studies. Randomization of hospital and patient-level exposures is often not feasible or presents ethical dilemmas. A number of statistical approaches are available for longitudinal data, and each method can inform distinct aspects of the infection or colonization process. Logistic regression is the most frequent method of analysis. Other methods include Poisson regression, survival analyses, or mechanistic (mathematical) models. Below we discuss approaches to the analysis of data from cohort studies and important considerations for each approach.

Logistic Regression

Commonly employed to analyze case-control studies, logistic regression (LR) has many other applications. Its use in healthcare epidemiology is widespread as it is easily performed and interpreted by the infection control community. In a review of observational studies published in 2014 from two major journals within the field (*American Journal of Infection Control* and *Infection Control and Hospital Epidemiology*), almost half (43%) reported longitudinal cohort designs for the study of HAI occurrence or pathogen-acquisition, as opposed to case-control or interrupted time series studies. Of these cohort studies, 63% used LR for the primary analysis (Figure 1). However, three important limitations require consideration when using LR for cohort studies in healthcare epidemiology.

First, LR necessitates the classification of outcomes (e.g. infected or not infected) at some arbitrary point in time (e.g. by time of discharge). This approach informs whether or not an event occurred, as opposed to when an event occurred [1, 2]. There are many scenarios where the timing of an event will have important implications and could be of primary interest. Consider the demonstrated impact of antibiotic exposure and risk of *Clostridium difficile* infection [3*]. In this case, we know the *if*, what we are missing is the *when*. The primary question of interest then may not be “is antibiotic exposure associated with infection?”, but instead “what antibiotics or patterns of antibiotic treatment are associated with faster progression to infection?” or “when is the period of highest risk post-exposure?” [4**]. These questions are not answered easily with LR.

Second, simple LR does not account for the change in risk imposed by an exposure over time. For example, risk factors and causative pathogens for early-onset sepsis in neonates (0–3 days of life) are notably different than those for late-onset sepsis (4–120 days of life), with the former associated with vertical transmission processes and the latter with hospital exposures [5]. Thus, without awareness of these time-dependent trends, a risk factor analysis for all sepsis types by time of discharge could be misleading.

Finally, LR does not account for varying lengths of stay (LOS) typically observed in a cohort of hospital patients. Investigators will often address this by adjusting for LOS in multivariable models to improve comparability of exposed and unexposed patients [6–8]. When doing so, investigators must carefully consider the composition of comparison groups at different time points during follow up. Comparisons made among those with a prolonged length of stay may differ from those made early in follow up among a potentially more

heterogeneous group of individuals who have not yet been removed from observation through death or discharge. In addition, LOS may be a common effect of a risk factor and healthcare-associated infectious outcome and not a confounder of the association [9–12]. Therefore, conditioning for LOS in LR, although intuitive, may introduce bias [9, 13, 14].

In general, results from LR may approximate those from alternative time to event approaches described below when the following conditions are met: the outcome event is rare, effect size is weak, and the follow-up period is short [15**, 16]. These conditions are unlikely to hold in many scenarios, particularly in the context of prolonged hospital stays or outbreaks. Therefore, an LR approach to hospital-based longitudinal studies is best applied to research questions targeting narrow time windows, in which risk is constant. Below we suggest alternative approaches to the analysis of data from cohort studies along with recommendations for their use. Though complexity increases with each of these approaches, overcoming obstacles inherent in hospital data is important to advance the science of infection prevention.

Poisson Regression

Poisson regression models can approximate estimates obtained from logistic or log-binomial regression models [17, 18]. Unlike LR, Poisson regression can account for time under observation by including a denominator of person-time at risk. These Poisson models estimate incidence rates that inform how quickly events are occurring in the population and account for each individual's contribution to person-time [19]. A common approach is to analyze aggregated numbers of events and total person-time. For example, Iwamoto et al. [20*] used a combination of Methicillin-resistant *Staphylococcus aureus* (MRSA) infection data from the Active Bacterial Core Surveillance system and US census-based person-time estimates and then applied Poisson regression to estimate changes in rates over time. Poisson regression can also be used with individual-level data [4**]. The Poisson model is highly adaptable and can be fit in a manner that mimics survival approaches (discussed below) [21].

The Poisson model has strict assumptions about the variability in event frequency that must be considered [19]. Hospital data often include rare events with greater variability than the model allows. This is particularly problematic when multiple mechanisms, some of which are not easily measured, are responsible for the variability in the outcomes. For instance, acquisition of MRSA in a neonatal intensive care unit may be related to healthcare worker-mediated transmission directly to neonates, healthcare worker-mediated patient-to-patient transmission, parent-mediated transmission to neonates, or spread from contaminated products or environment. If we cannot account for the mechanisms underlying transmission, a Poisson model may output misleading standard errors that underestimate the true variability and may impact inference [19, 22]. This phenomenon is referred to as overdispersion and can be overcome through the use of negative binomial or quasi-Poisson regression models [19].

Survival Models

Survival analysis is a powerful tool for analyzing the occurrence of HAIs or pathogen acquisition as well as the timing of these events [23]. Survival models can be particularly informative when risk of the outcome changes over time or is influenced by an exposure that changes over time. For hospital infection data, there is usually a time origin, often hospital or unit admission, that marks the beginning of the time a person is at risk for the outcome. In survival models, the time origin guides alignment and comparison of individuals with the same time at risk for the event of interest. We then can assess the likelihood, or hazard, of an event given survival to that point (i.e. being event-free and still under observation) [24].

Cox regression is a widely used survival model that makes no assumptions about the underlying hazard (which can vary over time) but assumes that the hazard ratio is constant over time. Latibeaudiere et al. [25**] assessed the hazard of *Acinetobacter baumannii* infection over the duration of admission among colonized versus non-colonized individuals. The findings estimated a 16 times higher hazard of infection in colonized individuals, and assumes that this effect was constant over the duration of hospitalization. Methods are available if this proportionality over time assumption is not met [26, 27]. For example, by accounting for non-proportional hazards, one study was able to report the time-varying effect of race on HAI risk. Specifically, they showed that non-Hispanic black patients compared to white patients had a higher hazard of HAI around the time of admission but this effect decreased with increasing length of stay [28*]. Parametric models are a useful alternative to Cox regression when the shape of the underlying hazard is of particular interest or when modeling time-varying effects [27].

By accounting for person-time, Poisson regression can approximate survival methods by estimating the average hazard. Typically, variables representing intervals of time will be added to the model, presuming a constant hazard in those intervals. The narrower the intervals, the more Poisson regression approximates continuous time survival approaches, such as Cox regression [21]. Both Poisson and Cox models have been used to report the time-dependent relationship between device dwell times and infections with consistent findings [29*–31*].

Time at risk needs to be carefully allocated in both survival analysis as well as Poisson regression. First, the time-varying nature of exposures is frequently overlooked in hospital-based analyses. Many hospital exposures, including mechanical ventilation, central line dwell time, and exposure to antibiotics vary over time. These variables are likely to be poorly characterized by their overall presence or absence and would be better defined by their duration, intensity, or particular timing in reference to outcome occurrence [32, 33**]. Note that time-varying exposures differ from the time-varying effects of exposures discussed above [26]. For example, survival analysis was used to assess the effect of time-varying antibiotic exposure on incident colonization with fluoroquinolone (FQ)-resistant *Escherichia coli* in a long-term care facility and found that the receipt of amoxicillin-clavulanate after colonization with FQ-susceptible *E. coli* was associated with increased risk of resistant colonization [34*]. This type of analysis can be accomplished with survival methods by constructing data in terms of person-periods, delineated by a change in exposure

status. Similarly, Brown et al. [4**] utilized a Poisson approach with a data structure in which each record represented one person-day under observation. Therefore, investigators were able to account for the time-varying nature of antibiotic treatment on *C.difficile* risk. With the application of the person-period based approaches, comes the need to account for the within-patient correlation in responses. This is typically accounted for during survival analysis set-up, but can also be addressed in Poisson regression through the use of generalized estimating equations [35], which can also be used to account for facility-level clustering in multi-center studies.

Another important consideration related to allocation of person-time is accounting for time deemed by the investigator as not at risk, sometimes called immortal person-time [36]. For example, if a 48-hour rule is utilized to distinguish community-acquired cases from hospital-acquired cases, this time should not be included as at-risk time in analyses of hospital-acquired events as a patient cannot have the event within this 48 hour period. The impact of immortal person-time has typically been studied in the context of chronic conditions [33**]. It has been shown that inclusion of person-time attributed to prevalent carriers of vancomycin-resistant enterococcus (VRE) notably inflated person-time denominators and, therefore, underestimated the incidence rate of VRE acquisition [37]. Additional research is needed to assess the impact of this issue in healthcare epidemiology.

A final issue to consider in time-to-event analyses is competing risks. A competing risk is an event that will preclude the observation of the outcome. In the context of hospital infection data, proposed competing events include hospital discharge and death as HAIs are often not captured after discharge and cannot occur after death [38**, 39]. This issue is of particular concern when the exposure is associated with the competing event [38**]. Consider a hypothetical study investigating the association between antibiotic exposure and risk of carbapenem-resistant Enterobacteriaceae (CRE) acquisition. Patients who receive antibiotics may have longer lengths of stay. If true, antibiotic exposure is associated with a competing event (discharge). Thus, simply censoring discharged individuals (who are less likely to be exposed to antibiotics) may result in an underestimation of the overall effect of antibiotic exposure on pathogen acquisition as it does not account for the fact that exposed individuals are more likely to have longer stays. Approaches are available to account for competing risks in survival analysis, including the estimation of the sub-distribution hazards based on the cumulative incidence function [40] or parametric mixture models [41]. A competing risk approach may give more accurate characterization of the risk that someone will experience after an exposure or intervention. Risk estimates that do not account for competing risks are not necessarily wrong, but do have a different meaning than those that do; hence, investigators should carefully consider whether a competing risk approach is appropriate.

Importantly, all of the approaches discussed above assume independence, that an individual's risk of pathogen acquisition in the hospital setting is independent of the colonization or infection status of others present on the unit. Due to the temporal and spatial relationships of healthcare-providers, patients, and the healthcare environment we know this assumption is often not met in the context of hospital infection data [42, 43]. One way to address this issue common to all regression-type models is to account for the unit-level

burden of colonization at a given time [44–46*]. Alternatively, mechanistic or mathematical models can be used as discussed below.

Mechanistic Models

Mechanistic models, which include models often termed ‘mathematical models’ and agent-based models, extend purely statistical approaches with some explicit representation of the disease process. These types of models are intended to describe the spread of nosocomial pathogens and inherently address the dependency issue discussed above by taking into account modes of transmission [47]. Mechanistic models have been used to describe pathogen transmission dynamics or to predict the effectiveness of a novel intervention given a range of scenarios [48**]. Models are also able to address limitations in available data, such as interval-censored observations of colonization status (i.e. weekly surveillance screening) or testing limitations (e.g. single anatomic site surveillance cultures compared to multiple anatomic sites).

A number of different types of mechanistic models have been applied to the study of HAIs. These models can be distinguished by whether they describe subgroups or individuals, the degree to which they are data-driven, and how they incorporate random variation in event occurrence (stochasticity) [49]. Compartmental models divide the study population into homogenous subgroups and assess the degree of movement between these compartments as a function of a series of differential equations. These equations reflect what is known about the process of transitioning between stages of colonization or infection. The aggregated impact of these transitions is estimated at the unit or population level. One such study modeled transitions from susceptible to colonized/infected states and found that environmental reservoirs may perpetuate transmission even when healthcare worker and patient reservoirs are no longer a factor, highlighting the need for thorough environmental disinfection [50].

Other approaches, such as agent-based models, explicitly represent the individual by accounting for heterogeneity in the model actors and characterizing the spatial aspects of transmission. Often these models simulate transmission for a set of parameters informed by available data or prior literature [51, 52*]. Highly data-driven approaches typically involve the estimation of unknown parameters (e.g. transmission rate) by fitting models directly to empirical data [49]. Finally, models can also be distinguished by the degree of stochasticity that is incorporated. In healthcare epidemiology, stochastic models are typically used to account for the role of chance in transmission processes within hospital units of relatively small sample size [49]. Markov process models form the basis of many stochastic approaches. A particular subtype, hidden Markov models, are particularly useful in the study of pathogen acquisition, as they can account for the largely unobserved process of transitioning from a non-colonized to colonized state [42, 53, 54].

Mechanistic approaches have been widely used in infectious disease epidemiology and are an ideal alternative to models that make assumptions that cannot hold under contagion-based processes. Careful crafting of models requires appropriate expertise in modeling methods and healthcare epidemiology. Collaboration is key to produce the most informed and

appropriately fit models of healthcare-associated transmission [47, 55]. Additional considerations for mechanistic approaches are discussed in Table 1.

Conclusion

There are numerous methodologies available for the analysis of epidemiologic cohort data. Poisson regression and survival analysis are able to exploit the highly granular nature of hospital-based data through the incorporation of time at risk, the ability to address time-varying exposures, and the flexibility to account for a hazard of infection or colonization that changes over time. Mechanistic models, though more complex in their development, offer benefits when compared to many out of box solutions in that they explicitly account for and model uncertainty in the transmission process. Although over-utilized in some contexts, logistic regression remains an important tool for risk factor analyses over short observation periods. The continued morbidity and mortality associated with HAIs and the threat posed by continually emerging, increasingly-antibiotic resistant organisms make uncovering etiology and intervening a priority. We hope that investigators will take full advantage of the data and statistical resources available to address these challenges.

Acknowledgments

Financial support: Research reported in this publication was supported by the National Institutes Of Allergy And Infectious Diseases of the National Institutes of Health (NIH) (R03AI117169) and the Agency for Healthcare Research Quality (AHRQ) (R01HS02287). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or AHRQ.

References

1. Singer, JD.; Willett, JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford Scholarship Online; 2003.
2. Carrieri M, Stolfi I, Moro M, Units I. Intercenter variability and time of onset: two crucial issues in the analysis of risk factors for nosocomial sepsis. The Pediatric Infectious Disease Journal. 2003; 22(7):599609.
3. Slimings C, Riley T. Antibiotics and hospital-acquired *Clostridium difficile* infection: update of systematic review and meta-analysis. J Antimicrob Chemother. 2014; 69(4):881–891. [PubMed: 24324224] Investigators performed a systematic review and metanalysis of observational studies investigating the effect of antibiotic exposure on risk for *C. difficile* infection. The majority were risk factor analyses using a case-control study approach, highlighting the need for longitudinal analyses to further elucidate the patterns and timing of exposures important for the development of *C. difficile* infection.
4. Brown K, Fisman D, Moineddin R, Daneman N. The Magnitude and Duration of *Clostridium difficile* Infection Risk Associated with Antibiotic Therapy: A Hospital Cohort Study. PLoS ONE. 2014; 9(8) Brown et al. assessed the pattern of risk for *C. difficile* infection during after after antimicrobial exposure through the use of weighted Poisson regression models. The time-varying nature of antibiotic exposure was considered as well as it's cumulative impact. Authors addressed dependency of outcomes through the inclusion of a measure of *C.difficile* burden. Through the use of this overall approach, which builds upon current survival methodologies, investigators were able to target specific times of elevated risk of *C. difficile* infection during and post-antibiotic exposure that could serve to inform intervention. In addition, the methodology used in this investigation could be applied to the study of many of time-varying exposures in the healthcare setting.
5. Hornik, CP.; Fort, P.; Clark, RH.; Watt, K.; Benjamin, DK.; Smith, BP., et al. Early human development [Internet]. Vol. 88. Elsevier; 2012. Early and late onset sepsis in very-low-birth-weight

infants from a large group of neonatal intensive care units; p. S69-S74. Available from: <http://www.sciencedirect.com/science/article/pii/S0378378212700191>.

6. Giuffrè M, Amodio E, Bonura C, Geraci D, Saporito L, Ortolano R, et al. Methicillin-resistant *Staphylococcus aureus* nasal colonization in a level III neonatal intensive care unit: Incidence and risk factors. *American Journal of Infection Control*. 2015
7. Chopra V, Ratz D, Kuhn L, Lopus T, Chenoweth C, Krein S. PICC-associated bloodstream infections: prevalence, patterns, and predictors. *Am J Med*. 2014; 127(4):319–328. [PubMed: 24440542]
8. Jeon C, Neidell M, Jia H, Sinisi M, Larson E. On the Role of Length of Stay in Healthcare-Associated Bloodstream Infection. *Infection Control and Hospital Epidemiology*. 2012; 33(12): 12131218.
9. Irala-Estévez J, Martínez-Concha D, Díaz-Molina C, Masa-Calles J, Castillo A, Navajas R. Comparison of different methodological approaches to identify risk factors of nosocomial infection in intensive care units. *Intensive Care Med*. 2001; 27(8):1254–1262. [PubMed: 11511936]
10. Kleef E, Green N, Goldenberg S, Robotham J, Cookson B, Jit M, et al. Excess length of stay and mortality due to *Clostridium difficile* infection: a multi-state modelling approach. *Journal of Hospital Infection*. 2014; 88(4):213217.
11. Leistner R, Hirsemann E, Bloch A, Gastmeier P, Geffers C. Costs and prolonged length of stay of central venous catheter-associated bloodstream infections (CVC BSI): a matched prospective cohort study. *Infection*. 2013; 42(1):3136.
12. Lye D, Earnest A, Ling M, Lee T-E, Yong H-C, Fisher D, et al. The impact of multidrug resistance in healthcare-associated and nosocomial Gram-negative bacteraemia on mortality and length of stay: cohort study. *Clinical Microbiology and Infection*. 2012; 18(5):502–508. [PubMed: 21851482]
13. Cole S, Platt R, Schisterman E, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010; 39(2):417–420. [PubMed: 19926667]
14. Hernán M, Hernández-Díaz S, Robins J. A Structural Approach to Selection Bias. *Epidemiology*. 2004; 15(5):615. [PubMed: 15308962]
15. Buffet-Bataillon S, Saunders L, Campillo-Gimenez B, Haegelen C. Risk factors for neurosurgical site infection after neurosurgery in Rennes, France: Comparison of logistic and Cox models. *American Journal of Infection Control*. 2013; 41(12):12901292. Authors compare the use of Cox regression to logistic regression for the analysis of risk factors for surgical site infection. Although the models target different questions (logistic regression is assessing whether an event occurred during some observation period and Cox regression, in contrast, assesses time until an event occurred), results are similar using the two methods. This may be due to a relatively short observation period and assumed complete follow up on individuals under observation. Cox regression is recommended in situations where longitudinal data is available and events are common.
16. Green M, Symons M. A comparison of the logistic risk function and the proportional hazards model in prospective epidemiologic studies. *J Chronic Dis*. 1983; 36(10):715–723. [PubMed: 6630407]
17. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology*. 2004; 159(7):702–706. [PubMed: 15033648]
18. Yelland L, Salter A, Ryan P. Performance of the Modified Poisson Regression Approach for Estimating Relative Risks From Clustered Prospective Data. *Am J Epidemiol*. 2011; 174(8):984–992. [PubMed: 21841157]
19. Cox S, West S, Aiken L. The analysis of count data: a gentle introduction to poisson regression and its alternatives. *J Pers Assess*. 2009; 91(2):121–136. [PubMed: 19205933]
20. Iwamoto M, Mu Y, Lynfield R, Bulens SN, Nadle J, Aragon D, et al. Trends in invasive methicillin-resistant *Staphylococcus aureus* infections. *Pediatrics*. 2013 Oct 2; 132(4):e817–e824. [PubMed: 24062373] This study informs trends in invasive MRSA infection through the use of Poisson regression applied to aggregated data and demonstrates lack of a downward trend among children, in contrast to that found in adults in previous studies. This highlights one of the benefits of Poisson approaches, that it can be applied when only aggregate data is available.

21. Callas P, Pastides H, Hosmer D. Empirical comparisons of proportional hazards, poisson, and logistic regression modeling of occupational cohort data. *Am J Ind Med.* wiley. 1998; 33(1):33–47.
22. Gardner W, Mulvey E, Shaw E. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull.* 1995; 118(3):392. [PubMed: 7501743]
23. Aiken A, Mturi N, Njuguna P, Mohammed S, Berkley J, Mwangi I, et al. Risk and causes of paediatric hospital-acquired bacteraemia in Kilifi District Hospital, Kenya: a prospective cohort study. *Lancet.* 2011; 378(9808):2021. [PubMed: 22133536]
24. Cox, DR.; Oakes, D. *Analysis of Survival Data.* Chapman & Hall; 1984.
25. Latibeaudiere R, Rosa R, Laowansiri P, Arheart K, Namias N, Munoz-Price L. Surveillance Cultures Growing Carbapenem-Resistant *Acinetobacter baumannii* Predict the Development of Clinical Infections: A Retrospective Cohort Study. *Clinical Infectious Diseases.* 2014; 60(3): 415422. Investigators assessed time to *Acinetobacter baumannii* infection according to colonization status. Multivariable Cox regression models were used. Variation in risk based on the timing of surveillance culture was considered. *Acinetobacter baumannii* colonization was identified as an important risk factor for subsequent infection, and, importantly, through the use of survival approaches, authors were able to report a notable difference in the proportion of patients that were *A. baumannii* infection-free at 30 days of admission comparing those colonized with non-colonized individuals.
26. Bellera C, MacGrogan G, Debled M, Lara C, Brouste V, Mathoulin-Pélissier S. Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol.* 2010; 10(1):20. [PubMed: 2023435]
27. Cox C, Chu H, Schneider M, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statist Med.* wiley. 2007; 26(23):4352–4374.
28. Jeon C, Muennig P, Furuya E, Cohen B, Nash D, Larson E. Burden of present-on-admission infections and health care-associated infections, by race and ethnicity. *American Journal of Infection Control.* 2014; 42(12):12961302. Investigators demonstrated the time-varying effect of race on HAI risk through the use interaction terms (race by time) within the context of Cox regression. They were able to demonstrate that race may be an important risk factor for HAI early in admission, but the risk imposed by race dissipates over time, perhaps implicating other mechanisms of HAI risk at later time points.
29. Milstone AM, Reich NG, Advani S, Yuan G, Bryant K, Coffin SE, et al. Catheter dwell time and CLABSI in neonates with PICCs: a multicenter cohort study. *Pediatrics.* 2013 Dec; 132(6):e1609–e1615. [PubMed: 24218474] In a multi-center study Milstone et al demonstrate the flexibility of the Poisson model to assess time-dependent effects of PICC dwell time on infectious risk. Investigators used a cubic spline approach to model fit and reported a period of elevated risk from 2 weeks post-insertion until removal.
30. Barrier A, Williams D, Connelly M, Creech C. Frequency of Peripherally Inserted Central Catheter Complications in Children. *Pediatr Infect Dis J.* 2012; 31(5):519. [PubMed: 22189533]
31. Jumani K, Advani S, Reich N, Gosey L, Milstone A. Risk Factors for Peripherally Inserted Central Venous Catheter Complications in Children. *JAMA Pediatr.* 2013; 167(5):429–435. [PubMed: 23549677] Investigators assessed the role of PICC dwell time on venous catheter complications, both infectious and non-infectious. Poisson regression was used with catheter dwell time modeled with cubic spline terms to address the nonlinear association with complications over time, highlighting the application of Poisson regression to time-varying hazards with informative visual depictions.
32. Shintani A, Girard T, Eden S, Arbogast P, Moons K, Ely E. Immortal time bias in critical care research: Application of time-varying Cox regression for observational cohort studies*. *Crit Care Med.* 2009; 37(11):2939. [PubMed: 19770737]
33. Hanley J, Foster B. Avoiding blunders involving “immortal time”. *International Journal of Epidemiology.* 2014; 43(3):949961. The concept of immortal person time is discussed. The mishandling of immortal time in survival analyses is a common threat to validity in a number of fields. The error comes in two forms: the misallocation of person-time when exposure status changes over time and the inclusion of an 'eligibility period' or person-time in which participants cannot, by definition, have the event of interest to be included in the study. This has implications

for both time-varying covariates as well as periods where patients are deemed not at risk for healthcare associated infections (e.g. 48 hour rule).

34. Han J, Maslow J, Han X, Xie S, Tolomeo P, Santana E, et al. Risk factors for the development of gastrointestinal colonization with fluoroquinolone-resistant *Escherichia coli* in residents of long-term care facilities. *J Infect Dis.* 2014; 209(3):420–425. [PubMed: 23986544] Han et al followed a cohort of long term care patients, all of whom were colonized with fluoroquinolone-susceptible *Escherichia coli* and assessed the role of antibiotic exposure on the development of fluoroquinolone-resistant *Escherichia coli*. Authors used Cox regression, accounting for the time-varying nature of exposures and found receipt of amoxicillin-clavulanate to be a risk factor for resistant colonization.
35. Hanley J, Negassa A, Edwardes M, Forrester J. Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. *Am J Epidemiol.* 2003; 157(4):364–375. [PubMed: 12578807]
36. Suissa S. Immortal Time Bias in Pharmacoepidemiology. *American Journal of Epidemiology.* 2008; 167(4):492–499. [PubMed: 18056625]
37. Huang S, Rifas-Shiman S, Pottinger J, Herwaldt L, Zembower T, Noskin G, et al. Improving the assessment of vancomycin-resistant enterococci by routine screening. *J Infect Dis.* 2007; 195(3): 339–346. [PubMed: 17205471]
38. Wolkewitz M, Cooper BS, Bonten MJ, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. *BMJ.* 2014 Jan 3;349:g5060. [PubMed: 25146097] A review and study of competing risks is presented in the context of hospital epidemiology. Competing risks are those that cause censoring and alter the chance of having the event of interest. Potential completing risks for studies with an outcome of HAI include death and discharge. Authors provide real data and scenario examples to demonstrate the effect of competing risks on interpretation of risk estimates. Competing risks have the most notable impact when the exposure is associated with the competing event.
39. Schumacher M, Allignol A, Beyersmann J, Binder N, Wolkewitz M. Hospital-acquired infections--appropriate statistical treatment is urgently needed! *Int J Epidemiol.* 2013 Oct 2; 42(5):1502–1508. [PubMed: 24038717]
40. Lau B, Cole S, Gange S. Competing Risk Regression Models for Epidemiologic Data. *Am J Epidemiol.* 2009; 170(2):244–256. [PubMed: 19494242]
41. Checkley W, Brower R, Muñoz A. Investigators for. Inference for Mutually Exclusive Competing Events Through a Mixture of Generalized Gamma Distributions. *Epidemiology.* 2010; 21(4):557. [PubMed: 20502337]
42. Cooper B, Lipsitch M. The analysis of hospital infection data using hidden Markov models. *Biostatistics.* 2004 Apr 4; 5(2):223–237. [PubMed: 15054027]
43. Halloran ME, Struchiner CJ. Study Designs for Dependent Happenings. *Epidemiology.* 2(5):331–338. [PubMed: 1742381]
44. Ajao AO, Harris AD, Roghmann M-CC, Johnson JK, Zhan M, McGregor JC, et al. Systematic review of measurement and adjustment for colonization pressure in studies of methicillin-resistant *Staphylococcus aureus* vancomycin-resistant enterococci, and *Clostridium difficile* acquisition. *Infect Control Hosp Epidemiol.* 2011 May; 32(5):481–489. [PubMed: 21515979]
45. Popoola VO, Carroll KC, Ross T, Reich NG, Perl TM, Milstone AM. Impact of colonization pressure and strain type on methicillin-resistant *Staphylococcus aureus* transmission in children. *Clin Infect Dis.* 2013 Nov 5; 57(10):1458–1460. [PubMed: 23943821]
46. Popoola VO, Milstone AM. Decolonization to prevent *Staphylococcus aureus* transmission and infections in the neonatal intensive care unit. *J Perinatol.* 2014 Nov 6; 34(11):805–810. [PubMed: 25010222] Popoola et al review the concept of colonization pressure in a larger discussion of the motivations behind decolonization treatment for MRSA prevention. Colonization pressure remains an important mechanism of controlling for dependency of outcomes in regression-type analyses.
47. Grundmann H, Hellriegel B. Mathematical modelling: a tool for hospital infection control. *The Lancet Infectious Diseases.* 2006; 6(1):3945.
48. Doan TN, Kong DC, Kirkpatrick CM, McBryde ES. Optimizing hospital infection control: the role of mathematical modeling. *Infect Control Hosp Epidemiol.* 2014 Dec; 35(12):1521–1530. [PubMed: 25419775] Authors review the effectiveness of infection control measures in the context

of mathematical modeling approaches. In particular, authors focused on compartmental models. A common approach to compartmental models in the study of healthcare-associated infections is to use a vector-host model, where a vector, typically healthcare workers, facilitate the transition from an uncolonized to a colonized state. Therefore, the effect of infection control measures can be assessed by varying model parameters, such as healthcare worker decolonization expected from a hand hygiene initiative.

49. van Kleef E, Robotham J, Jit M, Deeny S, Edmunds W. Modelling the transmission of healthcare associated infections: a systematic review. *BMC Infectious Diseases*. 2013; 13(1):294. [PubMed: 23809195]
50. Wang X, Xiao Y, Wang J, Lu X. A mathematical model of effects of environmental contamination and presence of volunteers on hospital infections in China. *Journal of Theoretical Biology*. 2012; 293:161173.
51. Temime L, Opatowski L, Pannet Y, Brun-Buisson C, Boëlle P, Guillemot D. Periparturient health-care workers as potential superspreaders. *Proc Natl Acad Sci USA*. 2009; 106(43):18420–18425. [PubMed: 19841263]
52. Barnes S, Morgan D, Harris A, Carling P, Thom K. Preventing the Transmission of Multidrug-Resistant Organisms: Modeling the Relative Importance of Hand Hygiene and Environmental Cleaning Interventions. *Infection Control and Hospital Epidemiology*. 2014; 35(9):11561162. Barnes et al use an agent-based model to assess the impact of hand hygiene and environmental disinfection on multi-drug resistant organism transmission. Relative to environmental disinfection, improvements in hand hygiene compliance were associated with higher reductions in pathogen-acquisition.
53. Khader K, Leecaster M, Greene T, Samore M, Thomas A. Improved hidden Markov model for nosocomial infections. *Mathematical Medicine and Biology*. 2013; 31(4):338352.
54. Cooper BS, Medley GF, Bradley SJ, Scott GM. An augmented data method for the analysis of nosocomial infection data. *Am J Epidemiol*. 2008 Sep 1; 168(5):548–557. [PubMed: 18635575]
55. Forrester B, Pettitt A. Use of Stochastic Epidemic Modeling to Quantify Transmission Rates of Colonization With Methicillin-Resistant *Staphylococcus aureus* in an Intensive Care Unit. *Infect Control Hosp Epidemiol*. 2010
56. Nemes S, Jonasson J, Genell A, Steineck G. Bias in odds ratios by logistic regression modelling and sample size. *BMC Med Res Methodol*. 2009; 9(1):56. [PubMed: 19635144]
57. Babyak MA. What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosomatic Medicine*. 2004; 66:411–421. [PubMed: 15184705]

Key Findings

- Cohort studies represent a large portion of investigations of healthcare-associated infectious outcomes. Despite the range of statistical methods available for longitudinal studies, logistic regression remains the dominant approach.
- Poisson models and survival analysis provide the opportunity to assess variation in time to the event of interest as well as address data complexities, such as time-varying exposures, time-varying hazards, and competing risks.
- Mechanistic models explicitly characterize the infection process, are able to account for the uncertainty in our measurements and observations, and can be flexibly modified to assess the impact of interventions or exposures for strategic purposes.
- Healthcare-associated infections remain a major threat to patient safety. Statistical methodologies are available to assess etiology of infection or colonization, inform transmission dynamics, and determine the impact of novel interventions.

Figure 1.

Summary of study designs for healthcare-associated infectious outcomes, which includes healthcare-associated infections (central line-associated blood stream infection, catheter-associated urinary tract infection, ventilator-associated pneumonia, and surgical site infection) and pathogen acquisition. Analytic approaches to cohort studies with healthcare-associated infectious outcomes are presented.

Abbreviations: ITS, interrupted time series; AJIC, *American Journal of Infection Control*; ICHE, *Infection Control and Healthcare Epidemiology*

Table 1**Approaches to longitudinal analysis of healthcare-associated infectious outcomes**

Model	Strengths	Considerations	Solutions
Logistic Regression	Easily performed in statistical packages Highly interpretable output	Does not allow for assessment of variation in time to event Exposure typically classified as time-fixed Does not consider censoring or varying contributions of person-time Model overfitting should be avoided in the context of small sample sizes or sparse events ^a	Consider Poisson regression, survival approaches, or mechanistic models
Poisson Regression	Easily performed in statistical packages Can be used for binary or count, aggregated or individual data Allows for inclusion of person-time for calculation of incidence rates Can approximate survival approaches Does not require clear time origin as is necessary for survival methods Can be modeled to incorporate time-varying effects of exposure (non-proportional hazard) Can incorporate time-varying nature of exposures	Has stringent distributional assumptions about variability of count data that may not hold Overdispersed data will result in inaccurate standard errors, which can impact inference Model overfitting should be avoided in the context of small sample sizes or sparse events ^a	Test for over/under dispersion Consider negative binomial models
Survival Methods	Easily performed in statistical packages Allows for assessment of time to event Can be modeled to incorporate time-varying effect of exposure (non-proportional hazard) Can account for competing risks Can incorporate time-varying nature of exposures	Requires indexing from a time origin when a meaningful time origin may not be present Cannot be used with non-individual, aggregated data Requires consideration of immortal person time Model overfitting should be avoided in the context of small sample sizes or sparse events ^a	Consider Poisson models that do not require a time origin and can accommodate grouped data If competing risks are of concern, subdistribution hazards approach or mixture models can be used Immortal person time should not be included as time at-risk
Mechanistic Models	Can more flexibly characterize the infection process, not restricted by model assumptions Can fully account for the dependencies between individuals in outcome status Can incorporate measurement uncertainty Can address interval-based measurements of outcome status (i.e. weekly surveillance culture)	Need to consider interaction between actors in model ^b Requires background epidemiologic analysis to understand mechanisms of pathogen acquisition ^b [48] Though tools are available for model implementation, computing skills required ^b Model predictions should be validated ^b [49] Model overfitting should be avoided in the context of small sample sizes or sparse events ^a	Consider and account for facility-specific factors that mediate interactions among patients, staff, and visitors (i.e. is spatial or provider cohorting used?) ^b [42] Prioritize model components & perform epidemiologic studies to inform and validate models ^b [48,49] Consider collaboration with colleagues experienced in the application of mathematical models

Superscript denotes that information was not addressed in text.

^a All statistical approaches are susceptible to overfitting due to the addition of too many variables, often despite a small sample size or a small number of events. Overfitting can lead to spurious findings that cannot be replicated [56,57].

^b Considerations and solutions for mechanistic models.