Stewarding antibiotic stewardship in intensive care units with Bayesian artificial intelligence [version 1; referees: 1 approved with reservations]

Tavpritesh Sethi¹-³, Shubham Maheshwari³, Aditya Nagori⁴, Rakesh Lodha¹

¹All India Institute of Medical Sciences, New Delhi, India
²Stanford University School of Medicine, Stanford, California, USA
³Indraprastha Institute of Information Technology, Delhi, New Delhi, India
⁴CSIR-Institute of Genomics and Integrative Biology, New Delhi, India

Abstract
Emerging antimicrobial resistance (AMR) is a global threat to life. Injudicious use of antibiotics is the biggest driver of resistance evolution, creating selection pressures on micro-organisms. Intensive care units (ICUs) are the strongest contributors to this pressure, owing to high infection and antibiotic usage rates. Antimicrobial stewardship programs aim to control antibiotic use; however, these are mostly limited to descriptive statistics. Genomic analyses lie at the other extreme of the value-spectrum, and together these factors predispose to siloing of knowledge arising from AMR stewardship. In this study, we bridged the value-gap at a Pediatric ICU by creating Bayesian network (BN) artificial intelligence models with potential impacts on antibiotic stewardship. Methods, actionable insights and an interactive dashboard for BN analysis upon data observed over 3 years at the PICU are described. BNs have several desirable properties for reasoning from data, including interpretability, expert knowledge injection and quantitative inference. Our pipeline leverages best practices of enforcing statistical rigor through bootstrapping, ensemble averaging and Monte Carlo simulations. Competing, shared and independent drug resistances were discovered through the presence of network motifs in BNs. Inferences guided by these visual models are also discussed, such as increasing the sensitivity testing for chloramphenicol as a potential mechanism of avoiding ertapenem overuse in the PICU. Organism, tissue and temporal influences on drug co-resistances are also discussed. While the model represents inferences that are tailored to the site, BNs are excellent tools for building upon pre-learnt structures, hence the model and inferences were wrapped into an interactive dashboard not only deployed at the site, but also made openly available to the community via GitHub. Shared repositories of such models could be a viable alternative to raw-data sharing and could promote partnering, learning across sites and charting a joint course for antimicrobial stewardship programs in the race against AMR.

Keywords
Antimicrobial Stewardship, ICU, Bayesian Networks, Artificial Intelligence, Modeling, Inference, Interactive Dashboard
Corresponding authors: Tavpritesh Sethi (tavsethi@stanford.edu), Rakesh Lodha (rlodha1661@gmail.com)

Author roles: Sethi T: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Maheshwari S: Formal Analysis, Methodology, Software, Writing – Original Draft Preparation; Nagori A: Data Curation, Validation; Lodha R: Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust/DBT India Alliance Early Career grant IA/CPHE/14/1/501504 to Tavpritesh Sethi.

Copyright: © 2018 Sethi T et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Sethi T, Maheshwari S, Nagori A and Lodha R. Stewarding antibiotic stewardship in intensive care units with Bayesian artificial intelligence [version 1; referees: 1 approved with reservations] Wellcome Open Research 2018, 3:73 (doi: 10.12688/wellcomeopenres.14629.1)

Introduction

The world is poised to enter the post-antibiotic era as per the World Health Organization report on Antimicrobial Resistance (AMR)\(^1\). This is an alarming responsibility, which needs radical and innovative approaches to tackle the steadily growing resistance in organisms. In the United States alone, AMR infections kill about 23,000 of at least 2 million lives affected each year\(^2\). The few statistics that are available for developing countries, such as India, indicate a much higher rate of AMR\(^3\). As a response to this threat, antimicrobial stewardship programs are being launched all over the world to measure, monitor and promote the judicious use of antibiotics. These programs are especially relevant to the intensive care units (ICUs) for three reasons. First, the widespread incidence and transmission of healthcare-associated infections (HAIs) coupled with high antibiotic usage creates a huge selection pressure upon micro-organisms to evolve resistance mechanisms by accumulating mutations. Secondly, stewardship programs are easier to implement under the monitored prescription and use of antibiotics, thus making ICUs an operationally convenient choice. Thirdly, the acute physiological decompensation in patients admitted to ICUs makes them especially vulnerable to rapid invasion and death due to sepsis. Together, these factors make ICUs a major breeding ground of AMR. Indeed, it is estimated that as many as two-thirds of bacteremia cases in ICUs are caused by multi-drug resistant (MDR) organisms\(^4\).

To this end, antimicrobial stewardship programs have advocated the regular use of antibiograms (i.e. charts and tables of data summarizing the sensitivity patterns of micro-organisms). Despite being a valuable tool, antibiograms fail to explicitly capture and model the co-occurrence of resistance among bacterial isolates. This is common, owing to the shared genetic mechanisms of acquiring resistance. Furthermore, resistance to a class of antimicrobial is inevitably linked with the compensatory overuse of a different class of antimicrobial, thus creating a dependency structure in the antibiogram. Network analysis, a popular method of representing complex data, can help gain qualitative understanding of resistance through visualization of associations. However, association networks do not allow quantitative reasoning with data or actionable decision-making.

Bayesian networks (BNs)\(^5\)\(^6\), on the other hand, are probabilistic graphical models that provide a single joint-multivariate fit on the data. A BN is a directed acyclic graph representing conditional independencies and their associated parameters estimated from underlying data (Figure 1). Learning a BN model directly from data is a two-stage process involving (i) structure learning and (ii) parameter fitting for inference. Structure encapsulates statistical concepts (such as direct effects, confounders, mediators and inter-causal effects) into intuitive graphical motifs. Fitting parameters then allows probabilistic (and sometimes causal) reasoning in observational data and causal inference in the presence of interventional data. Owing to these properties, BNs have been extensively used in medicine as artificial intelligence expert systems for guiding decisions\(^7\). However their use in guiding and creating continuously-learning antibiotic stewardship programs has been largely unexplored.

The first generation of BNs mostly relied upon specification of model structure through expert knowledge or using that as a starting point for the model\(^8\). Previous studies in modeling antimicrobial use have used such an approach to define a causal structure for optimal treatment of bacterial infections\(^9\),\(^10\). However, we reasoned that bacterial resistance mechanisms are complex, multi-factorial and the mechanisms of cross-tolerance are not completely elucidated. Thus, learning the structure directly from data is a reasonable starting point for guiding site-specific antibiotic stewardship and a learning AMR decision system should allow sanitization of the learnt structure using expert knowledge. Finally, the biggest opportunity waiting to be exploited is learning across sites through integrated systems and shared models\(^11\). To address these directions, here we report the methods and insights from constructing a data-driven BN for AMR modeling at our clinical site. We hypothesized that construction of such models could not only provide quantitative insights into resistance patterns but also create a tailored...

**Figure 1. Culture-data characteristics.** (a) Among culture-positive samples, a few organisms contribute to highest number of infections. (b) An example correlation plot for validating the shared sensitivity patterns in the cohort.
learning system for general and site-specific guidelines. This is, to the best of our knowledge, the first report demonstrating the utility structure-learning and BN inference for guiding antibiotic stewardship in the ICU.

Methods
Cohort description
The study was carried out the Pediatric ICU at the All India Institute of Medical Sciences (AIIMS), New Delhi, India, a tertiary care hospital in New Delhi, India, and is a part of the ongoing initiative to build a Sepsis Advanced Forecasting Engine for ICU (SAFE-ICU). Since the data analytics under the SAFE-ICU project do not involve direct patient contact or alteration of patient management, a waiver of consent was obtained from the AIIMS Institute Ethics Committee (Ref. No. IEC/NP-211/08.05.2015, AA-2/09.02.2017). The period of data extraction was from July 25, 2014 to February 17, 2018 and data were extracted as per a retrospective observational study design. The population of interest was children admitted to the Pediatric ICU.

Data extraction and exploratory data analysis
Queries were written to extract data tables from the PostgreSQL database containing the ICU data. Data merging, cleaning, reshaping and date-time handling were carried out in R v.3.4.3\(^1\), dplyr v.0.7.5\(^1\) and lubridate v.1.6.0\(^1\) libraries. Exploratory data analysis for the visualization of cohort demographics, isolate frequencies and correlations among the sensitivity patterns was conducted using ggplot2 v.2.2.1\(^1\). The final data consisted of date-time and nominal discrete variables, including sample description, organism isolates and antibiotic sensitivity results (sensitive, resistant, indeterminate and not tested).

Association network analysis
Since the data were nominal with two or more levels, pairwise Cramer’s V inter-correlations\(^1\) were used to calculate associations. The inter-correlation list was pruned at the Cramer’s V threshold of greater than 0.75 and the network was visualized using the visNetwork v.2.0.3 library\(^1\).

Structure learning and parametrization
Learning the BN structure and fitting the conditional probabilities involves a sequence of steps to ensure the recovery of valid models\(^1-3\). The structural independencies of antibiotic sensitivity were learnt directly from the data using an ensemble structure learning approach. An arbitrarily high number of 351 bootstraps (sampling with replacement) were constructed from data. An odd number of bootstraps ensured the breaking of ties, while constructing the ensemble averaged network structure. The hill-climbing method, a score-based structure learning algorithm, was applied on each of the bootstrap samples. An edge and its direction were accepted when both of these were present in at least 51% of the learnt structures. The stability of structures was assessed through agreement across multiple scoring criteria, such as Bayesian information criteria, Bayesian Dirichlet equivalent score and Akaike information criteria. For each of these criteria, the imaginary sample size was kept low (three), which allowed the algorithm to learn structure from data rather than relying heavily on priors. Parametrization of conditional probabilities was done using Bayesian parameter estimation. These computations were carried out using the bnlearn v.4.3 library\(^1\) in R.

Visualization, inference and dashboard construction
The learnt structure was visualized using a tree layout\(^7\) for the identification of root nodes, branches and terminal nodes in the AMR independencies. This visualization guided the expert evaluation and queries on the network. Queries in the AMR model were conducted by setting evidence followed by evaluation of change in conditional probabilities (inference) effected by the change in state of the evidence. Inference was carried out using both exact and approximate algorithms. Error bars were constructed using 25 rounds of approximate inference and the results were compared with exact inference. The network and its functionalities are deployed as an interactive dashboard for continuous exploration of inferences. The development version of the dashboard, complete with inferences, is available at GitHub and the version 1.0 release is available at Zenodo\(^20\) under a GNU General Public License v3.0.

Results
Exploratory data analysis of culture data
A total of 5061 cultures were ordered during the specified duration of the study, and included samples from sources such as the blood, cerebrospinal fluid (CSF), urine, broncho-alveolar lavage fluid and pus. A total of 549 cultures tested positive for one or more organisms (culture positive). Figure 1a shows the organisms isolated from the blood cultures that were positive irrespective of site. As expected, this followed a negative exponential trend, with a few organisms (Pseudomonas sp., Acinetobacter sp., Klebsiella sp., Escherichia sp.) contributing to the majority of cases of sepsis. Alarmingly, drug-resistant Acinetobacter has been posited to be an increasing threat in ICUs the world over, indicating the targeted importance of stewardship. Figure 1b shows an example inter-correlation heat map for drug-sensitivity observed for Pseudomonas species. Hierarchical clustering revealed blocks of shared drug resistance for Pseudomonas. The presence of inter-correlation patterns motivated us to carry out further exploration of associations and inferences using a combination of association networks and BN frameworks. The raw and analyzed data are available publicly on OSF\(^21\).

Association networks analysis for shared antibiotic resistance
We carried out an association networks analysis on the entire dataset to look for global patterns. Cramer’s V indices higher than 0.75 were preserved for the network (Figure 2). Since association networks are undirected, a star layout with the strongest hub represented at the center is a preferred for visualization. Notably, the network reveals that the organism is a central hub for drug associations. Whereas previous studies have looked at gene-level association networks for cross-resistance\(^22\), this network reveals gross microbiological correlations. These are more feasible and potentially more insightful for stewardship programs across sites. An association network is not a model, hence it is limited to insights but not quantitative decisions. Thus, in order to provide
actionable insights for the program, we resorted to a probabilistic graphical model (BN), which not only reveals the bare-bones skeleton of a network after correcting for confounding, but also allows the model to be queried by the clinicians to gain quantitative predictions.

Colliders (V-structures) in the AMR model and inferences

The ensemble structure learnt upon bootstrapped samples represents the optimum model explaining the data. A probabilistic graphical model (PGM) can be visualized as a network, with nodes as variables, edges as interactions and the directions as parent-child relationships. A BN is a class of PGMs known as directed acyclic graphs and it encodes the conditional probabilities and independency structure in the data. Relying on conditional probabilities makes it sparse, hence more interpretable than other graphical models. Directed edges permit inferences to be drawn in both observational and interventional settings. We use these features for reasoning about the antibiotic sensitivity patterns observed in our data. For example, the presence of a collider with two edges pointing towards a node represents competing hypotheses, and observing one can explain away the other (inter-causal reasoning). We observed an important collider in the ensemble network. As an example, a collider was observed between gentamicin (high-level), amikacin, and linezolid with gentamicin as the child node (Figure 3a). Model inference revealed that the resistance to amikacin increased the belief about resistance to gentamicin, as expected. Setting the state of the model as “gentamicin and amikacin resistant” revealed a 55% probability of the organism being sensitive to linezolid (Figure 3b). While this can be partially explained by the gram-positive and gram-negative spectrums of linezolid and the aminoglycosides, respectively, the collider informs that knowing the amikacin status is not enough to reason about linezolid in ambiguous situations. Observing gentamicin high level resistance is essential to open up the path for reasoning between these linezolid and amikacin. Notably, the remaining probability was not apportioned to “resistant”, but to “not-tested”, indicating that stewardship programs may advise more frequent testing of linezolid when there is a strong suspicion of an infection with gram-positive organisms. We chose this example because none of these three antibiotics was directly connected with the organism node. In another related collider formed between amikacin, netilmicin (both active on the gram-negative spectrum) and the organism tested we could reason that netilmicin sensitivity in the presence of amikacin resistance was higher for *Escherichia* (8.3%) than *Klebsiella* (3.4%). More complex inferences with multiple conditions are available for exploration on the dashboard.

Chains and forks in the AMR model and inferences

Chain and fork structures are other actionable motifs that are helpful in guiding BN inference. A chain represents a sequential flow of belief with mediating variables, whereas a fork indicates the presence of a confounder (a common parent) between two variables. An obvious chain obtained in the network was the parent-child relationship between piperacillin + tazobactam (PT) sensitivity and the organism isolated. Unlike an association network, where the organism isolated had connections with multiple antibiotics, the BN revealed this drug combination as the single parent of organism. Therefore, setting the state of the PT node to sensitive revealed that the organism isolated was most likely *Pseudomonas*, a common microbiological decision. On the other hand, setting the node to resistant revealed the most likely organism was *Acinetobacter*. This finding was confirmed on reversing the query direction, revealing a 90% probability of *Acinetobacter* isolates being PT resistant, as compared to a 29% probability of *Pseudomonas* isolates being resistant (Data not
Figure 3. Bayesian network model for guiding stewardship. An ensemble average structure from 351 bootstrapped networks was learnt. An illustrative confirmatory collider (a) was seen between linezolid (green), amikacin (red) and gentamicin high-level resistance, indicating that the learnt structure was able to capture gram-spectrum. Additionally it was able to provide quantitative inferences such as the need for increasing linezolid testing (b) in the presence of gentamicin + amikacin resistance as linezolid is a potential driver being close to the root of the network. Mediators (chains) and confounders (forks) are the other actionable motifs in a BN (c). Here, sample type description (red) was seen to have a confounder effect (fork) between chloramphenicol (green) and nitrofurantoin (blue) indicating that it needs to be controlled for drawing relationships between these two drugs. However, the same variable was an intermediate node (chain) in the flow of probabilistic influence between colistin and chloramphenicol, hence it should not be controlled if relationships between colistin and chloramphenicol need to be drawn. A simple inference query for chloramphenicol sensitivity shows that more peritoneal fluid samples need to be tested for this drug as isolates from this site show a sensitive to resistant ratio of 2:1 (d) whereas isolates from CSF has this ratio as 3:5 thus indicating the importance of site while taking decisions for testing.

shown; available on the Dashboard). Similarly a fork structure involving chloramphenicol provided insights about the patient-sample being tested (Figure 3c). Chloramphenicol was most tested in the CSF and peritoneal fluid. While the probabilities of chloramphenicol sensitivity were similar (33%) for CSF and peritoneal fluid (Figure 3d), a much higher number of peritoneal samples were not tested for chloramphenicol sensitivity, which may be a missed opportunity before resorting to severe antibiotics such as ertapenem, for which the resistance in both CSF and peritoneal fluid were similar. This indicates a potential advantage of increasing chloramphenicol sensitivity in stewardship programs. It is important to note that these inferences need not reflect the clinical decisions that would need a balance between chloramphenicol’s potentially hematotoxicity in children versus the rapid development of multi-drug resistance due to non-judicious use of ertapenem.

Dashboard
Even though the BN is sparse, there is a multitude of scenarios that can be set and are beyond the scope of results. Hence, we created an interactive dashboard, not only for the use at the clinical site, but also for public dissemination (Figure 4). Exploring the model developed at our site may help formulate shared policies and feedback. Moreover, since BNs allow continuous incorporation of prior knowledge, this model will serve as a template for our site with continuous validation over time. With
enough iterations and models shared over other ICU sites, it will be possible to derive tailored components for site-specific policy while preserving generalization of the global graphical structure.

**Discussion**

This study approaches the problem of antimicrobial stewardship in a new way by making it more actionable and transparent through a probabilistic machine learning strategy, termed BNs. The objective was to create methods, pipelines and tools for a continuously learning system that can allow expert interaction and inferences. BNs were chosen as a model for this strategy based upon the combination of data type (discrete nominal), interpretability, potential to inject expert knowledge at all stages of learning and maturity of computational methods to allow direct learning from the data. Unlike other popular methods of machine learning, such as random forests and deep learning, which are best suited to prediction problems, BNs are interpretable and aim to bring transparency to the hidden mechanisms that could have generated the data. Hence BNs offer the perfect combination when insights and predictions are both desirable. Although the interpretation of edge-directions learnt from the data has been a source of some confusion and debate, this is not difficult to resolve. In the absence of interventional data, the edge-directions usually do not represent causal influence. An exception to this is when special criteria such as the front-door and back-door criteria are fulfilled, an investigation beyond the scope of this work. Therefore in this work the directions are not ascribed a causal interpretation, i.e. we do not imply that resistance to Drug A causes resistance to Drug B in our ICU. Rather, even in the absence of causal inference, we use the discovered motifs to reason about the hidden process, thus discovering insights for informing future causal interventions in antibiotic use.

A major hurdle in the use of BNs in large scale data has been their computational complexity. However, theoretical and computational advances (e.g. the use of parallel computation for structure learning) in the recent years have made these tractable. Furthermore, theoretical advances in analytic methods has facilitated learning from interventional data, thus enabling network causal inferences. In this study, we used the state-of-the-art methods, supplementing these with our own wherever needed for reliable and interpretable modeling of complex AMR patterns. The novel analytic contributions of this work were stable approximate inference with error bars (through resampling), its comparison with exact inference, modularity detection and shared-dashboard.

The phenomenon of antibiotic class restriction causing compensatory overuse of other classes has been long known. This leads to the creation of dependency structures in AMR patterns. A classic example supporting this was the clinical trial conducted by Rahal et al. which showed that cephalosporin restriction in ICUs and hospitals gradually restored cephalosporin sensitivity at the cost of increased imipenem resistance in Pseudomonas aeruginosa. In addition to human factors, such as compensatory overuse, the mechanisms of drug response are also shared at the micro-organism level. Microbes use a limited number of available pathways to overcome drug sensitivity, such as drug efflux, drug
inactivation, target alteration, target protection and energy metabolism\(^\text{3}\). Most of the integrative analyses conducted thus far have focused on the genomics of resistance and cross-resistance\(^\text{23,27,28}\). The linkage of these mechanisms at an epidemiologic and genomic level was recently reported in a clinical trial collecting samples obtained from 12 community-based nursing homes across Michigan\(^\text{29}\). The study also conducted preliminary network analysis using the Cox proportional hazards model. In the present study, we reasoned that artificial intelligence and machine learning methods such as BNs are better suited to complex data with interactions, are more immune to unreasonable assumptions and have sufficiently developed heuristics (model averaging, scoring functions, search strategies). Our network structure captures the patterns in antimicrobial sensitivity that are expected across sites such as ceftazidime being the root of the network. This is reasonable because \textit{Pseudomonas} is the most common micro-organism affecting ICU patients across the world, a finding validated in our ICU (Figure 1a). Aggressive antibiotics, such as meropenem, imipenem, PT and cefotaxime, which are expected to be the drivers for selection pressure are found to be clustered near the root, whereas older-generation antibiotics, such as ampicillin, nitrofurantoin, chloramphenicol, cefixime and ceftriaxone, populate the terminal parts of the tree. This structure indicates that knowledge of resistance to these antibiotics does not change our belief about the resistance to the aggressive antibiotics, whereas knowledge of resistance to aggressive antibiotics indicates resistance to older-generation antibiotics.

Our model did show a temporal indication of meropenem and imipenem resistance, with the highest resistance rates in 2015, falling monotonically every year thereafter. Notably, these were the only two antibiotics that exhibited a direct temporal interaction. This may be indicative of restriction of carbapenem use even before the stewardship program was started towards the end of 2016. Our platform is not static and is designed for a learning stewardship system. To facilitate this we will document the continuous learnings from the network that have not been described in this paper in the development version on GitHub.

Hence, the BN model captures the microbiological and clinical factors involved in AMR. An obvious limitation of this study is the need for integration with patient phenotypes, diagnoses and disease evolution. The SAFE-ICU initiative\(^\text{30}\) has so far warehoused about 3,00,000 hours of continuous monitoring data from the pediatric ICU, along with treatment schedules and laboratory investigations. The future work on integrating the warehouse to create an artificial intelligence-enabled ICU is under development. Another limitation of the current analysis is that we have used BN as a reasoning tool rather than a causal inference tool. However, this work is a starting point for developing insights into the data-generative model and is expected to lead to better stewardship design through follow-up interventional models. Finally, we echo the sentiment of Mody et al.\(^\text{31}\), who state that the time is right for joint efforts in building integrated systems and shareable models that may be our only option for mitigating the global threat of AMR. AMR Dashboard is a local, yet a global-facing effort in this direction.

### Data availability

The raw and processed data associated with study are available on OSF: https://osf.io/57y98/\(^\text{32}\). Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).

### Software availability

**Software and source code available from:** https://github.com/SAFE-ICU/AMRsteward

** Archived source code at time of publication:** https://doi.org/10.5281/zenodo.1255584\(^\text{33}\).

**License:** GNU General Public License v3.

### Competing interests

No competing interests were disclosed.

### Grant information

This work was supported by the Wellcome Trust/DBT India Alliance Early Career grant IA/CPHE/14/1/501504 to Tavpritesh Sethi.

\textit{The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.}

### Acknowledgements

We acknowledge Dr. Nigam Shah for providing resources and support at the Biomedical Informatics Research Department, Stanford University School of Medicine. We also acknowledge Mr. Anant Mittal for his inputs on visualization and design of AMR dashboard.

### References

   Reference Source

   Reference Source


Open Peer Review

Sumeet Agarwal
Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India

This work addresses the very important issue of antimicrobial resistance in a large-scale data-driven fashion, using Bayesian network models to try and uncover drug-drug and drug-organism interactions in resistance. A useful dashboard to assist clinicians in antimicrobial stewardship is also developed, which will hopefully continue to get further refined as more data is gathered. There are however certain aspects of the methodology which are not entirely clear or appear to be insufficiently justified. I focus my remarks below on these.

1. "The final data consisted of date-time and nominal discrete variables...": It would be quite useful here to provide one or two example data points, so that the reader can clearly see the form of the data on which the Bayesian network models were trained.

2. "Structure learning and parametrization": Some rationale should be provided for the specific choice of 351 bootstraps. How sensitive is the learnt structure to this choice? Some sort of sensitivity analysis would help make such a choice more convincing.

For the exact structure learning algorithm used, some citations should be provided so that the reader can look up precise details.

"the imaginary sample size was kept low (three)": Again, it would be good to see some discussion of how varying this choice affects the results.

3. Figure 1b: This figure should include a colour bar so that the correspondence of colours to numerical values can be seen. Also, it appears that the ordering of drugs along the axes is top-to-bottom and right-to-left (not shown). Typically it is left-to-right, so this makes the plot a bit confusing to read initially. Any reason why it has been inverted here?

Also, how exactly are the correlation values depicted here computed? This does not seem to have been specified anywhere.

4. Figure 2 (and subsequently): It seems a bit odd to have a single 'Organism' node. Why not have separate nodes for each organism, just as there are for each drug? Wouldn't this allow us to see drug-organism interactions more clearly? And would also allow us to see organism-organism interactions, i.e., which organisms tend to co-occur and how that relates to their respective resistance patterns?
5. "the presence of a collider with two edges pointing towards a node represents competing hypotheses, and observing one can explain away the other": This doesn't sound quite correct. In such a situation, the competing hypotheses are marginally independent; they only become conditionally dependent when the node pointed to (the explanandum) has actually been observed. And when this happens, observing one hypothesis 'explains away' the explanandum (not 'the other' hypothesis). One could say that it makes the other hypothesis conditionally less likely, as a sufficient explanation for the explanandum is already available. So it would be good to re-word this sentence to clarify these aspects.

6. At various points the possibility of incorporating prior or expert knowledge in the Bayesian networks is mentioned, and also that "BNs allow continuous incorporation of prior knowledge". But it does not seem as if this has actually been attempted anywhere, or a specific methodology described for doing this. Some discussion of the details as to how this would happen, and whether such a possibility has been incorporated into the dashboard developed, would be helpful.

7. A key question in machine learning is always that of model validation and error analysis. These have not been attempted here in any systematic fashion. Some examples of observations/inferences from the learnt model have been provided, which seem to make sense. However, this is not really a sufficient basis for one to trust the model, or to be able to characterise its level of reliability. How many of the edges in the model are false positives? How many edges has it missed out on (false negatives)? How do these depend upon the hyperparameters used during training (point 2 above)? Some attempt to answer these sorts of questions would be expected of any such model. I realise this is hard, as no 'ground truth' is available here; but some systematic attempt at model validation would be helpful. For instance, it is standard to split the data into training and testing sets, and to try and validate the learnt model from the training set on the testing set to see how well it explains the latter. Something of that kind should also be included here, and would help quantify the accuracy/fidelity of the model much better. Such validation (or k-fold cross-validation, which is even better) is also helpful for selecting model hyperparameters, which at the moment (as mentioned in point 2 above) seem to have been determined arbitrarily.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Partly

Are sufficient details provided to allow replication of the method development and its use by others?
Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Partly

**Competing Interests:** No competing interests were disclosed.
Referee Expertise: Machine learning, computational/systems biology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.