Published in final edited form as:

IJSE Trans Healthc Syst Eng. 2017 ; 7: 121–128. doi:10.1080/24725579.2017.1339148.

Identifying Hospital Antimicrobial Resistance Targets via Robust Ranking

J. Paul Brooks, José H. Dulá, Amy L. Pakyz, and Ronald E. Polk

Virginia Commonwealth University, Richmond, VA, USA

Abstract

We develop a robust ranking procedure to uncover trends in variation in antibiotic resistance (AR) rates across hospitals for some antibiotic-bacterium pairs over several years. We illustrate how the method can be used to detect potentially dangerous trends and to direct attention to hospitals' management practices. A robust method is indicated due to the fact that some unusual reported resistance rates may be due to measurement protocol differences and not any real difference in AR rates. Our proposed method is less sensitive to outlier observations than other robust methods. The application on real AR data shows how a dangerous trend in a particular AR rate would be detected. Our results indicate the potential benefits of systematic AR rate collection and AR reporting systems across hospitals.

Keywords

hospital management; antimicrobial resistance; robust ranking; L1-norm line-fitting

2. Introduction

Antibiotic-resistant (AR) infections are life-threatening and costly, and currently there is no national surveillance system in place. The CDC estimates that 23,000 deaths and \$20 billion in health care costs per year are the result of related infections in hospitals (Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2013). The problem has been a concern in the health care field since the discovery of antibiotics. The Centers for Disease Control's National Healthcare Safety Network (NHSN, www.cdc.gov/nhsn) is just now beginning to implement a comprehensive surveillance program. Shulz et al. (2012) assert that advances are needed in understanding the relationship between drug use and resistance before the effects of interventions can be properly quantified.

We address the problem of discovering patterns in resistance rates that may be the result of hospital characteristics and management practices. The problem is addressed by introducing a robust ranking method based on the L1 norm that captures the variability of AR rates in hospital data while effectively handling outliers, and produces weights with revealing

patterns. Our data includes rates of resistance from hospital annual antibiograms for twelve antibiotic-bacterium (drug-bug) pairs for 51 hospitals for the years 2004 to 2008. The data contain outliers likely because uniform standards for producing antibiograms do not exist across hospitals.

Hospital antibiograms are rarely shared among hospitals and analyzed collectively. Notable exceptions are studies that are either focused on a single medical center or a single resistance rate. In a review by Shulz et al. (2012) of studies that analyzed antibiogram data, all but one of eight studies focused on trends in a single medical center, and the remaining study focused the effects on a single resistance rate across ICU units. Pakyz et al. (2012) analyze a single resistance rate from hospital antibiograms and its relationship to the prescription rates of several antibiotics. In our study, we simultaneously analyze multiple resistance rates from antibiograms from a large sample of hospitals.

Our aim was to develop a robust ranking procedure to facilitate the discovery of important trends in AR over time and to inform choices in hospital management. The method is based on the adaptation to higher dimensions of known approaches to fit lines in two dimensions using the L1 norm. Our analysis gives an indication of the potential insights and benefits that can be derived from systematic AR reporting.

3. Measuring Hospital-Level Antimicrobial Resistance

The hospital antibiogram is an annual summary of antimicrobial susceptibilities to bacteria most frequently isolated from patient samples submitted to the hospital's microbiology laboratory. Antibiograms are used by clinicians to assess local susceptibility rates as an aid in selecting empirical antibiotic therapy and in monitoring resistance trends over time within an institution. Antibiograms can also be used to compare susceptibility rates across institutions and track resistance trends.

VCU researchers requested antibiograms from 75 medical centers participating in the University Health Consortium (UHC, now part of Vizient, Inc.) for the years 2004 to 2008, and 51 provided an antibiogram for at least one year. Nearly all U.S. academic medical centers participate in the UHC, and our sample of 75 centers represents a significant portion of them.

We chose to analyze resistance rates for 12 antibiotic-bacterium pairs based on the virulence and transmissibility of infections and the availability of data from antibiograms. Table 1 contains the list of antibiotics and bacteria along with abbreviations that we will use through the remainder of the article.

Resistance rates are measured based on the number of *isolates* and not the number of patients with particular infections. Isolates are bacterial cultures obtained from patient samples in the hospital lab, and their resistance to different antibiotics is observed *in vitro*. Because more than one bacterium can be isolated from an infected patient, resistance rates are measured based on the number of isolates and not the number of patients with particular infections. A common treatment pattern is a trial-and-error approach with the initial prescription of a broadspectrum antibiotic, with changes to targeted antibiotics if the

infection does not respond. Due to the time between sample collection and generation of lab results, culture reports only influence choices in antibiotics after the first round of treatment. Further, if a particular antibiotic is effective, the number of susceptible isolates will decrease, causing an increase in the observed proportion of resistant isolates. Conversely, if an organism becomes resistant during therapy, then it may not be reflected in the antibiogram as only the first isolate is usually included, so that the observed resistance is less than the actual resistance. Therefore, the percentage of resistant isolates may not accurately reflect the true antibiotic-bacterium resistance rates. To overcome these biases, the proportion of resistant isolates is converted to the number of resistant isolates per 1,000 patient days (Schwaber et al., 2004).

We sought to facilitate the comparison of hospitals based on AR rates by creating a single summary measure. In a sample of hospitals with diverse hospital management and clinical practices, if the variation in resistance rates for an antibiotic-bacterium pair is low, then the data indicate that a change in practice is unlikely to reduce the infection rate for that antibiotic-bacterium pair. On the other hand, if the variation of resistance rates is high for a antibiotic-bacterium pair, then the variation may be a reflection of different practices, and the antibiotic-bacterium pair would be a target for understanding the differences between hospitals with high and low rates.

Figure 1 is a box plot of the 2008 AR data for twelve antibiotic-bacterium pairs. The whiskers for each resistance rate extend to the most extreme data point that is not more than 1.5 times the interquartile range from the box. We observe several outlier hospitals for all but two antibiotic-bacterium pairs (mrsa and vre faecium). Some of the outliers may be due to differences in data collection practices for antibiograms. As we will see, an analysis based on traditional PCA is unduly affected by the presence of these outliers. Therefore, we need a robust estimate that is insensitive to outliers and captures the patterns across most of the hospitals. The L1-norm is an outlier-resistant measure of error. These considerations are the motivation to develop a method for finding an estimate of the L1-norm best-fit line.

4. Finding Antimicrobial Resistance Loadings via L1-Line Fitting

Because our survey includes a diverse set of hospitals in terms of management and clinical practices, we can make inferences about the impact of choices by looking at variations in AR rates of antibiotic-bacterium pairs. Large variations in AR rates may indicate that certain choices can have an impact on antibiotic-bacterium pairs whereas low variation for a pair suggests that it is less affected by the range of currently-accepted practices. When these measurements are taken over time we can also detect trends about how resistance is progressing.

Our goal is to detect and measure the variation in AR rates across hospitals with different practices in hospital data containing outliers. Outliers may result from a lack of standardization in sample collection and processing that causes inflated values in resistance rates for some hospitals, or they may result from truly unusual resistance rates. Regardless of the origin, outliers can mask relationships among the majority of hospitals and hinder the ability to recover fundamental system properties.

One approach to measure each antibiotic-bacterium pair's variation relative to other pairs is standard principal component analysis (PCA). PCA generates a set of vectors – the principal component loadings – the first one of which is the direction with the most variation as measured by the variance of projected points onto a fitted line. Equivalently, this first principal vector defines geometrically a line that optimally fits the multidimensional data in the sense that the sum of the squared projection distances to the line are minimized (Jolliffe, 2002). It is therefore the “best-fit” line for the data according to the sum-of-squared error criterion. Furthermore, the loadings vectors contain a measure of the variability in the data in each dimension; the larger the value of a loading in the vector, the more variation there is in the corresponding coordinate. The loadings for the first principal component may be thought of as weights which can be used to rank the data points. This is done by adding the product of the weights with the points' values over all the dimensions. These inner products, one for each point, provide an ordering of the points. We will use this approach to rank hospitals.

Standard PCA is disproportionately and adversely affected by the presence of outliers. Therefore, there is a need for a robust method for calculating the first principal loading vector, or equivalently, the best-fit line through the data. Replacing the sum-of-squares measure associated with standard PCA with the L1 distance has been proposed to produce robust principal components (Ke and Kanade, 2003; Kwak, 2008; Nie et al., 2011; Candès et al., 2011; Brooks et al., 2013; Goldfarb et al., 2013; Markopoulos et al., 2014; Jiang et al., 2015; Visentin et al., 2016). Unfortunately, L1-norm PCA optimization problems have no known closed form solution method nor an efficient algorithm for calculating the first principal component for multi-dimensional data. One method for finding the weights for the first L1-norm principal component is to solve the L1-norm best-fit line problem directly. This is known to be NP-hard (Gillis and Vavasis, 2015) so the method we present next generates an estimate based on the solution to a series of weighted median problems.

Suppose we are given points $x_i \in \mathbb{R}^m$, $i = 1, \dots, n$. The L1-norm best-fit line through the origin can be found by solving

$$\min_{\substack{v \in \mathbb{R}^m, \alpha_i \in \mathbb{R}, \\ i = 1, \dots, n}} \sum_{i=1}^n \|x_i - v\alpha_i\|_1. \quad (1)$$

The vector v defines the line through the origin and for each point i , α_i indicates the location of the L1 projection of x_i on v . The problem is nonlinear, nonconvex, and nondifferentiable.

The problem above can be solved to optimality efficiently when data are in only two dimensions (Bloomfield and Steiger, 1980). That is because the problem reduces to finding the two weighted medians with respect to the angles the points make with each of the coordinate axes weighted by the magnitude of the points' corresponding components. The weighted median is the point in the “middle” of the data in terms of these angles, where the weights account for the distance from the origin. The required operations are simply calculating angles and sortings. The result is the line through the data with the property that

it goes through at least one of the points (Appa and Smith, 1973; Martini and Schöbel, 1998).

The heuristic we propose applies this result from the two-dimensional case. This idea is adapted to multidimensional data by fixing coordinate axes one at a time and measuring angles with respect to these fixed axes. The angles and magnitudes are calculated in each of the $m-1$ two-dimensional Cartesian planes that result from pairing the other axes with the one that is fixed. For each pair created with the fixed axis, the heuristic generates one of the components that will define a line in the full dimension of the data. This results in a line that is a composite of the $m-1$ individual two-dimensional best-fit lines. A total of m different lines (one for each fixed coordinate) that fit the data in the full space are generated and from these we select the one that performs best, that is, the one that has the smallest sum of the L1 errors.

Next we present the pseudo-code for this heuristic.

Algorithm 1

Heuristic for estimating an L1-norm best-fit line.

```

Given points  $x_i \in \mathbb{R}^m$  for  $i = 1, \dots, n$ .
1: Set  $z^* = \infty$ .
2: for ( $j = 1; j \leq m; j = j + 1$ )                                     /* Loop on fixed coordinate. */
3:   for ( $i = 1; i \leq n; i = i + 1$ )                                   /* Find each weight. */
4:     for ( $i = 1; i \leq n; i = i + 1$ )                                /* Calculate angles with fixed coordinate axis. */
5:       Set  $\theta_i = \begin{cases} \arctan \frac{x_{ij}}{x_{i\hat{j}}} & \text{if } x_{i\hat{j}} \neq 0, \\ 0 & \text{o.w.} \end{cases}$ 
6:     end for
7:     Find the weighted median  $\tilde{\theta}$  of  $\{\theta_i : i = 1, \dots, n\}$  with weights  $|x_{i\hat{j}}|$ .
8:     Set  $v_j = \tan \tilde{\theta}$ .                                           /* Set  $j^{\text{th}}$  weight. */
9:   end for
10:  Set  $z_{\hat{j}} = \sum_{i=1}^n \sum_{j=1}^m |x_{ij} - x_{i\hat{j}} v_j|$ .
11:  if  $z_{\hat{j}} < z^*$ ,
12:    Set  $z^* = z_{\hat{j}}$ ,  $v^* = v$ .                                       /* Update optimal solution. */
13:  end if
14: end for

```

Notes:

1. The heuristic fixes each coordinate in turn and for each of the other coordinates, calculates the angles the points make in those two axes. Each iteration of the middle loop generates a component of the vector v . After all components have

been generated, the vector v for the current fixed coordinate is fully defined and can be used to calculate the L1 error for the data. The vector from among all the choices of fixed coordinates that generates the smallest error is selected.

2. The procedure requires ratios, calculation of a trigonometric function, and sorting. A data set with n points and m dimensions requires $m(m-1)$ sortings of n angles each of which involves a ratio. This is a modest computational requirement.
3. In Step 5, the angle is set to zero if the datum is zero because although the angle is undefined, the weight in the weighted median step is zero.
4. The heuristic can generate the optimal solution for sets of points that project onto the optimal line fixing the same direction. Consider the following set of points

$$x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, x_3 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}.$$

The optimal solution is $v^* = (1,2,3)^T$, passing through point two, with objective function value $z^* = 2$. Points one and three project onto the optimal line by preserving the second coordinate. The global optimality of the solution is verified independently using Couenne (Belotti et al., 2009). In the next section we test the procedure against other approaches to fit lines using our hospital data. These results show that the heuristic performs well.

5. Table 2 demonstrates the heuristic using an example with the following seven points in three dimensions.

$$x_1 = \begin{pmatrix} 1.5 \\ 1.0 \\ 0.2 \end{pmatrix}, x_2 = \begin{pmatrix} 0.8 \\ 0.4 \\ 1.0 \end{pmatrix}, x_3 = \begin{pmatrix} 0.4 \\ -0.5 \\ 0.4 \end{pmatrix}, x_4 = \begin{pmatrix} 0.5 \\ 0.7 \\ -0.5 \end{pmatrix}, x_5 = \begin{pmatrix} 1.6 \\ -0.3 \\ 0.7 \end{pmatrix}, x_6 = \begin{pmatrix} 0.3 \\ 0.4 \\ -0.3 \end{pmatrix}, x_7 = \begin{pmatrix} 0.5 \\ -1.2 \\ -1.2 \end{pmatrix}.$$

The table includes the values at each iteration of the proposed algorithm. Since the data is in 3D, three separate estimates of the weights v are generated. Each of these has an objective function value that is the sum of the L1 distance from points to their projections. The best estimate is the vector with the smallest objective function value. This occurs with $v = (1,0.5,0.4)^T$. The objective function value for our estimate is within 1% of the solution obtained using Couenne.

A graphical demonstration is available in the supplementary file “11lineDemonstration.pdf” (an interactive version is available at <http://www.people.vcu.edu/~jpbrooks/11line/11lineDemonstration.html>). Notice how the two-dimensional lines fit at least one of the data points but the composite final line in three dimensions does not fit any of them. A graph of the objective function with our solution and the solution from Couenne is included.

The procedure to fit a line to multidimensional data developed here was applied to the hospital data. The next section reports on and interprets the results of this application. The section after that will analyze, test, and evaluate the procedure comparing to standard PCA and several other procedures involving the L1 norm. As we will see, the procedure introduced here is better than PCA and competitive with other procedures in terms of fitting the data and capturing the essential information contained in the non-outlier observations.

5. Discovering Antimicrobial Resistance Targets with L1-Norm Line Fitting

To prepare the data for analysis with our procedure the percentages of resistant isolates were first converted to rates of resistance per 1,000 patient days. Data that were missing are indications that the resistance rate was not measured in a hospital. Missing rates were replaced with the median resistance rate. The data was centered by subtracting the column-wise medians so that the data conforms to the specification that the best-fit line contains the origin.

Table 3 contains the best-fit line v for each year of data, 2004–2008 as generated by the heuristic. We can see that during the mid 2000s when high hospital infection rates and community-acquired mrsa were the focus of attention for hospitals, mrsa had the highest loading among the twelve antibiotic-bacterium pairs for the years 2004–2008. This means that the mrsa antibiotic-bacterium pair varied the most among the hospitals reflecting the intense activity on the part of hospitals in managing this infection with different rates of success. Recall from Figure 1 that mrsa is one of just two antibiotic-bacterium pairs without outliers for the 2008 data. The procedure was able to identify the high inherent variability in the data for this antibiotic-bacterium pair even though other pairs had outliers giving the appearance of exaggerated dispersion.

The changes in a loading through time can also provide valuable information. An illustration of this is the rise of the value of the loadings for fq ec. One interpretation of this is that while hospitals focused their energies on reducing mrsa rates this infection gradually became more conspicuous requiring increasing attention from the hospitals. This increase is corroborated by an independent survey of a different cohort of hospitals reported by The Center for Disease Dynamics, Economics and Policy. An online figure based on data collected by The Surveillance Network (The Center for Disease Dynamics, Economics & Policy, 2016) depicts the percentages of fq ec isolates, and the percentage more than doubles from 2004 to 2008. Indeed, fq ec is a major challenge for hospitals today (Johnson et al., 2013). This information obtained by looking at the change in variability is verified in other situations. For example, the loadings produced by the heuristic for pairs containing *P. aeruginosa* in Table 3 are consistently low from year to year and, indeed, the studies in (Johnson et al., 2013) confirm that antibiotic resistance rates did not change much in the 2004–2008 time frame.

The solution heuristic generates a vector which defines an estimate of the L1 best-fit line for the data. The components of this vector can be treated as weights and used to rank the hospitals. Rankings are a convenient way to compare complex entities such as hospitals. Hospitals are the object of many rankings with different purposes. One example, from

among many, of a ranking in the mass media is U.S. News & World Report's "America's Best Hospitals" (<http://health.usnews.com/best-hospitals>). Such rankings rely on "experts" to define the weights and attributes used to calculate the scores that define the position in the ranking. When experts are involved, the values of the weights reflect the relative importance given to the attributes by these experts and may prompt units' decision-makers to apply strategies and manage their operations to attain higher positions in the ranking (Pope, 2009).

Rankings for the 51 hospitals for the five years 2004–2008 using this criterion are in Table 4. A low number indicates lower resistance rates and therefore a better ranking. Hospital 13 consistently has the best ranking. This could be an indication that this hospital's management of resistance is superior for all rates, although it could also be a reflection of different measurement practices.

Recall that no hospital had resistance rates below the bottom whiskers in the box plot in Figure 1; all of the hospitals we identified as outliers had rates above at least one of the top whiskers. Additional interesting cases are hospitals 16 and 26. From 2005 to 2008, the ranking score improved relative to other hospitals. Hospitals 6, 8, 40, and 50 reveal patterns that reflect an apparent loss of control of resistance rates. In these cases, the rankings degrade over time. Lastly, observe the evolution of hospital 14. The relative rates dropped dramatically in 2005 and 2006, but increased again in 2007. The control of infections for this hospital should be reviewed closely.

It is possible to devise a strategy to improve the position in the ranking, even though AR rates are the result of complex bacterium-human-environment interactions, and are only partially controllable by choices in hospital management and/or clinical practice. Previous simulation studies have demonstrated that hospital characteristics and ICP policies can affect transmission and resistance rates (Hagtvedt et al., 2009; Barnes et al., 2010).

A result in (Bougnol and Dulá, 2013) states that, under certain assumptions, the optimal strategy to improve in a ranking is to cause the change in the value of a single attribute; specifically, the one with largest weight to cost ratio. This strategy can be implemented if the costs of changing the resistance rates for the antibiotic-bacterium pair are known. Let's imagine for a moment that all costs are the same. Then the optimal strategy would be to focus on affecting change on one specific pair which by our results would be mrsa in all five years since this is the attribute with the largest value in the loading vector. The result in Bougnol and Dulá (2013) requires an assumption that there are no synergies or antagonisms among the attributes. This theoretical optimal strategy may be what is observed in actual practices where the focus has been, almost exclusively in fighting mrsa resistance.

Antagonisms, however, are likely present in these situations; fighting to reduce resistance in one antibiotic-bacterium pair may result in neglect in some other pairs causing resistances to increase elsewhere. The data presented here seems to reflect this phenomenon in that while rates and weights for mrsa declined gradually over the years, the weights for fq ec increased. It is clear in all this that a sound and effective strategy to climb in this ranking would be to focus in reducing AR in an antibiotic-bacterium pair with a high weight to cost ratio while monitoring closely the impact on the other pairs and reacting appropriately when adverse effects are detected.

6. Performance of L1-Norm Line Fitting Heuristic

In this section we compare the ability of the proposed method for estimating the L1-norm best-fit line to standard PCA and other methods based on the L1 norm using the hospital-level AR data from 2008. If a hospital had at least one rate that falls outside the whiskers in Figure 1, we declared it an outlier. Eleven of the 51 hospitals were declared outliers by this criterion.

Our analysis is conducted in the R Environment for Statistical Computing (R Core Team, 2015). We compare our method to standard PCA in the function *prcomp()*.

In Figure 2 we demonstrate the performance of the proposed heuristic versus traditional PCA on a subset of data residing in two dimensions from the 2008 resistance data. The values for x3 kleb and x4 ps are plotted along with the fitted lines for traditional PCA and the proposed heuristic. Since the data reside in two dimensions, the heuristic provides a globally optimal solution Baccini et al. (1996); Brooks and Dulá (2013). In the figure, we see that traditional PCA is adversely affected by the outlier observations with high values for x3kleb. The proposed heuristic ignores the outliers and captures the variation in the data.

We used larger datasets to see if the benefits of the heuristic generalize and also compare our method to seven L1-norm PCA methods. One is implemented in the R package *pcaPP* (Filzmoser et al., 2009; Croux and Ruiz-Gazen, 2005) and is a method based on projection pursuit. The remaining methods are implemented in the R package *pcaL1* (Jot et al., 2017).

L1-PCA* (Brooks et al., 2013) is a method based on iteratively finding the L1-norm best-fit hyperplane, defining a new basis in the $k - 1$ dimensional space at iteration $m - k$, and repeating. L1-PCA (Ke and Kanade, 2003, 2005) is based on alternate optimization, fixing v and solving for a and fixing a and solving for v in (1). PCA-L1 (Kwak, 2008) is a method to find a line of maximum dispersion in the L1 sense. L1-PCAhp (Visentin et al., 2016) is a heuristic for estimating successive best-fit hyperplanes. wPCA and awPCA (Park and Klabjan, 2016) use iteratively re-weighted least squares algorithms to minimize the L1-norm distances to L2-norm projections.

Table 5 contains the scores - the results of applying the learned weights to each hospital - for the 2008 data. By comparing with Table 4, one can verify that the rankings for 2008 are based on these scores. By comparing the scores for the various methods, we see that different methods produce different rankings.

To evaluate the robustness of the various methods, consider the plots of the empirical cumulative distribution function of the residuals for non-outlier hospitals and all hospitals in Figure 3. Methods with smaller residuals will have curves at the top left of the graph. Our method is competitive with the other robust procedures and prevails for certain residuals for the non-outlier hospitals in Figure 3(a). In Figure 3(b), when the outliers are added, we see that the curves are slightly distorted on the right side of the plot. The hospitals with larger residuals, the outliers, are on the right side of the plot. Our proposed method has larger residuals for the most severe outlier observations while having smaller residuals for many of the non-outlier hospitals, demonstrating its robustness.

7. Conclusions

Currently, there is no national or regional data collection system concerning the types and amount of AR. Knowledge about AR rates may inform about effective ASPs, ICPs, and differences in patient populations.

In this study we have shown that important insights can be gained from hospital-level AR data from a group of academic medical centers. The practical implications of our research are that AR reporting, even with differences in measurement protocols, would have detected an important trend in fluorquinolone-resistant *E. coli* and that it could have informed choices in hospital management and clinical practice. Another contribution of this research is the development of a robust linefitting procedure based on the L_1 norm to overcome the effects of hospitals with unusual resistance rates and to capture trends across hospitals. Our method performs competitively with previously proposed methods and for some data points it is the best. We anticipate that our method would perform best compared to other methods when outliers are clustered leverage outliers, pulling models to particular directions in the space of the data. When outliers are sufficiently balanced, their influence is reduced and traditional methods will likely suffice. The proposed method has the added benefit that it is practical to implement.

Limitations of this study include the fact that we analyzed only hospital-level annual data and therefore we are not able to capture the dynamics of resistance on small time scales. Also, we do not have information on hospital management practices so we cannot match AR levels to particular management or policy decisions. A limitation of the method we developed is that it does not guarantee a global optimal solution; in fact, it does not even guarantee a local optimal solution. This quantitative study provides further evidence that AR surveillance systems can help to detect important trends in bacteria responsible for infectious diseases. Our results also indicate that measurement protocols should be standardized and data should be collected in a systematic fashion across hospitals. In future work, we plan to apply similar methods to uncover trends in data collected by the NHSN.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

8. Acknowledgements

We are grateful to reviewers who contributed generously to improving this work.

References

- Appa G and Smith C (1973). On L_1 and Chebyshev estimation. *Mathematical Programming*, 5:73–87.
- Baccini A, Besse P, and de Faguerolles A (1996). A L_1 -norm PCA and heuristic approach. In *Proceedings of the International Conference on Ordinal and Symbolic Data Analysis*, pages 359–368.
- Barnes S, Golden B, and Wasil E (2010). MRSA transmission reduction using agent-based modeling and simulation. *INFORMS Journal on Computing*, 22:635–646.
- Belotti P, Lee J, Liberti L, Margot F, and Wachter A (2009). “Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods and Software*, 24:597–634.

- Bloomfield P and Steiger W (1980). Least absolute deviations curve-fitting. *SIAM Journal on Scientific and Statistical Computing*, 1:290–301.
- Bougnol ML and Dulá JH (2013). A mathematical model to optimize decisions to impact multi-attribute rankings. *Scientometrics*, 95:785–796.
- Brooks JP and Dulá JH (2013). The L1-norm best-fit hyperplane problem. *Applied Mathematics Letters*, 26:51–55.
- Brooks JP, Dulá JH, and Boone EL (2013). A pure L1-norm principal component analysis. *Computational Statistics & Data Analysis*, 61:83–98. [PubMed: 23976807]
- Candès EJ, Li X, Ma Y, and Wright J (2011). Robust principal component analysis? *Journal of the ACM*, 58(3):11:1–11:37.
- Centers for Disease Control and Prevention, U.S. Department of Health and Human Services (2013). Antibiotic resistance threats in the United States. Technical report.
- Croux C and Ruiz-Gazen A (2005). High breakdown estimators for principal components: The projection-pursuit approach revisited. *Journal of Multivariate Analysis*, 95:206–226.
- Filzmoser P, Fritz H, and Kalcher K (2009). pcaPP: Robust PCA by projection pursuit.
- Gillis N and Vavasis SA (2015). On the complexity of robust PCA and L1-norm low-rank matrix approximation. *arXiv*, 1509.09236v2 [cs.LG].
- Goldfarb D, Ma S, and Scheinberg K (2013). Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, 141:349–382.
- Hagtvedt R, Griffin P, Keskinocak P, and Roberts R (2009). A simulation model to compare strategies for the reduction of health-care-associated infections. *Interfaces*, 39:256–270.
- Jiang F, Enqvist O, and Kahl F (2015). A combinatorial approach to L1-matrix factorization. *Journal of Mathematical Imaging and Vision*, 51:430–441.
- Johnson JR, Tchesnokova V, Johnston B, Clabots C, Roberts PL, Billig M, Riddell K, Rogers P, Qin X, Butler-Wu S, Price LB, Aziz M, Nicolas-Chanoine MH, DebRoy C, Robicsek A, Hansen G, Urban C, Platell J, Trott DJ, Zhanel G, Weissman SJ, Cookson BT, Fang FC, Limaye AP, Scholes D, Chattopadhyay S, Hooper DC, and Sokurenko EV (2013). Abrupt emergence of a single dominant multidrug-resistant strain of *Escherichia coli*. *The Journal of Infectious Diseases*, 207:919–928. [PubMed: 23288927]
- Jolliffe IT (2002). *Principal Component Analysis*. Springer, 2nd edition.
- Jot S, Brooks JP, Visentin A, Park YW, and Zhou YH (2017). pcaL1: L1-norm PCA methods.
- Ke Q and Kanade T (2003). Robust subspace computation using L1 norm. Technical report, Carnegie Mellon University, CMU-CS-03–172.
- Ke Q and Kanade T (2005). Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Kwak N (2008). Principal component analysis based on L1-norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:1672–1680. [PubMed: 18617723]
- Markopoulos PP, Karystinos KN, and Pados DA (2014). Optimal algorithms for L_1 -subspace signal processing. *IEEE Transactions on Signal Processing*, 62:5046–5058.
- Martini H and Schöbel A (1998). Median hyperplanes in normed spaces - a survey. *Discrete Applied Mathematics*, 89:181–195.
- Nie F, Huang H, Ding C, Luo D, and Wang H (2011). Robust principal component analysis with non-greedy l_1 -norm maximization. *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pages 1433–1438.
- Pakyz AL, Lee JA, Ababneh MA, Harpe SE, Oinonen MJ, and Polk RE (2012). Fluoroquinolone use and fluoroquinolone-resistant *pseudomonas aeruginosa* is declining in us academic medical centre hospitals. *Journal of Antimicrobial Chemotherapy*, 67:1562–1564. [PubMed: 22416053]
- Park YW and Klabjan D (2016). Iteratively reweighted least squares algorithms for L1-norm principal component analysis. *IEEE 16th International Conference on Data Mining (ICDM)*.
- Pope DG (2009). Reacting to rankings: Evidence from “America’s best hospitals”. *Journal of Health Economics*, 28:1154–1165. [PubMed: 19818518]

- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL <http://www.R-project.org>.
- Schwaber MJ, De-Medina T, and Carmeli Y (2004). Epidemiological interpretation of antibiotic resistance studies - what are we missing? *Nature Reviews Microbiology*, 2:979–983. [PubMed: 15550943]
- Shulz LT, Fox BC, and Polk RE (2012). Can the antibiogram be used to assess microbiologic outcomes after antimicrobial stewardship interventions? A critical review of the literature. *Pharmacotherapy*, 32:668–676. [PubMed: 23307516]
- The Center for Disease Dynamics, Economics & Policy (2016). http://www.cddep.org/projects/resistance_map/fluoroquinolone_resistant_escherichia_coli#field-widget-tab-1.
- Visentin A, Prestwich S, and Tarim SA (2016). Robust principal component analysis by reverse iterative linear programming. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 593–605.

Figure 1:

Box plot of resistance rates per 1,000 patient days for each antibiotic-bacterium combination for 2008. Table 1 contains a key for the abbreviations.

Figure 2:
Plot of x3 kleb and x4 ps resistance rates from 2008 along with fitted lines for traditional PCA and the proposed heuristic.

(a)

(b)

Figure 3:
Empirical cumulative distribution function plot of residuals for all hospitals for the 2008 data.

Table 1:

Table of Abbreviations for Antibiotic-Bacterium (Drug-Bug) Pairs

Abbreviation	Antibiotic	Bacterium
x3 ec	3rd generation cephalosporin-resistant	<i>Escherichia coli</i>
x3 kleb	3rd generation cephalosporin-resistant	<i>Klebsiella pneumoniae</i>
x4 ps	4th generation cephalosporin-resistant	<i>Pseudomonas aeruginosa</i>
carb ab	carbapenem-resistant	<i>Acinetobacter baumannii</i>
carb kleb	carbapenem-resistant	<i>Klebsiella pneumoniae</i>
carb ps	carbapenem-resistant	<i>Pseudomonas aeruginosa</i>
fq ec	fluoroquinolone-resistant	<i>Escherichia coli</i>
fq ps	fluoroquinolone-resistant	<i>Pseudomonas aeruginosa</i>
pt ps	piperacillin/tazobactam-resistant	<i>Pseudomonas aeruginosa</i>
mrsa	methicillin-resistant	<i>Staphylococcus aureus</i>
vre faecalis	vancomycin-resistant	<i>Enterococcus faecalis</i>
vre faecium	vancomycin-resistant	<i>Enterococcus faecium</i>

Table 2:

Calculations for Heuristic for Example in Three-Dimensions

Outer Loop	Inner Loop	Angles with $\mathbf{f}$ -axis	Weights	Objective Function Value
$\hat{j}$	j	θ	v_j	z_j
1	1	$\left(\frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}\right)$	1.0	8.1
	2	(0.59, 0.46, -0.90, 0.95, -0.19, 0.93, -1.18)	0.5	
	3	(0.13, 0.90, 0.79, -0.79, 0.41, -0.79, -1.18)	0.4	
2	1	(0.98, 1.11, -0.67, 0.62, -1.39, 0.64, -0.39)	0.7	9.4
	2	$\left(\frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}\right)$	1.0	
	3	(0.20, 1.19, -0.67, -0.62, -1.17, -0.64, 0.79)	0.2	
3	1	(1.44, 0.67, 0.79, -0.79, 1.16, -0.79, -0.39)	0.8	9.7
	2	(1.37, 0.38, -0.90, -0.95, -0.40, -0.93, 0.79)	0.4	
	3	$\left(\frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}, \frac{\pi}{4}\right)$	1.0	

Table 3:

Loadings for 2004–2008 Antimicrobial Resistance Data

Drug-Bug*	v2004	v2005	v2006	v2007	v2008
x3 ec	0.02	0.06	0.05	0.03	0.01
x3 kleb	0.03	0.03	0.02	0.01	0.02
x4 ps	0.00	0.00	0.10	0.00	0.08
carb ab	0.02	0.00	0.00	0.00	0.02
carb kleb	0.00	0.00	0.00	0.00	0.00
carb ps	0.03	0.00	0.13	0.00	0.08
fq ec	0.19	0.21	0.22	0.29	0.38
fq ps	0.08	0.10	0.20	0.01	0.16
pt ps	0.00	0.04	0.05	0.00	0.06
mrsa	0.98	0.97	0.94	0.96	0.90
vre faecalis	0.00	0.00	0.00	0.00	0.00
vre faecium	0.02	0.00	0.00	0.00	0.00

* Table 1 contains a key for antibiotic-bacterium abbreviations.

Table 4:

Rankings of Hospitals using Loadings as Weights for 2004–2008 Antimicrobial Resistance Data

Hospital		Year				
Index	2004	2005	2006	2007	2008	
1	44	44	44	48	41	
2	9	10	45	43	40	
3	18	10	33	18	20	
4	11	40	35	40	38	
5	15	42	32	15	16	
6	12	10	41	44	42	
7	16	51	50	49	18	
8	6	10	17	37	36	
9	42	7	39	45	44	
10	49	41	31	51	33	
11	18	10	19	18	20	
12	43	47	48	47	39	
13	1	1	1	1	1	
14	10	3	7	13	14	
15	47	45	38	38	37	
16	14	10	2	2	2	
17	18	10	12	7	19	
18	8	4	16	18	13	
19	17	10	19	18	20	
20	45	43	40	18	20	
21	50	50	30	12	51	
22	37	46	43	41	34	
23	13	10	19	18	20	
24	38	10	19	18	45	
25	3	2	5	16	8	
26	40	48	18	11	11	
27	48	10	19	18	20	
28	51	10	47	39	48	
29	4	6	13	14	31	
30	41	49	49	50	50	
31	2	10	19	18	20	
32	46	8	36	42	49	
33	39	10	19	18	20	
34	7	9	9	17	15	
35	5	5	11	9	7	
36	18	10	42	8	4	
37	18	10	34	35	32	
38	18	10	15	5	5	

Hospital Index	Year				
	2004	2005	2006	2007	2008
39	18	10	8	18	20
40	18	10	14	36	35
41	18	10	10	10	10
42	18	10	46	34	12
43	18	10	6	6	9
44	18	10	19	18	47
45	18	10	4	4	6
46	18	10	51	18	20
47	18	10	37	33	17
48	18	10	3	3	3
49	18	10	19	18	30
50	18	10	19	46	46
51	18	10	19	18	43

Table 5:

Scores of Hospitals using Loadings as Weights for 2008 Antimicrobial Resistance Data

Hospital	Heuristic	L1-PCA*	L1-PCA	PCA	pcaPP	PCA-L1	L1-PCAhp	wPCA	awPCA
21	-7.2	-7.8	-7.9	-8.1	-7.7	-8.0	-7.7	-7.6	-7.6
30	-4.9	-4.3	-4.0	-4.2	-3.8	-4.3	-4.5	-4.4	-4.4
32	-4.8	-4.1	-3.9	-3.7	-3.7	-4.0	-4.3	-4.3	-4.3
28	-4.3	-4.9	-4.9	-5.3	-4.7	-5.2	-4.8	-4.6	-4.6
44	-3.8	-3.3	-3.1	-3.1	-2.7	-3.2	-3.4	-3.4	-3.4
50	-3.8	-3.5	-3.4	-3.3	-3.3	-3.4	-3.6	-3.6	-3.6
24	-3.6	-5.5	-5.9	-6.0	-6.3	-5.6	-5.1	-5.2	-5.3
9	-3.3	-3.2	-3.0	-2.9	-3.0	-3.0	-3.2	-3.3	-3.3
51	-2.9	-3.2	-3.2	-3.1	-3.3	-3.1	-3.1	-3.2	-3.2
6	-2.7	-2.3	-2.1	-2.1	-2.1	-2.2	-2.4	-2.4	-2.4
1	-2.7	-3.9	-4.4	-4.5	-3.8	-4.2	-3.6	-3.5	-3.5
2	-2.5	-2.3	-2.3	-2.4	-2.0	-2.5	-2.4	-2.3	-2.3
12	-1.9	-2.1	-2.0	-2.2	-2.0	-2.2	-2.1	-2.0	-2.0
4	-1.8	-1.8	-1.5	-1.7	-1.8	-1.8	-1.9	-1.8	-1.8
15	-1.5	-2.3	-2.6	-2.6	-2.3	-2.4	-2.1	-2.1	-2.1
8	-1.4	-1.3	-1.2	-1.3	-1.1	-1.4	-1.4	-1.3	-1.3
40	-0.9	-1.4	-1.5	-1.5	-1.5	-1.4	-1.3	-1.3	-1.3
22	-0.9	-0.7	-0.7	-0.7	-0.5	-0.9	-0.9	-0.8	-0.8
10	-0.8	-1.3	-1.5	-1.5	-1.4	-1.4	-1.2	-1.2	-1.2
37	-0.4	-0.6	-0.7	-0.7	-0.7	-0.7	-0.6	-0.6	-0.6
29	-0.3	-0.1	-0.0	0.0	0.0	-0.0	-0.1	-0.1	-0.1
49	-0.3	-0.5	-0.5	-0.5	-0.6	-0.5	-0.4	-0.5	-0.5
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
39	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	0.1	-0.2	-0.4	-0.4	-0.2	-0.3	-0.1	-0.1	-0.1
7	0.1	0.2	0.2	0.3	0.1	0.3	0.2	0.1	0.1
47	0.5	0.9	1.0	1.0	1.1	0.9	0.8	0.8	0.8
5	0.5	0.2	0.2	0.0	0.2	0.1	0.3	0.3	0.3
34	0.8	0.6	0.6	0.5	0.5	0.5	0.6	0.7	0.7
14	1.0	1.2	1.3	1.2	1.2	1.2	1.2	1.2	1.2
18	1.1	1.6	1.8	1.8	1.7	1.7	1.5	1.5	1.5

Hospital	Heuristic	L1-PCA*	L1-PCA	PCA	pcaPP	PCA-L1	L1-PCAh _p	wPCA	awPCA
42	1.2	1.1	1.1	1.0	1.1	1.0	1.1	1.1	1.1
26	1.2	1.2	1.2	1.2	1.1	1.2	1.2	1.2	1.2
41	1.5	2.0	2.1	2.1	2.1	2.1	1.9	1.9	1.9
43	2.2	1.9	1.7	1.7	1.7	1.8	2.0	2.0	2.0
25	2.5	2.3	2.2	2.2	2.2	2.3	2.4	2.4	2.4
35	2.6	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
45	2.9	3.0	2.9	2.9	3.0	2.9	3.0	3.1	3.1
38	3.1	3.0	2.9	2.9	2.9	2.9	3.0	3.1	3.1
36	3.3	3.4	3.4	3.4	3.3	3.4	3.4	3.4	3.4
48	3.6	3.7	3.7	3.6	3.6	3.6	3.7	3.7	3.7
16	4.1	4.1	4.0	3.9	3.9	4.0	4.1	4.1	4.1
13	4.5	4.7	4.7	4.6	4.7	4.7	4.7	4.7	4.7