JAC Antimicrob Resist
doi:10.1093/jacamr/dlab016

JAC-Antimicrobial Resistance

Could simulation methods solve the curse of sparse data within clinical studies of antibiotic resistance?

James C. Hurley [1,2]* and David Brownridge[3]

[1]*Department of Rural Health, Melbourne Medical School, University of Melbourne, Australia;* [2]*Division of Internal Medicine, Ballarat Health Services, Ballarat, Victoria, Australia;* [3]*Pharmacy, Ballarat Health Services, Ballarat, Victoria, Australia*

*Corresponding author. E-mail jamesh@bhs.org.au or hurleyjc@unimelb.edu.au

Infectious disease (ID) physicians and ID pharmacists commonly confront therapeutic questions relating to antibiotic resistance. Randomized controlled trial data are few and meta-analytic-based approaches to develop the evidence-base from several small studies that might relate to an antibiotic resistance question are not simple. The overriding challenge is the sparsity of data which is problematic for traditional frequentist methods, being the paradigm underlying the derivation of '*P* value' inferential statistics. In other sparse data contexts, simulation methods enable answers to key questions that are meaningful, quantitative and potentially relevant. How these simulation methods 'work' and how Bayesian-based methods, being not '*P* value based', can facilitate simulation are reviewed. These methods are becoming increasingly accessible. This review highlights why sparse data is less of an issue within Bayesian versus frequentist paradigms. A fictional pharmacokinetic study with sparse data illustrates a simplistic application of Bayesian and simulation methods to antibiotic dosing. Whether within epidemiological projections or clinical studies, simulation methods are likely to play an increasing role in antimicrobial resistance research within both hospital and community studies of either rare infectious disease or infections within specific population groups.

'You know the greatest danger facing us is ourselves, an irrational fear of the unknown. But there's no such thing as the unknown — only things temporarily hidden, temporarily not understood'. Starfleet Captain, James Tiberius Kirk [Star Trek; Season One, Episode 10; 'The Corbomite Manoeuvre']

Introduction

Antibiotic-resistant bacteria are responsible for significant morbidity and mortality, as well as escalating economic costs.[1] Early and appropriate antimicrobial therapy, one of the key elements of managing patients with known or suspected infection, becomes an increasing and evolving challenge in the age of antimicrobial resistance. With decreased susceptibility to current antimicrobials, and a lack of new agents with novel mechanisms of action, infectious diseases (ID) physicians and ID pharmacists face the increasing challenge of selecting optimal and effective regimens for patients with resistant infections using limited data. The purpose of this review is to broadly outline Bayesian and simulation methods that might help meet these data challenges together with a tutorial demonstration of their application. Also, several examples of the widening research and clinical applications of these methods towards research relating to antibiotic resistance generally are provided.

RCT data, a luxury?

The limited availability of robust randomized controlled trial (RCT) data to guide treatment decisions has long been problematic in many areas of infectious diseases, as with medicine generally. Moreover, RCTs that address therapeutic aspects relating to antibiotic resistance, are rare. Two examples of recently published RCTs, the ARREST trial[2] and the MERINO study,[3] help to illustrate the rare and exceptional nature of these types of studies.

The ARREST trial[2] addressed the question of whether adjunctive rifampicin would provide additional benefit over standard antibiotic therapy for adults with *Staphylococcus aureus* bacteraemia. This trial recruited patients from 26 UK hospital groups. It is notable that prior to this study, published in 2018, the RCT evidence base in relation to the optimal management of *S. aureus* bacteraemia was based on fewer than 1500 patients recruited to 16 controlled trials of antimicrobial therapy.[4] MRSA bacteraemias were not excluded from the ARREST trial but, with MRSA bacteraemias constituting only 47 of the 758 (6%) bacteraemias, any inference for optimal management of MRSA bacteraemias from this study is limited.

The MERINO study[3] was an RCT of piperacillin/tazobactam versus meropenem to evaluate the use of piperacillin/tazobactam as carbapenem-sparing therapy in patients with bloodstream infections (BSI) caused by ceftriaxone-resistant *Escherichia coli* or *Klebsiella pneumoniae*. This RCT recruited hospitalized patients

from 26 sites in nine countries. Of note, the prior evidence was somewhat inconclusive as a propensity matched comparison of 150 patients with ESBL-producing *E. coli* and *K. pneumoniae* BSI treated empirically with piperacillin/tazobactam versus a carbapenem in a multicentre study undertaken in Singapore found similar 30 day mortality among the two groups.[5]

The findings of these two RCTs are valuable and each progressed the evidence base from what had been available before. The ARREST RCT found that adjunctive rifampicin provided no overall benefit over standard antibiotic therapy in adults with *S. aureus* BSI. The MERINO RCT findings did not support piperacillin/tazobactam compared with meropenem for these patients with BSI caused by ceftriaxone-resistant *E. coli* or *K. pneumoniae*. Of note, neither *S. aureus* bacteraemia nor ESBL BSI are rare infections and both studies recruited across a wide network. However, despite this, patient recruitment to both studies was slow, taking 4 years in each case.

These examples illustrate the difficulties in obtaining RCT data even for common infections among broadly selected patients prospectively recruited across multiple sites. In many cases, the evidence for antibiotic-resistant infections is based on either subgroups of prospective studies or retrospective case series from single centres. Obtaining data relating to the treatment of multidrug-resistant infections, niche patient populations, such as those with altered renal function, critical illness and patients outside usual adult age and weight intervals, will be especially challenging.

Meta-analysis

Given the difficulty and cost in mounting large robust RCTs, what other approaches are there? To some extent, the application of various methods of meta-analysis enable the results from multiple smaller RCTs that have examined similar topics to be summed to provide a broader evidence base.[6,7] There are several challenges here, the main one being whether a sufficient number of truly comparable quality RCTs can be identified. The minimum number of studies required is debatable but generally at least ten studies are required to adequately appraise the amount of heterogeneity among the study results. Generating a summary with fewer than ten studies carries the risk of conveying a false sense of homogeneity.[7] A good meta-analysis will estimate the heterogeneity associated with the results and acknowledge that heterogeneity is the reality of the world outside of RCTs. If patient level data are available, it might be possible to undertake a patient level meta-regression. When there are at least ten studies available, it becomes possible to estimate the contribution of various study-level factors to variation in the study results using meta-regression techniques.

Network meta-analysis (NMA) is emerging as a technique to enable indirect comparisons of multiple interventions that have not been studied head-to-head within any one RCT. A key assumption of NMA is the transitivity between trials. Transitivity is generally an untestable assumption. There are other challenges. For example, a recent NMA to evaluate the comparative effectiveness of fifteen antimicrobial treatments used for drug-resistant *Acinetobacter baumannii* pneumonia in 2118 critically ill patients within 23 studies illustrates some of the challenges.[8] Only four of the included studies were RCTs. The accuracy and classification of drug-resistant *A. baumannii* pneumonia likely varied between the studies. The safety profiles of the treatments were not evaluated due to insufficient data. Novel therapies and combination therapies could not be evaluated due to the lack of opportunity to add these to the network.[8]

The above difficulties in undertaking meta-analyses of any type, need to be appreciated. Moreover, in addition to these limitations, for studies of antibiotic resistance, the subgroups of patients with resistant infections within larger RCTs may or may not have been reported. Even where they have been reported, there may be unrecognized confounding from other risk factors between patient populations with susceptible versus resistant infections. For example, the relative attributable mortality associated with MRSA bacteraemia versus MSSA bacteraemia is controversial. An early meta-analysis of 31 published studies reported a significant increase in hospital mortality associated with MRSA bacteraemia (OR, 1.93; 95% CI, 1.54–2.42) although the results were associated with significant heterogeneity.[9–13] However, the length of hospital stay (LOS) likely differs for patients with the different bacteraemia types and this might be an unrecognized confounder (Figure 1) that may be difficult to adequately control for. Likewise, the LOS may also confound the relationship between the mortality risk associated with vancomycin-resistant enterococcal bacteraemia versus vancomycin-susceptible enterococcal bacteraemia.[14] A multi-state study of a Scottish hospital cohort of patients acquiring

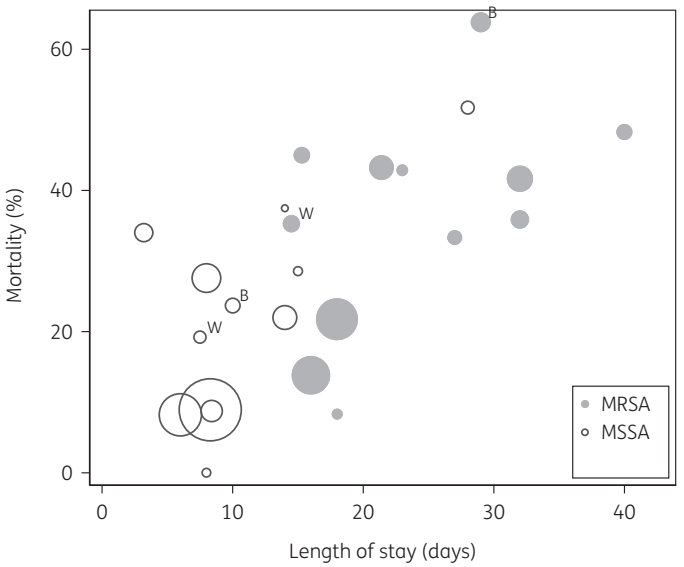

Figure 1. The relationship between mortality proportion among patients with *S. aureus* bacteraemia versus length of hospital stay (LOS) among groups of patients with either methicillin-resistant *S aureus* (MRSA; closed symbols) or methicillin-susceptible *S. aureus* (MSSA; open symbols) bacteraemia, as reported in 12 studies for which data on LOS was available (Table S1). The size of each group is represented by the corresponding size of the symbols (<30, 31–50, 51–109, 110–299, and >300 patients, respectively, from smallest to largest;). Data for recent studies by Blot *et al.*[12] [B] and Wolkewitz *et al.*[13] [W] are indicated. This Figure is modified from an original in Hurley,[11] with permission from the publisher.

S. aureus bacteraemia whilst in hospital found that MRSA was acquired approximately 7 days later than MSSA bacteraemia. As a consequence, the increased mortality associated with MRSA versus MSSA bacteraemia became non-significant after accounting for differences in temporal exposures and competing risks.[13]

The curse of sparse data

The above examples illustrate some traditional approaches to addressing key therapeutic questions where there is limited study data. However, in relation to therapeutic questions specific to antibiotic resistance, sufficient data simply may not be available. Sparse data is a limiting issue in many areas, not just in relation to therapeutic questions specific to antibiotic resistance. Where sparse data is limiting, simulation techniques together with Bayesian inference may be used to fill the gap. Understanding these methods is not simple for most ID practitioners. As a starting point, the two prevailing inferential data analysis paradigms, the Frequentist and the Bayesian, need to be described and contrasted (Table 1).[15–17]

The frequentist paradigm

Most ID practitioners will be familiar with the traditional frequentist paradigm which underlies the methodology for deriving *P* values as commonly used in RCTs and also meta-analyses (Table 1). The methodology originated in the early 20th century when all statistical analysis was undertaken 'long hand' and was geared towards null hypothesis significance testing to support binary inference decisions. That is, the null hypothesis was either accepted or rejected. Improving computing facilities in the second half of the 20th century facilitated the calculation of confidence intervals (CI) using the same analytic methods on pocket calculators and, more recently, personal computers. Frequentist methods have become established as the traditional analytic approach, and they will remain in common use, with software for their calculation being widely available.

Frequentist methods conceptualize the uncertainty in the data sample in the context of a long-run view of repeated imaginary samples. The fundamental limitation within frequentist methods is that they cannot provide the probability of any specific result despite the *P* value and the 95% CI being commonly misinterpreted in this way. The correct interpretation of a 95% CI is that if the same experiment were repeated many times with confidence intervals computed for each experiment, then 95% of those intervals would be expected to contain the 'true' mean. Note that this relates to the location and imprecision in where the true mean MIGHT be located. Moreover, this expectation relates to the mean value, not to the location of individual observations of potential interest.

A further limitation of frequentist methods is that a *P* value which is judged 'significant' for the primary finding of any study gives no indication of the certainty of the result. Many studies may provide findings that, while statistically significant, may not be robust due to the small numbers of patients studied. This is an issue for small randomized studies but more so for non-randomized studies due to possible imbalances in unrecognized confounders. This uncertainty can be quantified by a fragility index, which is defined as the minimum number of patients whose status if changed from a non-event (not experiencing the primary end point) to an event (experiencing the primary end point) would render the study results statistically insignificant.[18] A smaller fragility index indicates a less robust RCT result. For example, a recent study of oral vancomycin prophylaxis (OVP) for patients requiring systemic antibiotic therapy randomly assigned 50 patients to OVP versus 50 to no prophylaxis. The findings that zero of 50 patients (0%) receiving OVP versus 6 of 50 (12%) developed healthcare facility–onset *Clostridioides difficile* infection led to a conclusion that OVP appeared to be protective ($P = 0.03$, Fisher's exact test).[19] This result has a fragility index of only one because if one of the fifty

Table 1. Comparison of frequentist and Bayesian methods

Key elements	Frequentist	Bayesian
Underlying assumptions		
Observed data is:	From an imaginary repeatable random sample	A 'singleton'
Model parameters are:	Unknown but fixed across imaginary repeat samples	Random conditional on prior and observed data
Estimate of interest	Point estimate of parameter	Distribution of parameter
Estimation method	Various, application specific	Bayes' rule
Input	Limited to data at hand	Incorporates data at hand with prior knowledge and beliefs
Typical output	Estimate of a mean (or median) value	Estimate is a posterior distribution
Representation of imprecision in estimate of interest	Confidence interval	Credible interval
Sparse data	Problematic for parametric methods	Generally, not as problematic
Computation	Iterative or non-iterative closed-form equations	More complex and simulation usually required to 'bypass' intractable integrals
Validation	Assumption checking required	Model checking required
Inference	Based on presumption of infinite imaginary resampling from the same data model	Based on posterior distribution derived from the prior knowledge and beliefs with data at hand

patients (2%) receiving OVP had developed healthcare facility-onset *C. difficile* infection, the *P* value would now become no longer significant ($P = 0.11$, Fisher's exact test) on a recalculation. By contrast, with several hundred patients in each of the ARREST[2] and the MERINO[3] studies, these were powered for the derivation of robust findings using frequentist methods.

The Bayesian paradigm

Bayesian methods solve several obstacles problematic for frequentist methods arising from sparse and fragile data by augmenting the observed data with prior knowledge. ID practitioners may be less familiar with the Bayesian than the frequentist paradigm. However, Bayesian logic is largely intuitive and resembles how ID physicians make diagnostic inferences.[16] For example, in formulating a clinical diagnosis of possible measles for a patient presenting with a fever and a rash, there are two approaches. A physician with a frequentist approach would consider only the patient's clinical features as they might correspond to descriptions of the various potential candidate diagnoses as catalogued in a medical textbook. Using only this information, an ID physician using a frequentist perpsective would make a confident diagnostic inference as to whether the diagnosis was measles or was not measles as follows. For the next hypothetical 100 patients with a fever and rash syndrome identical to the presenting patient, the diagnosis made by the frequentist ID physician would apply to 95 and would not apply to 5. However, the frequentist ID physician would not be able to state to which 95 among this hypothetical century the diagnosis was applicable nor whether the presenting patient in question was one of the 95 to which the diagnosis was applicable or one of the 5 to which it was not applicable. Moreover, this prediction would rest on an assumption that the most appropriate textbook had been consulted.[20] By contrast, an ID physician using a Bayesian perspective would considerably modify the probabilities of the candidate diagnoses with knowledge that a measles outbreak was current in the community. The knowledge of a measles outbreak is key information but is external to the patient. The Bayesian ID physician will integrate both the patient-specific and the external information in an intuitive way creating a presumptive diagnosis of measles based on a subconscious probability estimate.

In the measles diagnosis analogy, the Bayesian ID physician's sub-conscious probability estimate has been shaped by a subliminal calculation. This calculus underlies Bayesian analysis and is formalized as the 'Bayes factor' or 'Bayes rule'.[17] The Bayes factor in this example is ratio of the diagnostic probability of measles with versus without the knowledge that there is a measles outbreak in the community. For example, the Bayesian ID physician might (subliminally) estimate the probability of measles for the specific individual patient in question, as being between say 2% to 5% with this estimate increasing to a likelihood of 90% to 95% in the presence of a known measles outbreak in the community. These estimates, with and without knowing that an outbreak is extant could be termed in Bayesian parlance, 'The prior' and 'The posterior', respectively. Of note, in the Bayesian paradigm, the prior and posterior estimates are each usually range estimates rather than point estimates of probability. Estimates informed by greater amounts of prior data will be narrower (more precise) than vague estimates based merely on a 'hunch'.

The Bayesian paradigm, and the 'Captain Kirk' style of thinking, has been noted at the intersection of critical care and ID patient management within the context of modelling the response to novel therapies for patients with sepsis.[21] This intersection is also where the difficult therapeutic questions relating to antimicrobial resistance are commonly encountered.[21–24]

In the Bayesian paradigm, the observed data is taken to be a 'singleton' and stands in contrast to the frequentist paradigm which assumes that the data arises as part of repeated imaginary sampling. This data 'singleton' gains inferential meaning when placed within the context of prior knowledge. In this way, Bayesian techniques conceptualize any uncertainty as incomplete knowledge, somewhat as implied by the remark from Star Trek's Captain Kirk, as quoted above. By contrast, frequentist methods struggle with sparse data because there is no recourse to external data.

Limitations

Bayesian statistics, which rely on a user-supplied prior, have been criticized for lacking the objectivity of frequentist statistics, which rely only on the observed data subject to some underlying assumptions.[16] The counterpoint to this is that Bayesian analysis makes any subjectivity explicit, whereas with traditional frequentist methods there may be 'hidden' subjectivity and fragility within the untested assumptions underlying the analytic model used.[15] The computational complexity of Bayesian and simulation methods is much less of a limitation with the increasing power of personal computers.

The concepts outlined above will be illustrated using a simplified case study which is not specific to antimicrobial resistance. The case study relates to the use of Bayesian methods to facilitate the prediction of the probability of target attainment (PTA) of an antimicrobial regimen from a sparse dataset of plasma concentrations observed in a small study.

A PK/PD case study

The optimal dosing of antibiotics for individual patients is a key therapeutic strategy for ID physicians and ID pharmacists. Moreover, optimal dosing for niche patient groups (such as those with altered renal function, critical illness and patients outside usual adult age and weight intervals) is challenging and needs to be addressed. Typically, the available data is limited. The concepts and the technology underlying the literature on population pharmacokinetic/pharmacodynamic (PK/PD) methods have evolved and these methods are possibly confusing.[25–27]

There are at least three common measures derived from population PK/PD methods in relation to the MIC of the antibiotic of interest versus the bacterial agent of interest. These are: (1) the percentage of the dosing-time interval that a drug concentration remains above the MIC; (2) the ratio of the maximal drug concentration to the MIC; and (3) the ratio of the AUC to the MIC. In other contexts, across the broad range of antimicrobial agents, other serum levels may apply. For example, in the context of HIV suppression, the question may relate to maintaining antiretroviral trough levels, being the lowest level during the dosing interval, above an acceptable nadir (which is a level below which the risk

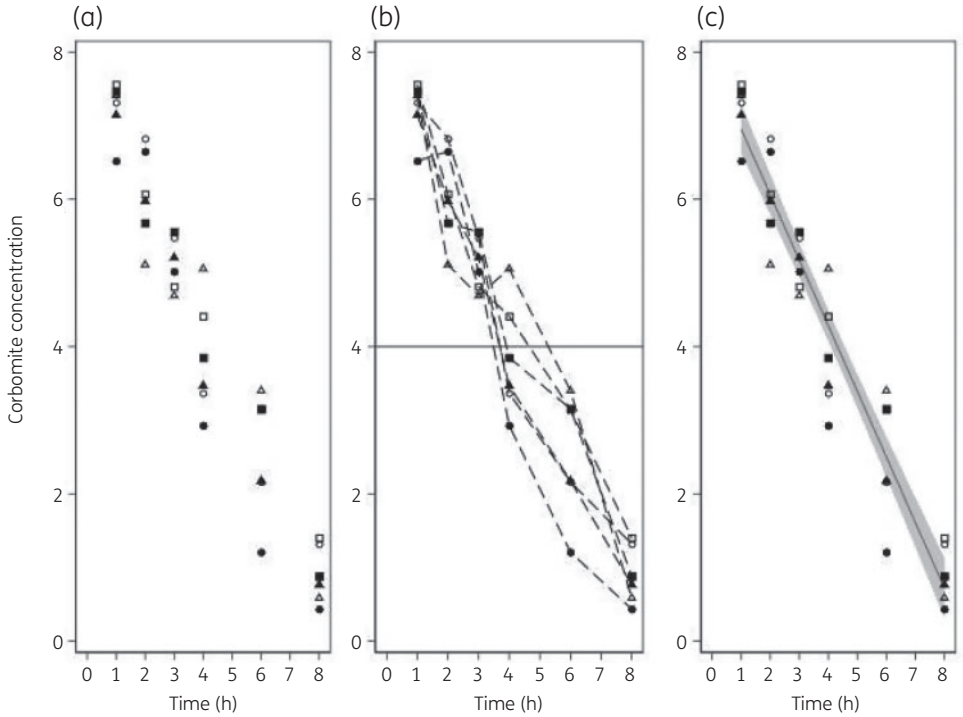

Figure 2. Time course of serum concentration of 'Corbomite' from a fictional pharmacokinetic study presented as raw data (Figure 2a, left), as correlated data (Figure 2b, middle) and as a linear regression of the whole data with 95% CIs (Figure 2c, right). The horizontal line at 4 is the therapeutic target and the individual symbols correspond to individual fictional subjects.

of failure of suppression of viral replication is too high) despite the possibility of missed doses.[28]

Whatever the measure, the fundamental question is usually 'what is the PTA that our antibiotic dosing regimen will achieve in relation to the MIC in this patient population?'. The MIC treatment target is typically that of the defined antibiotic versus either common causative pathogens for the infection syndrome or the MIC versus specific bacterial isolates. Answering the PTA question requires an ability to forecast the relationship between serum levels achieved with antibiotic dosing within our target patient group versus the MIC treatment target over the dosing interval.

A fictitious pharmacokinetic study

Methods to address the PTA question will be illustrated using data from a small PK/PD study of 'Corbomite', a fictional antibiotic (Figure 2; Table S2 available as Supplementary data at *JAC-AMR* Online). The study has observed plasma levels at six timepoints following a defined dose of Corbomite administered to six individuals (Figure 2a). This data has been generated using random numbers as described in the Supplement data. To simplify, the adsorption and protein binding issues are ignored and a one-compartment model with exclusive renal clearance is assumed. The six individuals might be either normal healthy adult volunteers within a Phase 1 study or they might be patients meeting specific inclusion criteria receiving the antibiotic as therapy. That there are only six subjects simplifies the example here. However, the reality is that data available for a particular patient group is often limited. For example, most PK/PD studies of have 60 or fewer subjects and key studies may have fewer than 20.[29]

Data issues

The fictitious pharmacokinetic study data illustrates several issues that require thought. Fundamentally, how do we extrapolate from such sparse data to answer the PTA question for the niche population of which the six individuals are representative? The strengths and limitations of available inference methods and simulation techniques need to be considered.

With the sparse data and the PTA question at hand, frequentist methods have several limitations. With only six observations per timepoint, whether the data is normally distributed is uncertain and with only six timepoints, whether the time course is linear, log-linear, or follows some other distribution is unclear. The six individuals have been sampled repeatedly and their observations will be correlated (Figure 2b). By contrast, sampling of six randomly selected individuals from the broader population at each timepoint, would produce data that is less precise with wider 95% CIs. In one sense, this correlation is a nuisance to the analysis. However, this correlation is pertinent to our PTA question and ideally our analysis should allow for it.

The use of 95% CIs to summarize the data as a whole (Figure 2c) is a typical use of frequentist methods. However, these confidence intervals cannot be used to forecast drug levels in possible future individual patients. The confidence limits represent the uncertainty in the population mean estimate were we to repeat this same study with the same dosing in the same subjects multiple times.

Multi-level non-linear models could be used to model the correlated data for the six individuals. These models are mixed in that they might include either or both of fixed and random effects.

Fixed effects is a term that would describe the six individuals as being of special note or identifiable within the analysis, whereas random effects is a term that would describe the six individuals as being merely representative of the broader population from which they have been drawn. Being multi-level, they attempt to estimate the variability in this broader population that is additional to that observed among the six individuals, although as in the meta-analysis examples, with fewer than ten observations the true variability in the broader population is likely to be underestimated. The intent here is analogous to the intent underlying hierarchical models in the meta-analysis of published results of studies of diagnostic tests[30,31] and the application of multi-level techniques to cluster randomized controlled trials.[32] In each case, the intent of these methods is to use the available data to estimate for a broader population, for which the available data is taken to be representative. With only six individuals in the fictional study, we cannot be certain that this intent will be met.

Simulation methods

One simple generally applicable simulation technique is the bootstrap, which is used to derive robust statistics whenever the actual shape or form of the distribution underlying any limited data is uncertain. The bootstrap principle is that our best guess for the underlying distribution is contained within the six observations we have. If we created a large sample containing only these six observations each replicated many times, 500 replicates for example, and then made random draws each of six observations from this large sample, we would have samples each containing various combinations of the original observations with each observation appearing once, more than once or not at all in these replicate samples. In this simple data set, the mean and 95% CIs derived for Corbomite levels at the first (60 min) timepoint with (6.9; 6.2–7.5) versus without (6.9; 6.0–7.8) bootstrap sampling of the six observations are similar. The use of the bootstrap for our sample illustrates a simple and widely used simulation technique which enables the derivation of robust statistics. However, for the PTA question, the bootstrap does not solve the inferential limitations that arise with the application of frequentist methods to this sparse data.

Bayesian methods

So what simulation methods might help address the PTA question here? We may have prior knowledge that Corbomite serum levels at 60 min after comparable dosing in similar settings are typically 5 (95% CI; 3–7). A Bayesian analytic approach will integrate this information, termed 'the prior' with the observed study data to give a result termed 'the posterior'. For simple data, this may not require simulation. The mean of this simulated posterior will be located between the mean of the prior and the mean of our observed data with the exact location dependent on the relative weighting of each within the analysis. For the example here, the posterior for the Corbomite serum levels at the 60 min timepoint is 6.6 (95% credible interval; 5.5–7.3), which falls between the observed concentration and the prior. Note that each of these entities, the prior, the posterior and the study data, represent data as statistical distributions rather than as point estimates. In this way, the posterior is able to answer the PTA question with a probablity estimate relevant to the individual members of our target population. This is represented as a 95% credible interval, which, in contrast to the 95% confidence interval, has the valid interpretation of containing 95% of the population values. Alternatively, we could derive an interval between the 2.5% and 97.5% percentiles of the posterior data by bootstrapping to derive an interval containing 95% of the observations. These 95% credible and bootstrap-derived intervals are derived by stochastic processes and, in contrast to a confidence interval, will often not be symmetrical and will vary slightly with repeated simulations. Where simulation is used to produce the posterior, even for the integration of the prior and the data at a single timepoint, as here, the simulation is slightly more complex than the simple random sampling underlying the bootstrap. Typically, >1000 simulations might be required.

However, we can do better. We may have knowledge of Corbomite pharmacokinetics which, for the purposes of demonstration, is determined to be unusually simple. This fictional antibiotic undergoes renal clearance with zero order elimination kinetics, that is a constant amount (e.g. so many milligrams) of drug is eliminated per unit time, and the volume of distribution is limited to the intravascular space. Hence, measures of renal function together with indices of intravascular space, such as height and weight, would serve as a logical basis to model the variation in Corbomite serum levels in a simulated population of several thousand patients using indices similar to our patient group of interest. These multiple predictive indices, including estimates of between-individual variation, each have separately defined underlying statistical distributions. Deriving a posterior by integrating these various parameters will require simulation techniques that are substantially more complex.

Monte Carlo simulation techniques provide approximate solutions to intractable quantitative model-based problems using the results from repeated simulations of the model with the input randomly varying each time. Monte Carlo simulation differs from traditional simulation in that the model parameters are treated as stochastic or random variables, rather than as fixed values.[33] Stanislaw Ulam, a mathematician, conceived the method to enumerate by simulation the number of possible outcomes of the card game 'solitaire' using newly developed computers developed for the United States nuclear project during World War II.[34]

Monte Carlo Markov chain (MCMC) is a technique that conjoins random sampling (Monte Carlo) in sequence according to an algorithm (Markov chain). More specifically, MCMC generates the results that form the posterior in sequence by a random process satisfying the Markov property being that the next result over thousands of simulations depends only on the current result and not on any of the previous results. In this way the MCMC will optimally generate the posterior, which otherwise is an unknowable distribution, one result at a time. How the MCMC generates the posterior is best understood from watching two moving image demonstrations available online.[35,36]

Executing an MCMC is computer intensive and 5000 iterations would generally be regarded as a minimum. How many iterations might be required depends on the efficiency of the MCMC and the complexity of the posterior distribution that is being generated. Also note that by generating results in sequence, the MCMC creates some autocorrelation between the results in the posterior distribution. This autocorrelation is highest within the initial iterations of

any simulation and typically, the first 1000 simulations, termed the 'burn in', are usually discarded. Some MCMC sampling algorithms may be more efficient than others, but all require model checking.

Figure 3 shows the result of a simulation model of the PK/PD study as described in Figure 2 using a prior which indicates our pre-existing knowledge of the slope and intercept values. This simulation will require model checking for all three estimated parameters; the slope (Figure 4 and Figure S1), the intercept (Figure S2) and the variance (Figure S3).

Figure 4 demonstrates the model checking of the slope parameter generated in association with the previous simulation. This checking is undertaken by reference to graphical diagnostics generated as part of running the MCMC algorithm to ensure that the MCMC has converged (Figure 4a), that the algorithm has adequately explored the parameter space (Figure 4b), and there is minimal residual autocorrelation following the 'burn in' (Figure 4c) within the posterior distribution. If not, a larger 'burn in' may be required. That the density obtained after the first half (5000 iterations) of the simulation is similar to that obtained after the second half and overall (10 000 iterations) suggest that a sufficient number of iterations was executed (Figure 4d). More so than frequentist methods, model checking is essential for all parameters generated by MCMC simulation to ensure that we 'know' our hitherto 'unknown' posterior within our Bayesian analysis.

At the end of this model building, simulation and model checking, the model is ready to be validated using additional data derived from observation. The validation data is used to assess the goodness of fit of the simulation model data by visual predictive checking. Most simply, this is done by overlaying the serum concentrations of the validation data versus the model-generated data. A good fit would be indicated by observing the validation data to fit within the 95% credible intervals of the model data. This would serve to validate the model. This type of analysis is an example of 'confronting the model with the data'.[37,38] Using simulations from this validated model enables forecasting the relationship between serum levels of Corbomite for our population of interest in relation to the MIC of interest and the PTA question can be answered.

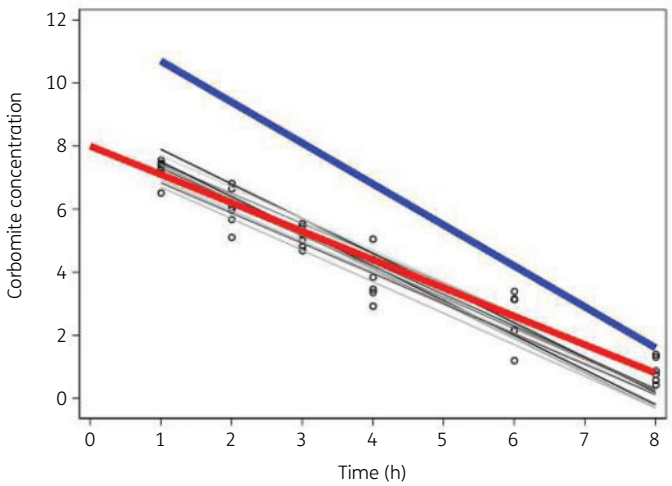

Figure 3. MCMC simulation of serum levels of Corbomite for the entire time course of a fictional pharmacokinetic study as described in Figure 2. The prior (solid blue line) has a y intercept at 12 and a slope of $-1.3x$. The simulated 'observations' are the open symbols and these have been derived by addition of random noise to a line having a y intercept at 8 and a slope of $-0.9x$ (solid red line). The first ten MCMC simulations from the posterior are represented by the thin regression lines (grey). Note that these ten lines do not align with the 'observations' but have slope and intercept values that are between those of the prior and the 'observations'.

This example illustrates several limitations of simulation methods. Firstly, there needs to be a well-defined model on which to base the simulation. There may even be more than one candidate model. For example, even for an extensively studied antibiotic such as gentamicin, models that are one-, two-, and even three-compartment have been described.[39] Secondly, simulation results may be influenced by the starting values and how long to run a simulation is difficult to define. The diagnostics for convergence, beyond a simple visual inspection of the trace plot, are not simple. Complex models may require multiple concurrent MCMC chains with a comparison of the results derived from the different chains.

Emerging applications

The concepts underlying the PTA question are also applicable to a range of ID-related research questions that are otherwise difficult to address. These concepts are most useful for rare diseases or outcomes. Bayesian analysis can be applied to questions where the available data is sparse. By contrast, in this context, traditional frequentist methods, being dependent on large-sample approximation as a key underlying assumption, may suffer from sparse data bias.[40]

The application of Bayesian methods is not limited to the PTA question. For example, Bayesian techniques can use historical study data to inform the setting of non-inferiority margins for clinical trials of antibiotics,[41] where to set MIC breakpoints for new antibiotics[42] and can address safety issues such as the question of whether the mortality risk associated with cefepime use for patients with febrile neutropenia is unsually high.[43]

Within the hospital setting, there are examples relating to hospital-based infections such as the application of MCMC and Bayesian methods to predict causative pathogens and target antibiotic therapy for patients with ventilator-associated pneumonia[44] and to infer hospital transmission dynamics from imperfect data.[45]

There are increasing examples relating to the incidence rates of infectious diseases and antibiotic resistance in community studies. MCMC-based modelling has been used to identify factors other than Zika virus infection that might help map the spread of microcephaly in Brazilian municipalities,[46,47] and how *Wolbachia*-infected mosquitoes could best be targeted to reduce dengue virus transmission in Indonesia.[48] Another example used a multi-state model of tuberculosis together with a model-based PK/PD approach to increase the power to forecast drug activity as part of expediting the clinical evaluation of tuberculosis drug development.[49]

In an application to the study of antibiotic resistance epidemiology, Bayesian methods have been used to define and quantify a progressive increase in the MIC (termed MIC creep) within the non-resistant population within *Salmonella enterica* over a period of 15 years of surveillance. This creep would not otherwise have been evident from the MIC values as dichotomized into susceptible and resistant, as conventionally reported.[50]

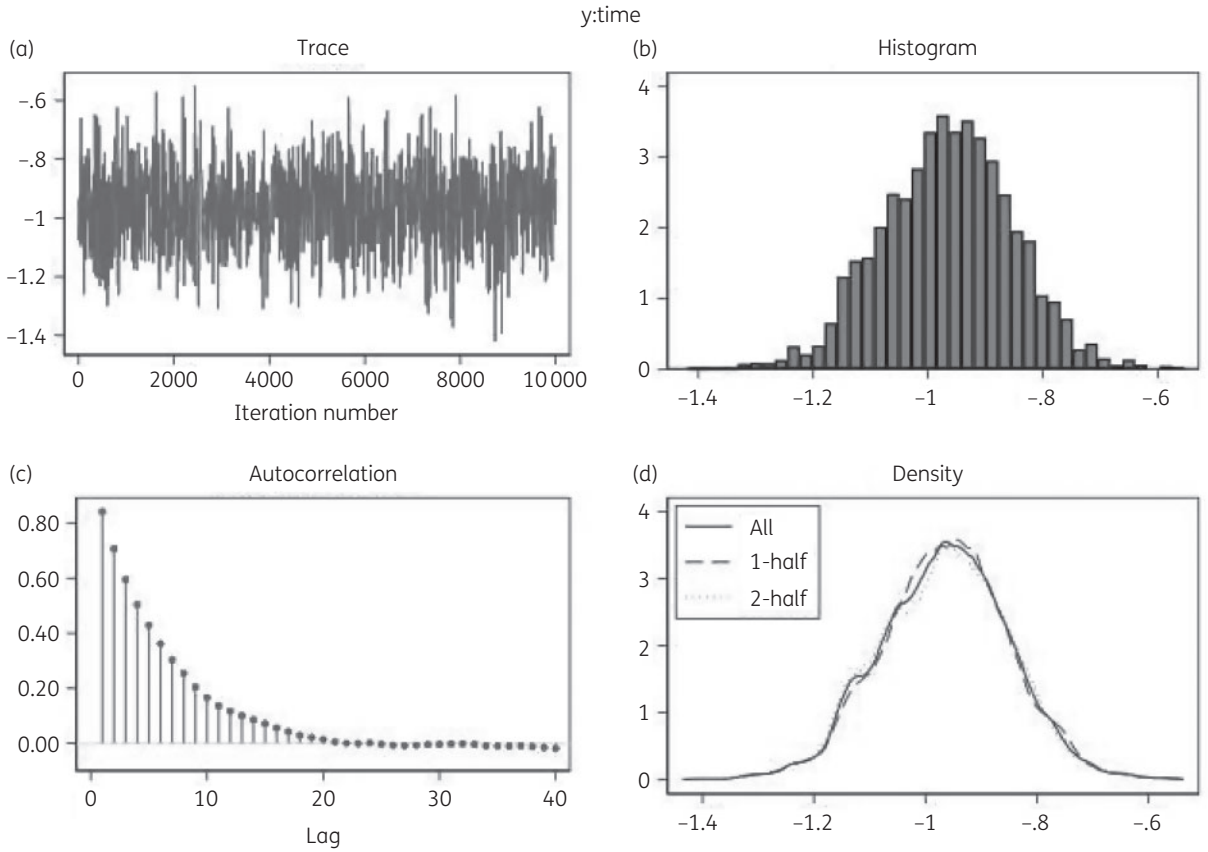

Figure 4. Diagnostic plots for the derivation of the slope parameter by MCMC simulation in Figure 3. (a) The MCMC trace plot shows the simulated values versus the MCMC iteration number and demonstrates a well-mixing parameter which traverses the posterior domain rapidly with a nearly constant mean and variance resulting in a 'hairy caterpillar' appearance to the trace. (b) The histogram depicts the general shape of the marginal posterior distribution. (c) The autocorrelation trace decreases from an initial positive value toward 0 as the simulation progresses. Autocorrelation that fades rapidly indicates a well-mixing MCMC chain. (d) Kernel density plots show three density curves as smoothed histograms (with the first-half, the second-half, and overall density obtained using the first half, the second half and the whole of the MCMC sample, respectively). The three density curves lying close to each other indicates that the chain converged early and with good mixing.

These examples illustrate the increasing use of simulation in combination with Bayesian methods to address ID research questions that would otherwise be difficult to address. Moreover, they represent examples of confronting models with data, to enable answers to key research and policy questions that are meaningful, quantitative and potentially relevant.[51]

Funding

This work was supported by the Australian Government Department of Health and Ageing through the Rural Clinical Training and Support (RCTS) program.

Transparency declarations

None to declare.

Author contributions

J.H. performed the statistical analysis and wrote the manuscript. J.H. and D.B. read and approved the final manuscript and are the guarantors of the paper.

Supplementary data

Tables S1 and S2 and coding information are available as Supplementary data at *JAC-AMR* Online

References

1 Shallcross LJ, Davies SC. The World Health Assembly resolution on antimicrobial resistance. *J Antimicrob Chemother* 2014; **69**: 2883–5.

2 Thwaites GE, Scarborough M, Szubert A *et al.* Adjunctive rifampicin for *Staphylococcus aureus* bacteremia (ARREST): a multicentre, randomised, double-blind, placebo-controlled trial. *Lancet* 2018; **391**: 668–78.

3 Harris PN, Tambyah PA, Lye DC *et al.* Effect of piperacillin-tazobactam vs meropenem on 30-day mortality for patients with *E. coli* or *Klebsiella pneumoniae* bloodstream infection and ceftriaxone resistance: a randomized clinical trial. *Jama* 2018; **320**: 984–94.

4 Thwaites GE, Edgeworth JD, Gkrania-Klotsas E *et al.* Clinical management of *Staphylococcus aureus* bacteremia. *Lancet Infect Dis* 2011; **11**: 208–22.

5 Ng TM, Khong WX, Harris PN *et al.* Empiric piperacillin-tazobactam versus carbapenems in the treatment of bacteremia due to extended-spectrum β-lactamase-producing Enterobacteriaceae. *PLoS One* 2016; **11**: e0153696.

6 Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Stat Med* 2008; **27**: 625–50.

7 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**: 1539–58.

8 Jung SY, Lee SH, Lee SY *et al*. Antimicrobials for the treatment of drug-resistant *Acinetobacter baumannii* pneumonia in critically ill patients: a systemic review and Bayesian network meta-analysis. *Crit Care* 2017; **21**: 319.

9 Cosgrove SE, Sakoulas G, Perencevich EN *et al*. Comparison of mortality associated with methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* bacteremia: a meta-analysis. *Clin Infect Dis* 2003; **36**: 53–9.

10 Hurley JC. Comparison of mortality associated with methicillin-susceptible and methicillin-resistant *Staphylococcus aureus* bacteremia: an ecological analysis. *Clin Infect Dis* 2003; **37**: 866–8.

11 Hurley JC. Risk of death from methicillin resistant *Staphylococcus aureus* bacteremia: a meta-analysis. *Med J Aust* 2002; **176**: 188–9.

12 Blot SI, Vandewoude KH, Hoste EA *et al*. Outcome and attributable mortality in critically ill patients with bacteremia involving methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*. *Arch Intern Med* 2002; **162**: 2229–35.

13 Wolkewitz M, Frank U, Philips G *et al*. Mortality associated with in-hospital bacteremia caused by *Staphylococcus aureus*: a multistate analysis with follow-up beyond hospital discharge. *J Antimicrob Chemother* 2011; **66**: 381–6.

14 Hurley JC. Mortality due to vancomycin-resistant enterococcal bacteremia versus vancomycin-susceptible enterococcal bacteremia: an ecological analysis. *Clin Infect Dis* 2005; **41**: 1541–2.

15 McElreath R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press, 2020.

16 Goodman SN. Toward evidence-based medical statistics. 1: The *P* value fallacy. *Ann Intern Med* 1999; **130**: 995–1004.

17 Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999; **130**: 1005–13.

18 Walsh M, Srinathan SK, McAuley DF *et al*. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol* 2014; **67**: 622–8.

19 Johnson SW, Brown SV, Priest DH. Effectiveness of oral vancomycin for prevention of healthcare facility–onset *Clostridioides difficile* infection in targeted patients during systemic antibiotic exposure. *Clin Infect Dis* 2020; **71**: 1133–9.

20 Nolen L. How medical education is missing the bull's-eye. *N Engl J Med* 2020; **382**: 2489–91.

21 Kalil AC. Deciphering the sepsis riddle: We can learn from Star Trek. *Crit Care Med* 2013; **41**: 2458–60.

22 Roberts JA, Kirkpatrick CM, Roberts MS *et al*. Meropenem dosing in critically ill patients with sepsis and without renal dysfunction: intermittent bolus versus continuous administration? Monte Carlo dosing simulations and subcutaneous tissue distribution. *J Antimicrob Chemother* 2009; **64**: 142–50.

23 Jager NG, van Hest RM, Xie J *et al*. Optimization of flucloxacillin dosing regimens in critically ill patients using population pharmacokinetic modelling of total and unbound concentrations. *J Antimicrob Chemother* 2020; **75**: 2641–9.

24 Roberts JA, Paul SK, Akova M *et al*. DALI: defining antibiotic levels in intensive care unit patients: are current β-lactam antibiotic doses sufficient for critically ill patients? *Clin Infect Dis* 2014; **58**: 1072–83.

25 Bonate PL. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer, 2011.

26 Drusano GL. Pharmacokinetics and pharmacodynamics of antimicrobials. *Clin Infect Dis* 2007; **45** Suppl 1: S89–S95.

27 Nielsen EI, Friberg LE. Pharmacokinetic-pharmacodynamic modeling of antibacterial drugs. *Pharmacol Rev* 2013; **65**: 1053–90.

28 Olagunju A, Schipani A, Bolaji O *et al*. Evaluation of universal versus genotype-guided efavirenz dose reduction in pregnant women using population pharmacokinetic modelling. *J Antimicrob Chemother* 2018; **73**: 165–72.

29 McAleenan A, Ambrose PG, Bhavnani SM *et al*. Methodological features of clinical pharmacokinetic–pharmacodynamic studies of antibacterials and antifungals: a systematic review. *J Antimicrob Chemother* 2020; **75**: 1374–89.

30 Hurley JC. Does gram-negative bacteremia occur without endotoxaemia? A meta-analysis using hierarchical summary ROC curves. *Eur J Clin Microbiol Infect Dis* 2010; **29**: 207–15.

31 Hurley J. Meta-analysis of clinical studies of diagnostic tests: developments in how the receiver operating characteristic "works". *Arch Pathol Lab Med* 2011; **135**: 1585–90.

32 Hurley JC. How the Cluster-randomized Trial "Works". *Clin Infect Dis* 2020; **70**: 341–6.

33 Bonate PL. A brief introduction to Monte Carlo simulation. *Clin Pharmacokinet* 2001; **40**: 15–22.

34 Metropolis N, Ulam S. The monte carlo method. *J Am Stat Assoc* 1949; **44**: 335–41.

35 Stata HC, Introduction to Bayesian statistics. Part 1: the basic concepts. 2016. https://www.youtube.com/watch?v=0F0QoMCSKJ4&feature=youtu.be.

36 Stata HC, Introduction to Bayesian statistics, part 2: MCMC and the Metropolis-Hastings algorithm. 2016. https://www.youtube.com/watch?v=OTO1DygELpY&feature=youtu.be.

37 Cooper BS. Confronting models with data. *J Hosp Infect* 2007; **65**: 88–92.

38 Hurley JC. Structural equation modeling the "control of gut overgrowth" in the prevention of ICU-acquired Gram-negative infection. *Crit Care* 2020; **24**: 1–2.

39 Llanos-Paez CC, Hennig S, Staatz CE. Population pharmacokinetic modelling, Monte Carlo simulation and semi-mechanistic pharmacodynamic modelling as tools to personalize gentamicin therapy. *J Antimicrob Chemother* 2017; **72**: 639–67.

40 Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *Bmj* 2016; **352**: i1981.

41 Ambrose PG, Hammel JP, Bhavnani SM *et al*. Frequentist and Bayesian pharmacometric-based approaches to facilitate critically needed new antibiotic development: overcoming lies, damn lies, and statistics. *Antimicrob Agents Chemother* 2012; **56**: 1466–70.

42 Drusano GL, Preston SL, Hardalo C *et al*. Use of preclinical data for selection of a phase II/III dose for evernimicin and identification of a preclinical MIC breakpoint. *Antimicrob Agents Chemother* 2001; **45**: 13–22.

43 Kalil AC. Is cefepime safe for clinical use? A Bayesian viewpoint. *J Antimicrob Chemother* 2011; **66**: 1207–9.

44 Forrester ML, Pettitt AN, Gibson GJ. Bayesian inference of hospital-acquired infectious diseases and control measures given imperfect surveillance data. *Biostat* 2007; **8**: 383–401.

45 Visscher S, Kruisheer EM, Schurink CA *et al*. Predicting pathogens causing ventilator-associated pneumonia using a Bayesian network model. *J Antimicrob Chemother* 2008; **62**: 184–8.

46 Amaral P, Resende de Carvalho L, Hernandes *et al*. Geospatial modeling of microcephaly and zika virus spread patterns in Brazil. *PloS One* 2019; **14**: e0222668.

47 Bautista LE. Maternal Zika virus infection and newborn microcephaly—an analysis of the epidemiological evidence. *Annal Epidemiol* 2018; **28**: 111–8.

48 O'Reilly KM, Hendrickx E, Kharisma DD *et al*. Estimating the burden of dengue and the impact of release of wMel Wolbachia-infected mosquitoes in Indonesia: a modelling study. *BMC Med* 2019; **17**: 4.

49 Svensson RJ, Gillespie SH, Simonsson US. Improved power for TB Phase IIa trials using a model-based pharmacokinetic–pharmacodynamic approach compared with commonly used analysis methods. *J Antimicrob Chemother* 2017; **72**: 2311–9.

50 Zhang M, Wang C, O'Connor AA. Hierarchical Bayesian latent class mixture model with censorship for detection of linear temporal changes in antibiotic resistance. *PloS One* 2020; **15**: e0220427.

51 Hurley JC. *Candida–Acinetobacter–Pseudomonas* Interaction Modelled within 286 ICU Infection Prevention Studies. *J Fungi* 2020; **6**: 252.