Phenome-Based Analysis as a Means for Discovering Context-Dependent Clinical Reference Ranges

Jeremy L. Warner, MD, MS[1,2], Gil Alterovitz, PhD[3]

[1]Division of Hematology/Oncology, Vanderbilt University, Nashville, TN; [2]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; [3]Center for Biomedical Informatics, Harvard Medical School, Boston, MA

Abstract

Robust electronic medical records (EMR's) have made large-scale phenome-based analysis feasible. The context-dependent phenome of a large ICU-based EMR database (MIMIC II) was explored, as a function of a clinical feature: white blood cell count (WBC). Phenome visualization led to the discovery that peak WBC in the range 15-45 K/μl was highly associated with the diagnoses of Clostridium difficile and bacterial sepsis; thus, it is conceivable that clinicians might delay ordering targeted antimicrobials towards C. difficile for patients with peak WBC in this range. This hypothesis was confirmed, with significant delays in this group (median 135 vs. 85 hours, p = 0.002). These delays could be associated with adverse effects on patient health and high hospitalization costs (e.g. an additional $3,000,000 for the MIMIC II cohort). In conclusion, context-dependent clinical reference ranges are critical to clinical decision making; furthermore, important findings can be discovered through EMR-driven phenome association studies.

Introduction and Background

Phenomics, which is the mapping of molecular or clinical features to an explicitly defined phenotypic "universe," is a promising application of clinical informatics to large datasets[1]. The increasing availability of medical information in electronic format has made it possible to evaluate relationships, associations, and correlations which were previously obscure. For example, the association of celecoxib with excess cardiac toxicity was discovered in part through large dataset evaluation[2]. Linking molecular data such as gene expression and clinical data such as outcome and phenotype is not just possible but is increasingly practicable[3-5]. In this paper, we describe a new method for analyzing phenomic associations using continuously varying clinical features, in the contextual element of critical illness.

Most previous phenomics work has focused on features that can be defined as binary or dichotomous variables, such as single-nucleotide polymorphisms (SNPs)[6, 7]. However, many clinical features have a wide dynamic range, as do molecular measures such as quantitative gene expression. It is possible that significant information from these features can be lost in the process of dichotomization, even with optimal cut-point selection. Therefore, we propose a novel method for the visualization of phenomic associations across such features. This method is then illustrated using a contextual clinical use case. Specifically, we investigated how the phenotypic spectrum of a common clinical lab test, the white blood cell (WBC) count, might be affected by the context of critical illness. Subsequently, a hypothesis generated through this use case was investigated by examining the timing of initiation of directed antimicrobial therapy.

Methods

Use Case: Highly elevated WBC counts (leukocytosis) are traditionally known to be associated with malignant leukemias and with situations of severe systemic inflammation, such as septic shock. *Clostridium difficile*, an infection usually confined to the large bowel, is also known to induce leukocytosis, often before the clinical appearance of diarrhea[8]. Given that leukocytosis often accompanies critical illness, we evaluated the phenotypic spectrum of a well-characterized cohort of critically ill patients by their maximum observed (peak) WBC count during their hospitalization.

We then examined the timing between hospital admission and the initiation of antimicrobial therapy directed against *C. difficile*, as a function of WBC count, for those patients who were assigned the International Classification of

Diseases, Clinical Modification (ICD-9-CM) code for intestinal infection caused by *C. difficile* (008.45). It is conventional to treat critically ill patients with empiric antibiotics, usually those directed against gram negative bacilli and methicillin-resistant *Staphylococcus aureus* (MRSA). However, very few antibiotics (metronidazole administered orally [PO] or intravenously [IV], or vancomycin administered PO or rectally [PR]) adequately treat *C. difficile*, whereas other antibiotics can actually worsen the infection[9]. Therefore, increased time between admission and directed treatment for *C. difficile* increases the risk of exposure to incorrect and/or deleterious antibiotics, in addition to increasing the cost of hospitalization.

Data Source: MIMIC II, a comprehensive electronic medical record (EMR) database of over 30,000 critically ill patients admitted to Beth Israel Deaconess Medical Center between 2001 and 2007, was used as the primary data source[10]. All patients in the MIMIC II database spent at least some portion of their hospitalization in the intensive care unit. The ICD-9-CM codes recorded at the time of hospital discharge were used to define patient phenotypes. In MIMIC II, the sequence of ICD-9-CM codes is also recorded: primary, secondary, and so forth. All investigators completed appropriate human subjects training prior to accessing MIMIC II data.

Clinical Feature Phenome Map: The base phenomic spectrum was defined as the cumulative distribution of all ICD-9-CM codes from all admissions recorded in the MIMIC II database, where the WBC count was measured at least once.

Next, the WBC count was divided into one hundred equally spaced segments spanning from 0 K/μl to 100 K/μl. Each cutoff defined by this segmentation was used as a lower bound (e.g. ≥ 50 K/μl) to define a subset of the patient population. Admissions where the WBC count measurement exceeded the lower bound on at least one occasion were included in the subset, and the ICD-9-CM codes recorded from these admissions comprised a subset of the base phenomic spectrum.

A "phenome map" was created by calculating p-values on the phenome-wide association for each subset defined by the 100 discrete cutoffs, and each of these calculations was displayed as a horizontal slice of a two-dimensional graph. The horizontal axis of this graph comprises the ICD-9-CM codes, divided into separate chapters by color, with V codes and E codes on the far right. The vertical axis corresponds to the cutoffs. Each significant p-value, as defined below, was recorded as a point on the map, with the size of the point proportionate to the negative logarithm of the p-value.

Statistical Analysis: For each ICD-9-CM code with a non-zero entry, the exact binomial test was performed, comparing the observed frequency of the code occurrence in the subset to the expected frequency in the base phenomic spectrum. In order to reject the null hypothesis, the incidence of the ICD-9-CM code in the subset had to be greater than the incidence in the base set (one-sided alternative hypothesis). Only p-values with significance less than 0.05 divided by the total number of non-zero ICD-9-CM codes in the subset (the Bonferroni correction) were considered to be significant[11]. A sensitivity analysis was carried out by repeating the analysis using only the primary ICD-9-CM code, only the primary and secondary ICD-9-CM codes, and so forth.

For admissions associated with ICD-9-CM code 008.45, the time to initiation of targeted therapy for *C. difficile* was defined as the difference, in hours, between the time of admission and the timestamp of first order entry for either 1) PO or IV metronidazole; or 2) PO or PR vancomycin. For those admissions with ICD-9-CM code 008.45 which did not have any of these antibiotics ordered, the time interval was defined as the difference between the time of admission and the time of discharge. All admissions and discharges in the MIMIC II dataset are recorded as occurring at 00:00 hours of the day of admission or discharge. The group of patients with *C. difficile* with peak WBC count between 15-45 K/μl was compared to the group with peak WBC count < 15 K/μl or ≥ 45 K/μl using the two-sided Mann-Whitney U test, with significance defined as p-value < 0.05[11].

Results

The base characteristics of the total MIMIC II population, as well as a subpopulation meeting the criteria of at least one WBC count $\geq$ 50 K/µl during their hospitalization are shown in Table 1. The phenome-wide association of this subpopulation compared to the total population is shown in Figure 1, with the ten most significant ICD-9-CM codes labeled. At this particular cutoff, 26 ICD-9-CM codes were significantly more likely to occur in the subpopulation (p-value < 5.7 x 10^{-5}). The ten most significant codes are shown in Table 2, with associated p-values.

Table 1. Baseline demographics of the MIMIC II dataset, and an example of a subset defined by WBC $\geq$ 50 K/µl.

Data from MIMIC-II v6	Number of Instances	Comments
Total Admissions	36,095	
Distinct ICD-9-CM Codes	5,675	33.8% of all possible codes
ICD-9-CM Instances	314,663	mean of 8.7 per admission
WBC Instances	506,659	
WBC $\geq$ 50 K/µl Instances	1,462	0.29% of all WBC instances
Unique Admits with WBC $\geq$ 50 K/µl	270	0.75% of all admissions
Distinct ICD-9-CM Codes of Subset	875	5.1% of all possible codes
ICD-9-CM Instances of Subset	3,351	mean of 12.4 per admission

Table 2. Ten most significant ICD-9-CM codes, when at least one WBC count >= 50 K/µl during the course of hospitalization, in the MIMIC II cohort.

ICD-9-CM Code	ICD-9-CM Description	p-value
204.10	Chronic lymphoid leukemia (CLL)	4.3 x 10^{-33}
205.00	Acute myeloid leukemia (AML)	4.0 x 10^{-29}
995.92	Severe sepsis	5.3 x 10^{-20}
286.6	Disseminated intravascular coagulation (DIC)	3.7 x 10^{-16}
785.52	Septic shock	7.3 x 10^{-16}
205.10	Chronic myeloid leukemia (CML)	7.7 x 10^{-16}
008.45	Intestinal infection due to *Clostridium difficile*	1.7 x 10^{-15}
038.3	Septicemia due to anaerobes	6.0 x 10^{-11}
038.9	Unspecified septicemia	7.2 x 10^{-11}
238.70	Neoplasm of uncertain behavior of other lymphatic and hematopoietic tissues	3.4 x 10^{-10}
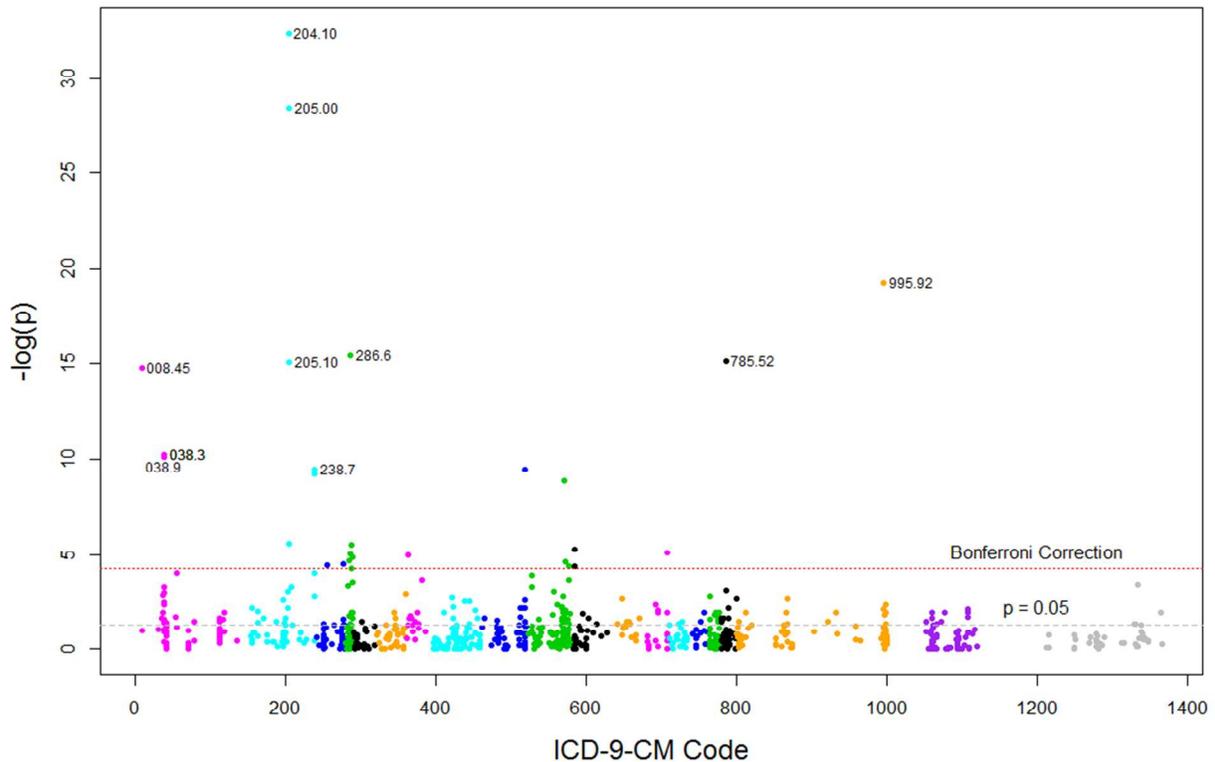
Figure 1. Phenome-wide association of ICD-9-CM codes in the subpopulation where WBC ≥ 50 K/µl on at least one occasion during a hospitalization. The ten most significant are labeled, and are also shown in Table 2. Each chapter of the ICD-9-CM coding schema is shown in a separate color, with V- and E- codes shown in purple and gray, on the right.

The phenome map obtained when the lower-bounded WBC cutoff was varied over the range from ≥ 0 K/µl to ≥ 100 K/µl is shown in Figure 2. Significant associations were not observed until the WBC count was ≥ 10 K/µl, except for ICD-9-CM code V30.00 (single liveborn, born in hospital, delivered without mention of cesarean section), which was significant at ≥ 9 K/µl. As expected, the codes for the various leukemias were significant over a wide range of elevated WBC counts. For example, 204.10 (chronic lymphocytic leukemia, CLL) was significant from ≥ 16 to ≥ 100 K/µl, and 205.00 (acute myeloid leukemia, AML) was significant from ≥ 22 to ≥ 100 K/µl. Multiple codes for severe infections and septic shock were significant. For example, 038.9 (unspecified septicemia) was significant from ≥ 11 to ≥ 68 K/µl, and 008.45 (intestinal infection due to *C. difficile*) was significant from ≥ 12 to ≥ 88 K/µl.

2-D Phenome Map, Showing Significant Diagnoses
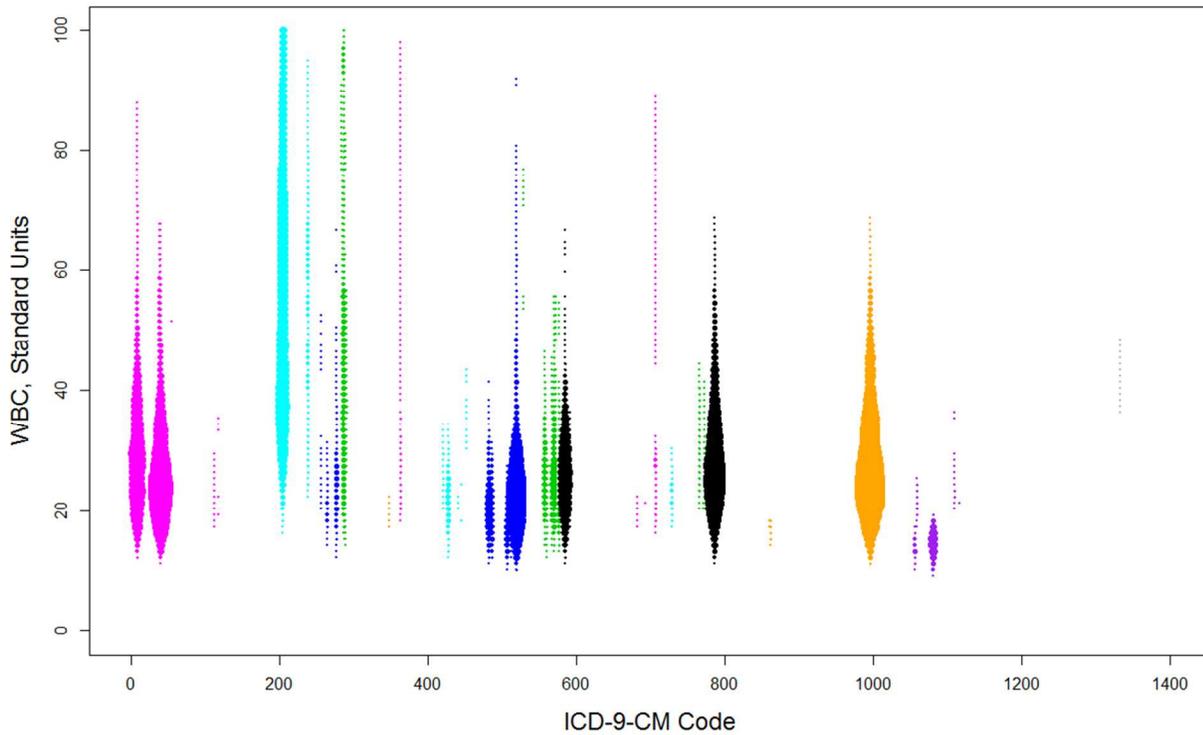
Figure 2. Phenome map as a function of peak WBC. ICD-9-CM codes for leukemias (200-208 range) are significant from ~20 K/μl to the top of the range ($\geq$ 100 K/μl). *Clostridium difficile* (008.45), sepsis (038.x), septic shock (785.5x), and SIRS (995.90) are all highly significant from approximately $\geq$ 15k to $\geq$ 45k. Each chapter of the ICD-9-CM coding schema is shown in a separate color, with V- and E- codes shown in purple and gray, on the right.

Sensitivity Analysis: In order to determine whether the findings were robust to omission of ICD-9-CM codes, we repeated the analysis with varying exclusions of non-primary ICD-9-CM codes. As shown in Figure 3, when only the primary ICD-9-CM was used, the number of significant associations was considerably lower. The phenome map becomes more informative with the addition of the secondary ICD-9-CM code (data not shown).

2-D Phenome Map, Showing Significant Diagnoses
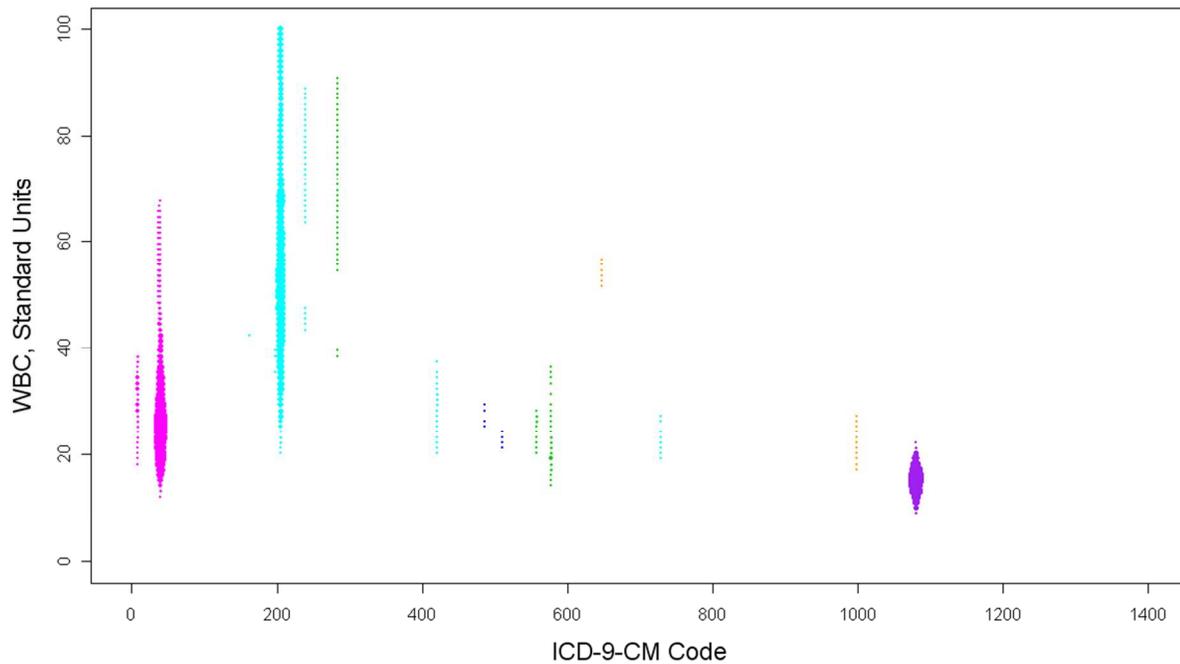
Figure 3. Sensitivity analysis, limited to primary ICD-9-CM code only. Most of the fidelity of the phenome map is lost, with only a few diagnostic codes retaining significance over the range.

Timing to Directed Antimicrobial Therapy: 723 admissions in the MIMIC II cohort were assigned an ICD-9-CM code of 008.45. Of these, 74.0% had orders for one or more of the specified antibiotics by one of the specified routes (Table 3). 67.4% of the admissions (N = 487) were associated with a peak WBC count in the 15 to 45 K/μl range, and 32.6% with a peak WBC count < 15 or ≥ 45 K/μl. There was no statistical difference in the rate of ordering of specified antibiotics in the two groups (74.2% vs. 73.9%, p = 1). The median time to ordering of one of the specified antibiotics was significantly longer in the 15 to 45 K/μl group than the < 15 or ≥ 45 K/μl group: 135 hours (interquartile range, 23 to 232 hours) vs. 85 hours (interquartile range, 22 to 194.8 hours), p = 0.002.

Table 3. Antimicrobial ordering patterns for the admissions assigned ICD-9-CM code 008.45, intestinal infection due to *C. difficile*.

Antibiotic	# of cases	% of total (N = 723)
PO metronidazole	373	51.6%
IV metronidazole	266	36.8%
PO vancomycin	247	34.2%
PR vancomycin	29	4.0%
Any of the above	535	74.0%

Conclusion and Discussion

In this paper, we have demonstrated a new method for phenome visualization over continuously varying clinical features, using peak WBC as a use case. We found that, in the MIMIC II cohort, diagnostic codes did not achieve statistical significance until peak WBC counts exceeded 10 K/µl (except for V30.00), similar to the conventional definition of leukocytosis[12]. We also found that certain diagnosis codes, such as septic shock, are significant over a defined range of WBC values, suggesting that reference ranges above the upper limit of normal can define important subpopulations, in the context of critical illness. This finding led to the generation of a hypothesis that critically ill patients who were likely to be diagnosed with intestinal infection due to *C. difficile* might have a delay in initiation of appropriate antimicrobial therapy. This hypothesis was based on the overlap of the elevated WBC range associated with *C. difficile* with that associated with sepsis and septic shock, as well as the clinical observation that leukocytosis can precede diarrhea by several days in *C. difficile* infection[8]. This hypothesis was confirmed, with an average delay of 50 hours to the initiation of directed antimicrobial therapy against *C. difficile* in patients with a peak WBC count in the range of 15 to 45 K/µl. This finding suggests that these patients may be at increased risk for adverse effects, both from a potential delay in effective treatment as well as from possibly increased time spent in the hospital, with concomitantly increased risks of iatrogenic infection and other complications. This finding also has implications for resource utilization. Given that 24 hours in the intensive care unit costs in excess of $3,000[13], this delay could represent costs in excess of $3,000,000 in the MIMIC II cohort alone.

There are several limitations to this study. Although our database was large and comprehensive, it is from a single institutional cohort and may therefore have limited generalizability. This cohort also includes both adult and neonatal patients, as evidenced by the fact that ICD-9-CM code V30.00 (single liveborn, born in hospital, delivered without mention of cesarean section) was significant over the range 9 to 19 K/µl - in keeping with the higher normal WBC range (10-30 K/µl) for infants[12]. Interestingly, this suggests that V30.00 serves as a useful "internal positive control" in the case of peak WBC.

We chose the criteria of peak WBC count for relative simplicity; other, more complicated criteria could easily be chosen and could lead to other phenotypic sub-spectrums. For example, a patient with a high peak WBC count followed by a very low nadir WBC count is likely to have presented with an acute leukemia followed by marrow ablation as a result of induction chemotherapy. Conversely, a patient with a very low nadir WBC count followed by a high peak WBC count is likely to have presented with agranulocytosis which responded to granulocyte-colony stimulating factor (G-CSF). These and other complex patterns of continuous variables, which can vary over time, will be the subject of future investigation.

Additionally, our choice of ICD-9-CM code as phenotypic definition bears further discussion. Billing data, such as: ICD-9-CM; diagnosis-related group (DRG) codes; and current procedural terminology (CPT) codes, are readily available as structured elements in EMR data. For this reason, they are commonly used to define the phenotypic "universe" of interest for phenome-based analysis. Because of the complexity of critical illness, most discharges involve at least five and as many as thirty distinct ICD-9-CM codes, creating a potentially rich phenotype. Phenotypic definitions can also be enriched through the combination of billing data with clinical structured elements and with information derived from natural language processing (NLP) or other methods[14, 15]. ICD-9-CM codes, in particular, are known to be recorded with variable accuracy; the presence of an internal positive control (V30.00) in our analysis is reassuring. However, ICD-9-CM codes are recorded only once, at the end of a hospitalization, and generally based on information available in a discharge summary. As such they are subject to the recency effect, suggesting that important diagnostic events occurring early in a prolonged admission may not be captured. It is unclear whether our findings would be improved through the addition of further phenotypic features such as those derived from NLP, which are still primarily research tools in development[16].

We observed a loss of significance for many ICD-9-CM codes as the cutoff became increasingly high. This could be due to a true loss of association or could be a result of under-powering, as the subset becomes increasingly small. However, it is notable that despite a loss of power with high cutoffs, codes representing the leukemias remained significant throughout the range.

Finally, the decision to use the Bonferroni correction to determine significance was based on generally accepted criteria for genome-wide association studies, to correct for multiple hypotheses. However, others have pointed out that the Bonferroni correction may be too rigorous[6, 17, 18]. As a result, it is possible that clinically significant associations were not displayed on the phenome map.

Future work will focus on automated feature selection using Bayesian methods, combination of clinical features into feature sets, and integration of continuously variable molecular data such as quantitative gene expression.

In conclusion, visualization of complex phenomic data is a powerful tool for knowledge discovery in large context-dependent EMR databases with robust phenome information. We have shown that the creation of a "phenome map" for a continuously variable clinical feature is feasible and have also demonstrated that a hypothesis can be generated and evaluated by using this visual information. This may represent a paradigm shift in the way that clinical reference ranges are understood, in the context of critical illness.

References

1. Houle D, Govindaraju DR, Omholt S. Phenomics: The next challenge. Nat Rev Genet. 2010;11(12):855-66.
2. Brownstein JS, Sordo M, Kohane IS, Mandl KD. The tell-tale heart: Population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction. PLoS One. 2007;2(9):e840.
3. Butte AJ, Kohane IS. Creation and implications of a phenome-genome network. Nat Biotechnol. 2006;24(1):55-62.
4. Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011;12(6):417-28.
5. Murphy S, Churchill S, Bry L, Chueh H, Weiss S, Lazarus R, et al. Instrumenting the health care enterprise for discovery research in the genomic era. Genome Res. 2009;19(9):1675-81.
6. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010;26(9):1205-10.
7. Pendergrass SA, Brown-Gentry K, Dudek SM, Torstenson ES, Ambite JL, Avery CL, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. Genet Epidemiol. 2011;35(5):410-22.
8. Bulusu M, Narayan S, Shetler K, Triadafilopoulos G. Leukocytosis as a harbinger and surrogate marker of Clostridium difficile infection in hospitalized patients with diarrhea. Am J Gastroenterol. 2000;95(11):3137-41.
9. Leffler DA, Lamont JT. Treatment of Clostridium difficile-associated disease. Gastroenterology. 2009;136(6):1899-912.
10. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, et al. Multiparameter Intelligent Monitoring in Intensive Care II: A public-access intensive care unit database. Crit Care Med. 2011;39(5):952-60.
11. Dawson B, Trapp RG. Basic & Clinical Biostatistics. 4th ed. New York: Lange Medical Books/McGraw-Hill, Medical Pub. Division; 2004.
12. Alberts B. Molecular Biology of the Cell 2005. Available from: http://www.ncbi.nlm.nih.gov/books/bv.fcgi?highlight=leukocyte,functions&rid=mboc4.table.4143.
13. Dasta JF, McLaughlin TP, Mody SH, Piech CT. Daily cost of an intensive care unit day: The contribution of mechanical ventilation. Crit Care Med. 2005;33(6):1266-71.
14. Gundlapalli AV, South BR, Phansalkar S, Kinney AY, Shen S, Delisle S, et al. Application of natural language processing to VA electronic health records to identify phenotypic characteristics for clinical and research purposes. Summit on Translat Bioinforma. 2008:36-40.

15. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, et al. Data mining in healthcare and biomedicine: A survey of the literature. J Med Syst. 2011.
16. Ohno-Machado L. Realizing the full potential of electronic health records: The role of natural language processing. J Am Med Inform Assoc. 2011;18(5):539.
17. Streiner DL, Norman GR. Correction for multiple testing: is there a resolution? Chest. 2011;140(1):16-8.
18. Rice TK, Schork NJ, Rao DC. Methods for handling multiple testing. Adv Genet. 2008;60:293-308.