



# Patient Similarity: Emerging Concepts in Systems and Precision Medicine

Sherry-Ann Brown \*

Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN, USA

**Keywords:** patient similarity, patient similarity analytics, computational medicine, big data analytics, clinical decision support

## INTRODUCTION

Healthcare data generates a huge *volume* of information in *various* formats at high *velocity* with sometimes questionable *veracity* (Barkhordari and Niamanesh, 2015) (4V). As a result, big data tools such as patient similarity are necessary to facilitate analytics, which reduces costs (Srinivasan and Arunasalam, 2013) and improves healthcare systems (Jee and Kim, 2013). Patient similarity investigates distances between a variety of components of patient data, and determines methods of clustering patients, based on short distances between some of their characteristics. Although patient similarity is in its early stages, ultimately information about diseases, risk factors, lifestyle habits, medication use, co-morbidities, molecular and histopathological information, hospitalizations, or death are compared with laboratory investigations, imaging, and other clinical data assessing medical evidence of human behavior (**Figure 1**). Such analytics consist of efficient computational analyses with patient stratification by multiple co-occurrence statistics, based on clinical characteristics. Algorithms create subgroups of patients based on similarities among their electronic avatars. Among electronic avatars found to be similar, subgroups of patients can be evaluated by further stratification guided by individual diagnoses, risk factors, medications, and so on. Because of the multiple networks of subgroups of patients, patient similarity can be considered an application of network medicine, with the output termed “patient similarity networks.” Thus, data mining extracts clinically relevant information hidden in clinical notes and embedded in other areas of the electronic health record (EHR) coupled with International Classification of Disease codes. The result is a systematic individualized analysis of a subset of patients that can improve outcome prediction and help guide management for a particular patient currently being cared for by a clinician (Lee et al., 2015). The communication or output from the algorithms can be used to identify and predict disease correlations and occurrence, and potentially for clinical decision support at the point of care. Patient similarity analytics are not restricted to global findings from large clinical trials consisting of somewhat heterogeneous patient populations (Roque et al., 2011). In this way, patient similarity represents a paradigm shift that introduces disruptive innovation to optimize personalization of patient care. Some promising examples are regarding mental and behavioral disorders (Roque et al., 2011), infectious diseases (Li et al., 2015), cancers (Wu et al., 2005; Teng et al., 2007; Chan et al., 2010, 2015; Klenk et al., 2010; Cho and Przytycka, 2013; Li et al., 2015; Wang, 2015; Bolouri et al., 2016; Wang et al., 2016), endocrine (Li et al., 2015; Wang, 2015), and metabolic diseases (Zhang et al., 2014; Ng et al., 2015). Others involve diseases of the nervous system (Lieberman et al., 2005; Carreiro et al., 2013; Cho and Przytycka, 2013; Qian et al., 2014; Buske et al., 2015a; Li et al., 2015; Bolouri et al., 2016; Wang et al., 2016), eyes (Buske et al., 2015a; Li et al., 2015), skin (Buske et al., 2015a; Li et al., 2015), heart (Wu et al., 2005; Tsymbal et al., 2007; Syed and Guttag, 2011; Buske et al., 2015a; Li et al., 2015; Panahiazar et al., 2015a,b; Wang, 2015; Björnson et al., 2016), liver (Chan et al., 2015), intestines (Buske et al., 2015a), musculoskeletal system

## OPEN ACCESS

### Edited by:

Krasimira Tsaneva-Atanasova,  
University of Exeter, UK

### Reviewed by:

Pietro Lio,  
University of Cambridge, UK  
Gustavo Glusman,  
Institute for Systems Biology, USA

### \*Correspondence:

Sherry-Ann Brown  
brown.sherryann@mayo.edu

### Specialty section:

This article was submitted to  
Computational Physiology and  
Medicine,  
a section of the journal  
Frontiers in Physiology

**Received:** 31 August 2016

**Accepted:** 07 November 2016

**Published:** 24 November 2016

### Citation:

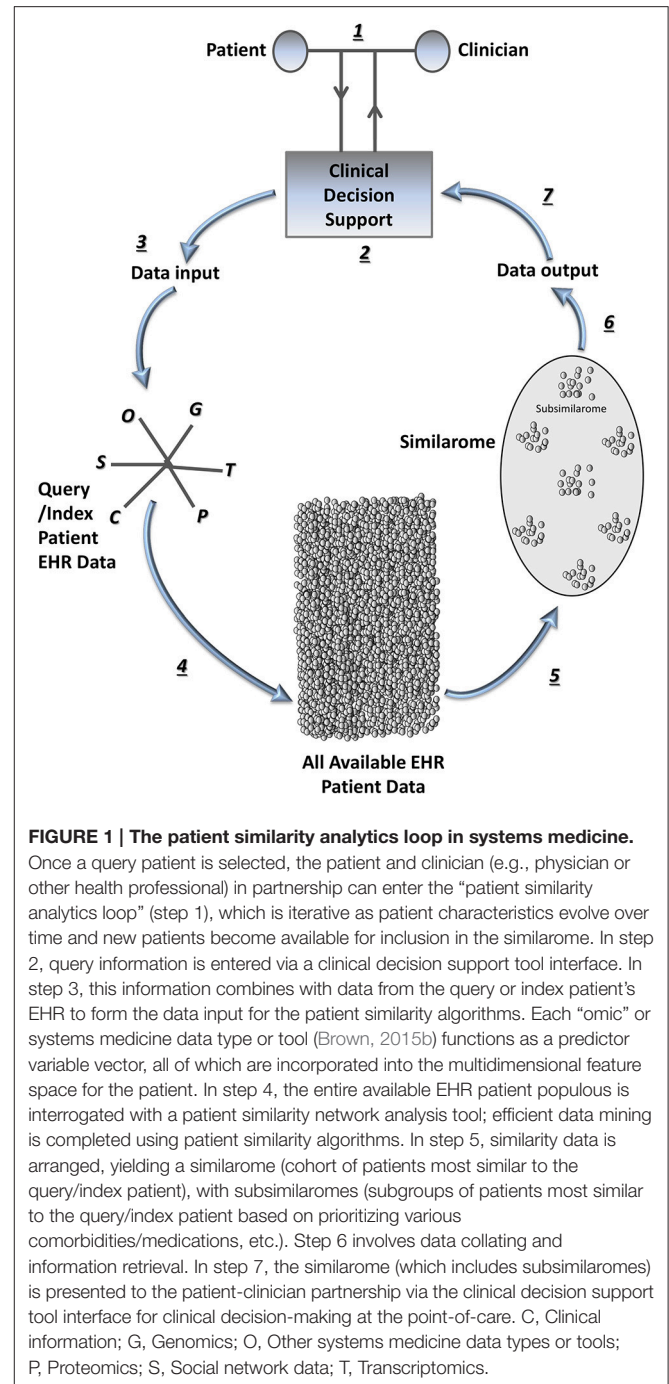
Brown S-A (2016) Patient Similarity:  
Emerging Concepts in Systems and  
Precision Medicine.  
Front. Physiol. 7:561.  
doi: 10.3389/fphys.2016.00561

(Buske et al., 2015a), congenital malformations (Buske et al., 2015a), and various other conditions or factors influencing health status (Gotz et al., 2012; Subirats et al., 2012; Ng et al., 2015).

## PATIENT SIMILARITY IN SYSTEMS MEDICINE

Patient similarity is just starting to spread its wings and has the potential to transform Systems Medicine, which is Systems Biology applied to health care. Systems Biology studies the characteristics of cells, tissues, organisms, or other comprehensive biological units as whole systems. Systems Biology seeks to determine how changes in one part of the system can affect the behavior of the whole system, and often focuses on predictive modeling of the system in a perturbed state. Patient similarity analytics could be developed to bring together characteristics of the patient as a whole human system, and compare these to a multitude of similar patients. Accordingly, patient similarity analytics should in the near future incorporate genomics, transcriptomics, proteomics, microbiomics, and other “omics” and diverse components of systems medicine. In addition, simulation of physiology at the level of the molecule, cell, tissue, organ, and organism should be consolidated as a comprehensive similarity feature to give a broader view of interactions among organ systems. Patient similarity analytics could provide predictive models of a patient’s outcome in the setting of disease perturbations or diagnoses relevant to the index patient. Making adjustments in the query data that serve as input for the predictive models would allow for assessment of how new diagnoses or therapies could impact the overall behavior and phenotype of the whole patient.

Beyond the reasoning above, integrating the majority of these systems medicine tools into patient similarity analytics is potentially the next frontier in Systems Medicine, for at least a few reasons. First, patient similarity analytics embrace a systems view by assessing a myriad of characteristics for hundreds or thousands of patients to produce a meaningful and useful result. Second, patient similarity analytics are analogous to various “omics” that in part compose Systems Biology. Just as transcriptomics refers to generation of messenger RNA expression profiles (Briefing, 1999), one could consider a term similaromics referring to generation or identification of patients similar to an index patient. Similaromics is also akin to phenomics, proteomics, and genomics, among others. Phenomics refers to cataloging the observable characteristics conferred by a gene and proteomics describes the generation of proteins expressed by a cell (Briefing, 1999). One might argue that patient similarity is not quite analogous to genomics, since an individual’s genome is thought to be constant throughout their lifetime. However, this is no longer necessarily the case, due to the current progress of genome editing tools. Indeed, patient similarity is analogous to these various omics, all with the potential to change over the lifetime of the individual. Thus, just as a genome is the complement of all DNA within a cell, a similarome is the complement of patients found to be similar to an index patient. Within the similarome, one can further



distinguish subgroups of patients that are most similar to an index patient, based on preferentially assigning preeminence to comorbidities or medications of most interest or relevance to the index patient, e.g., during a focused shared decision-making session with a clinician. Similar to genotyping then, which determines the presence or absence of a particular gene feature, simotyping would allocate the presence or absence of a particular similarity feature, for example, a diagnosis of diabetes. In this context then, a similarity-wide association study (SiWAS) has the

goal of discovering clusters of patients similar to an index patient and identifying similar features that associate with specific outcomes, such as complications, procedures, hospitalizations, or death. For example, investigating whether in patients most similar to an index patient diabetes is more likely to associate with non-healing leg ulcers, critical limb ischemia, or gangrene leading to limb amputation.

Third, patient similarity analytics have the potential to bring together a variety of omics and other systems medicine tools, if we can do so in a way that is effective, accurate, consistent, and computationally efficient (Brown, 2015a). Indeed, several groups have proposed methods of aggregating omics and monitoring these over time for individual patients, and perhaps even using comprehensive patient avatars. Integrating these methods with patient similarity has the potential to launch systems medicine further into a future where medicine is even more precisely individualized. Patient similarity will likely become and persist as a useful tool in systems medicine.

## MATHEMATICS IN PATIENT SIMILARITY ANALYTICS

For illustration of the utility of patient similarity in medicine, only briefly presented here are a few selected examples of patient similarity analytics used for diabetes and cancer, which are common chronic or terminal diseases, respectively, currently addressed in public health. In some studies, a patient similarity metric is determined as follows (Lee et al., 2015; Li et al., 2015). A patient can be represented by a Euclidean vector. Predictor variables such as laboratory test results or vital signs can define a multi-dimensional feature space. The cosine of the angle between two patients' vectors can define the associated patient similarity metric. A dot product can facilitate the calculation. This can be termed the "cosine similarity," defining the patient similarity metric as follows:

$$PSM(P_1, P_2) = \frac{P_1 \cdot P_2}{\|P_1\| \|P_2\|},$$

$$= \frac{\sum_{i=1}^n P_{1i} \times P_{2i}}{\sqrt{\sum_{i=1}^n P_{1i}^2} \times \sqrt{\sum_{i=1}^n P_{2i}^2}}$$

where  $P_{1i}$  and  $P_{2i}$  represent a single predictor variable vector for two separate patients,  $\cdot$  represents the dot product, and  $\| \cdot \|$  represents the Euclidean vector magnitude, as shown. Since the patient similarity metric is an angle cosine, it normalizes between  $-1$  (considered minimum possible similarity) and  $1$  (considered maximum possible similarity). As expected, two predictor variable vectors pointing in the exact opposite direction to each other would have a  $180^\circ$  angle between them, and would therefore calculate to a patient similarity metric of  $-1$ . Conversely, two perfectly overlapping vectors would have an angle of  $0^\circ$  between them, and would therefore calculate to a patient similarity metric of  $1$ . Accordingly, before calculating the total patient similarity metric, the product for each predictor variable vector would be normalized to the range of  $-1$  to  $1$  in the

multidimensional feature space, if continuous (Lee et al., 2015). The product for categorical/binary predictor variable vectors would be assigned a value of  $-1$  or  $1$ . The patient similarity metric would be calculated for each patient in a given data set, relative to an index patient  $P_1$ . The  $N$  most similar patients to the index patient would be utilized as a training data set for testing in a validation data set, with prediction of prognosis, morbidity, or mortality. After successful validation, the predictive model could be used for epidemiologic or clinical studies. For example, an algorithm using cosine similarity successfully identified three subgroups of patients with diabetes (Li et al., 2015). The first subgroup included patients with diabetic nephropathy (diabetes-related kidney disease) and diabetic retinopathy (diabetes-related eye disease). The second subgroup included several patients with cancer and cardiovascular diseases. The third subgroup included many patients who also had cardiovascular diseases, along with neurological diseases, allergies, and HIV infection. Various single nucleotide polymorphisms mapped to these three subgroups that were confirmed in the EHR, suggesting clinical relevance for patient similarity in precision medicine. Jaccard similarity, another metric that can be leveraged after assigning binary attributes to each patient's multifeature vector space, was useful to analyze features underlying deviant responses to therapeutics in patients with diabetes (Zhang et al., 2014).

Alternatively, unsupervised clustering of patients based on their clinical predictor variables could be used to produce a patient-patient network. The network could be organized using L-infinity centrality, which is the maximum distance from each point from any other point in a given data set. L-infinity centrality produces a detailed and succinct description of any data set yielding more information than scatter plots (Lum et al., 2013). Large values for L-infinity centrality correspond to data points at large distances from the center of the data set (Li et al., 2015). Other pattern analysis and cluster algorithms (Daemen and De Moor, 2009; Chan et al., 2010; Liu et al., 2013a; Mabotuwana et al., 2013; Sundar et al., 2014), or algorithms incorporating distance metric learning (Wang et al., 2011; Bian and Tao, 2012), locally supervised metric learning (Sun et al., 2012; Ng et al., 2015), local spline regression (Wang et al., 2012), or visual analytics (Tsymbal et al., 2009; Ebadollahi et al., 2010; Gotz et al., 2011; Perer, 2012; Heer and Perer, 2014; Bolouri et al., 2016; Ozery-Flato et al., 2016), can also be used for patient similarity to predict diabetes onset, develop treatment recommendations tailored to each patient, or predict survival after chemotherapy (Chan et al., 2010; Liu et al., 2013a; Ng et al., 2015; Ozery-Flato et al., 2016), among other applications. SNOMED CT and other medical terminology frameworks can be used to facilitate communication across platforms in various studies (Melton et al., 2006). There are also algorithms to incorporate a time series into patient similarity analysis, to predict trends over time among patients (Wu et al., 2005; Hartge et al., 2006; Ebadollahi et al., 2010; Carreiro et al., 2013; Alaa et al., 2016). For example, a patient similarity time series algorithm has been used to fine-tune radiation treatment planning for patients with head and neck cancers (Wu et al., 2005).

## CHALLENGES IN PATIENT SIMILARITY

There are certain challenges in patient similarity, such as network bottlenecks, low hardware performance (processing power and memory), and data locality (Osman et al., 2013; Karapiperis and Verykios, 2014; Barkhordari and Niamanesh, 2015). Given the observational or retrospective nature of patient similarity, interpretation of data analysis will be imperfect. Confounder control and treatment selection bias are inherent limitations in such studies. However, groups have developed strategies to manage the potential for confounders, such as restriction, stratification, matching, inverse probability weighting, and covariate adjustment (Gallego et al., 2015). Several groups have also proposed solutions for other challenges that enable large scale patient indexing and accurate and efficient clinical data retrieval (Wang, 2015). Some have devised algorithms to address the complexity of clinical data and limited transparency of many existing clinical case retrieval decision support systems (Tsymbal et al., 2009), as well as integration of data from various heterogeneous omics studies (Wang et al., 2014, 2016; Gligorijević et al., 2016) and physician input and feedback (Wang et al., 2011; Sun et al., 2012; Fei and Sun, 2015). Others have produced algorithms that address scalability and uncertainty, by requiring parallel or distributed algorithm implementations built to scale, and enhancing interpretability by conveying the certainty of results presented (Feldman et al., 2015). One such algorithm or platform is scalable and distributable patient similarity (ScaDiPaSi), a dynamic method for investigating patient similarity that spreads the algorithm over several self-sufficient hardware nodes to process query data from various sources of different formats simultaneously (Barkhordari and Niamanesh, 2015). Another tool, MapReduce, employs several optimization techniques, such as job scheduling and cascading work flows over multiple interdependent hardware nodes (Dean and Ghemawat, 2008). Use of all of these technological solutions for patient similarity in precision medicine will be facilitated by bridging gaps among different scientific, technological, and medical cultures, through interdisciplinary collaborations among experts in medicine, biology, informatics, engineering, public health, economics, and the social sciences (Kuhn et al., 2008).

## REFERENCES

- Alaa, A., Yoon, J., Hu, S., and van der Schaar, M. (2016). "Personalized risk scoring for critical care patients using mixtures of Gaussian Process Experts" in *Proceedings of the 33rd International Conference on Machine Learning*, vol 48 (New York, NY).
- Barkhordari, M., and Niamanesh, M. (2015). ScaDiPaSi: an effective scalable and distributable MapReduce-Based method to find patient similarity on huge healthcare networks. *Big Data Res.* 2, 19–27. doi: 10.1016/j.bdr.2015.02.004
- Briefing (1999). Proteomics, transcriptomics: what's in a name? *Nature* 402:715.
- Bian, W., and Tao, D. (2012). Constrained empirical risk minimization framework for distance metric learning. *IEEE Trans. Neural Netw. Learn. Syst.* 23, 1194–1205. doi: 10.1109/TNNLS.2012.2198075
- Björnson, E., Borén, J., and Mardinoglu, A. (2016). Personalized cardiovascular disease prediction and Treatment-A review of existing strategies and novel systems medicine tools. *Front Physiol.* 7:2. doi: 10.3389/fphys.2016.00002

## CONCLUSION

Various patient similarity algorithms have been deployed and have been found beneficial by improving clinical efficiency (Wang et al., 2015), enabling secure identification of similar patients and records sharing by clinicians and rare disease scientists (Buske et al., 2015a,b), predicting patients' prognosis or trajectory over time (Ebadollahi et al., 2010; Subirats et al., 2012; Wang et al., 2012; Gallego et al., 2015), providing clinical decision support (Daemen et al., 2009; Wang et al., 2011; Subirats et al., 2012; Sun et al., 2012; Gottlieb et al., 2013; Liu et al., 2013b; Gallego et al., 2015), tailoring individual treatments (Zhang et al., 2014), preventing unexpected adverse drug reactions (Hartge et al., 2006; Yang et al., 2014), flagging patients deserving more attention due to poor response to therapies (Zhang et al., 2014; Ozery-Flato et al., 2016), and pursuing comparative effectiveness studies (Wang et al., 2011), among other applications. In general, clinical guidelines often do not supply evidence on risks, secondary therapy effects, and long-term outcomes (Gallego et al., 2015). In this setting, patient similarity analytics can provide a cheaper, portable alternative or in fact adjunct to evidence-based clinical guidelines and randomized controlled trials, particularly if trial data are unavailable for conditions or patient characteristics specific to a query individual (Longhurst et al., 2014; Gallego et al., 2015). Synthesizing current patient similarity algorithms with systems medicine tools could provide actionable insights in precision medicine.

## AUTHOR CONTRIBUTIONS

SB conceived, analyzed, designed, drafted, critically revised, approved, and agreed to be accountable for this submitted work.

## ACKNOWLEDGMENTS

The author is grateful to Dr. Joerg Herrmann of Mayo Clinic in Rochester, Minnesota for reading the manuscript.

- Bolouri, H., Zhao, L. P., and Holland, E. C. (2016). Big data visualization identifies the multidimensional molecular landscape of human gliomas. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5394–5399. doi: 10.1073/pnas.1601591113
- Brown, S. A. (2015a). Principles for developing patient avatars in precision and systems medicine. *Front Genet.* 6:365 doi: 10.3389/fgene.2015.00365
- Brown, S. A. (2015b). Building SuperModels: emerging patient avatars for use in precision and systems medicine. *Front. Physiol.* 6:318. doi: 10.3389/fphys.2015.00318
- Buske, O. J., Girdea, M., Dumitriu, S., Gallinger, B., Hartley, T., Trang, H., et al. (2015b). PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *Hum Mutat.* 36, 931–940. doi: 10.1002/humu.22851
- Buske, O. J., Schiettecatte, F., Hutton, B., Dumitriu, S., Misyura, A., Huang, L., et al. (2015a). The Matchmaker Exchange API: automating patient matching through the exchange of structured phenotypic and genotypic profiles. *Hum Mutat.* 36, 922–927. doi: 10.1002/humu.22850

- Carreiro, A., Madeira, S., and Francisco, A. (2013). "Unravelling communities of ALS patients using network mining," in *KDD-DMH'13* (Chicago, IL).
- Chan, L., Chan, T., Cheng, L., and Mak, W. (2010). "Machine learning of patient similarity: a case study on predicting survival in cancer patient after locoregional chemotherapy," in *2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops* (Hong Kong), 467–470.
- Chan, L. W., Liu, Y., Chan, T., Law, H. K., Wong, S. C., Yeung, A. P., et al. (2015). PubMed-supported clinical term weighting approach for improving inter-patient similarity measure in diagnosis prediction. *BMC Med Inform Decis Mak.* 15:43. doi: 10.1186/s12911-015-0166-2
- Cho, D. Y., and Przytycka, T. M. (2013). Dissecting cancer heterogeneity with a probabilistic genotype-phenotype model. *Nucleic Acids Res.* 41, 8011–8020. doi: 10.1093/nar/gkt577
- Daemen, A., and De Moor, B. (2009). Development of a kernel function for clinical data. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 5913–5917. doi: 10.1109/iembs.2009.5334847
- Daemen, A., Gevaert, O., Ojeda, F., Debucquoy, A., Suykens, J. A., Sempoux, C., et al. (2009). A kernel-based integration of genome-wide data for clinical decision support. *Genome Med.* 1:39. doi: 10.1186/gm39
- Dean, J., and Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Commun. A.C.M.* 51, 107–113. doi: 10.1145/1327452.1327492
- Ebadollahi, S., Sun, J., Gotz, D., Hu, J., Sow, D., and Neti, C. (2010). Predicting patient's trajectory of physiological data using temporal trends in similar patients: a system for near-term prognostics. *AMIA Annu. Symp. Proc.* 2010, 192–196.
- Fei, W., and Sun, J. (2015). PSF: a unified patient similarity evaluation framework through metric learning with weak supervision. *IEEE J. Biomed. Health Inform.* 19, 1053–1060. doi: 10.1109/JBHI.2015.2425365
- Feldman, K., Davis, D., and Chawla, N. V. (2015). Scaling and contextualizing personalized healthcare: a case study of disease prediction algorithm integration. *J. Biomed. Inform.* 57, 377–385. doi: 10.1016/j.jbi.2015.07.017
- Gallego, B., Walter, S. R., Day, R. O., Dunn, A. G., Sivaraman, V., Shah, N., et al. (2015). Bringing cohort studies to the bedside: framework for a 'green button' to support clinical decision-making. *J. Comp. Eff. Res.* 11, 1–7. doi: 10.2217/ce.15.12
- Gligorijević, V., Malod-Dognin, N., and Pržulj, N. (2016). Integrative methods for analyzing big data in precision medicine. *Proteomics* 16, 741–758. doi: 10.1002/pmic.201500396
- Gottlieb, A., Stein, G. Y., Ruppig, E., Altman, R. B., and Sharan, R. (2013). A method for inferring medical diagnoses from patient similarities. *BMC Med.* 11:194. doi: 10.1186/1741-7015-11-194
- Gotz, D., Stavropoulos, H., Sun, J., and Wang, F. (2012). ICDA: a platform for Intelligent Care Delivery Analytics. *AMIA Annu. Symp. Proc.* 2012, 264–273.
- Gotz, D., Sun, J., Cao, N., and Ebadollahi, S. (2011). Visual cluster analysis in support of clinical decision intelligence. *AMIA Annu. Symp. Proc.* 2011, 481–490.
- Hartge, F., Wetter, T., and Haefeli, W. E. (2006). A similarity measure for case based reasoning modeling with temporal abstraction based on cross-correlation. *Comput. Methods Progr. Biomed.* 81, 41–48. doi: 10.1016/j.cmpb.2005.10.005
- Heer, J., and Perer, A. (2014). Orion: a system for modeling, transformation and visualization of multidimensional heterogeneous networks. *Informat. Visualizat.* 13, 111–133. doi: 10.1177/1473871612462152
- Jee, K., and Kim, G. H. (2013). Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Health Inform. Res.* 19, 79–85. doi: 10.4258/hir.2013.19.2.79
- Karapiperis, D., and Verykios, V. (2014). A distributed near-optimal LSH-based framework for privacy-preserving record linkage. *Sci. Inf. Syst.* 11, 745–763. doi: 10.2298/CSIS140215040K
- Klenk, S., Dippon, J., Fritz, P., and Heidemann, G. (2010). "Determining patient similarity in medical social networks," in *MEDEX 2010 Proceedings*, Vol. 572 (Raleigh, NC), 6–13.
- Kuhn, K., Knoll, A., Mewes, H.-W., Schwaiger, M., Bode, A., Broy, M., et al. (2008). Informatics and medicine: from molecules to populations. *Methods Inf. Med.* 47, 283–295. doi: 10.3414/ME9117
- Lee, J., Maslove, D. M., and Dubin, J. A. (2015). Personalized mortality prediction driven by electronic medical data and a patient similarity metric. *PLoS ONE* 10:e0127428. doi: 10.1371/journal.pone.0127428
- Li, L., Cheng, W. Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., et al. (2015). Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* 7:311ra174. doi: 10.1126/scitranslmed.aaa9364
- Lieberman, M. A., Winzelberg, A., Golant, M., Wakahiro, M., DiMinno, M., Aminoff, M., et al. (2005). Online support groups for Parkinson's patients: a pilot study of effectiveness. *Soc. Work Health Care.* 42, 23–38. doi: 10.1300/J010v42n02\_02
- Liu, H., Xie, G., Mei, J., Shen, W., Sun, W., and Li, X. (2013a). An efficacy driven approach for medication recommendation in type 2 diabetes treatment using data mining techniques. *Stud. Health Technol. Inform.* 192, 1071.
- Liu, L., Tang, J., Agrawal, A., Liao, W., and Choudhary, A. (2013b). *Mining Diabetes Complication and Treatment Patterns for Clinical Decision Support*. San Francisco, CA.
- Longhurst, C. A., Harrington, R. A., and Shah, N. H. (2014). A 'green button' for using aggregate patient data at the point of care. *Health Aff. (Millwood)*. 33, 1229–1235. doi: 10.1377/hlthaff.2014.0099
- Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., et al. (2013). Extracting insights from the shape of complex data using topology. *Sci. Rep.* 3:1236. doi: 10.1038/srep01236
- Mabotuwana, T., Lee, M. C., and Cohen-Solal, E. V. (2013). An ontology-based similarity measure for biomedical data-application to radiology reports. *J. Biomed. Inform.* 46, 857–868. doi: 10.1016/j.jbi.2013.06.013
- Melton, G. B., Parsons, S., Morrison, F. P., Rothschild, A. S., Markatou, M., and Hripcsak, G. (2006). Inter-patient distance metrics using SNOMED CT defining relationships. *J. Biomed. Informat.* 39, 697–705. doi: 10.1016/j.jbi.2006.01.004
- Ng, K., Sun, J., Hu, J., and Wang, F. (2015). Personalized predictive modeling and risk factor identification using patient similarity. *AMIA Jt. Summits Transl. Sci. Proc.* 2015, 132–136.
- Osman, A., El-Refaay, M., and Elnaggar, A. (2013). "Towards real-time analytics in the cloud," in *2013 IEEE Ninth World Congress on Services (SERVICES)* (Santa Clara, CA: IEEE).
- Ozery-Flato, M., Ein-Dor, L., Parush-Shear-Yashuv, N., Aharonov, R., Neuvirth, H., Kohn, M. S., et al. (2016). Identifying and investigating unexpected response to treatment: a diabetes case study. *Big Data* 4, 148–159. doi: 10.1089/big.2016.0017
- Panahiazar, M., Taslimitehrani, V., Pereira, N. L., and Pathak, J. (2015b). Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics. *Stud. Health Technol. Inform.* 210, 369–373.
- Panahiazar, M., Taslimitehrani, V., Pereira, N., and Pathak, J. (2015a). Using EHRs and machine learning for heart failure survival analysis. *Stud Health Technol Inform.* 216, 40–44.
- Perer, A. (2012). *Healthcare Analytics for Clinical and Non-Clinical Settings*. CHI'12. Austin, TX.
- Qian, B., Wang, X., Cao, N., Li, H., and Jiang, Y. (2014). A relative similarity based method for interactive patient risk prediction. *Data Min. Knowl. Disc.* 29, 1070–1093.
- Roque, F. S., Jensen, P. B., Schmock, H., Dalgaard, M., Andreatta, M., Hansen, T., et al. (2011). Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS Comput. Biol.* 7:e1002141. doi: 10.1371/journal.pcbi.1002141
- Srinivasan, U., and Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs. *IT Professional.* 15, 21–28. doi: 10.1109/MITP.2013.55
- Subirats, L., Ceccaroni, L., and Miralles, F. (2012). Knowledge representation for prognosis of health status in rehabilitation. *Future Internet* 4, 762–775. doi: 10.3390/fi4030762
- Sun, J., Wang, F., Hu, J., and Ebadollahi, S. (2012). Supervised patient similarity measure of heterogeneous patient records. *SIGKDD Explorations.* 14, 16–24. doi: 10.1145/2408736.2408740
- Sundar, I. K., Yao, H., Huang, Y., Lyda, E., Sime, P. J., Sellix, M. T., et al. (2014). Serotonin and corticosterone rhythms in mice exposed to cigarette smoke and in patients with COPD: implication for COPD-associated neuropathogenesis. *PLoS ONE* 9:e87999. doi: 10.1371/journal.pone.0087999
- Syed, Z., and Guttag, J. (2011). Unsupervised similarity-based risk stratification for cardiovascular events using long-term time-series data. *J. Mach. Learn. Res.* 12, 999–1024.

- Teng, C., Shapiro, L., Kalet, I., Rutter, C., and Nurani, R. (2007). "Head and neck cancer patient similarity based on anatomical structural geometry," in *2007 IEEE International Conference on Bioinformatics and Biomedicine Workshops* (San Jose, CA).
- Tsymbal, A., Huber, M., Zillner, S., Hauer, T., and Zhuo, K. (2007). "Visualizing patient similarity in clinical decision support," in *LWA 2007: Lernen - Wissen - Adaption, Workshop Proceedings*, Vol. 6 (Halle: Martin-Luther-University Halle-Wittenberg), 304–311.
- Tsymbal, A., Zhou, S. K., and Huber, M. (2009). Neighborhood graph and learning discriminative distance functions for clinical decision support. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 5617–5620. doi: 10.1109/iembs.2009.5333784
- Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., et al. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11, 333–337. doi: 10.1038/nmeth.2810
- Wang, F. (2015). Adaptive semi-supervised recursive tree partitioning: the ART towards large scale patient indexing in personalized healthcare. *J. Biomed Inform.* 55, 41–54. doi: 10.1016/j.jbi.2015.01.009
- Wang, F., Hu, J., and Sun, J. (2012). "Medical prognosis based on patient similarity and expert feedback," in *21st International Conference on Pattern Recognition (ICPR 2012)* (Tsukuba), 1799–1802.
- Wang, F., Sun, J., and Ebadollahi, S. (2011). "Integrating distance metrics learned from multiple experts and its application in patient similarity assessment," in *Proceedings of the 2011 SIAM International Conference on Data Mining* (Philadelphia, PA: Society for Industrial and Applied Mathematics), 59–70. doi: 10.1137/1.9781611972818.6
- Wang, H., Zheng, H., Wang, J., Wang, C., and Wu, F. (2016). Integrating Omics Data With a Multiplex Network-Based Approach for the Identification of Cancer Subtypes. *IEEE Transact. Nanobiosci.* 15, 335–342. doi: 10.1109/TNB.2016.2556640
- Wang, Y., Tian, Y., Tian, L. L., Qian, Y. M., and Li, J. S. (2015). An electronic medical record system with treatment recommendations based on patient similarity. *J. Med. Syst.* 39:55. doi: 10.1007/s10916-015-0237-z
- Wu, H., Salzberg, B., Sharp, G., Jiang, S., Shirato, H., and Kaeli, D. (2005). "Subsequence matching on structured time series data," in *Proceedings of the ACM SIGMOD International Conference on Management of Data* (Baltimore, MD), 682–693.
- Yang, F., Yu, X., and Karypis, G. (2014). "Signaling adverse drug reactions with novel feature-based similarity model," in *IEEE Conference on Bioinformatics and Biomedicine* (Belfast).
- Zhang, P., Wang, F., Hu, J., and Sorrentino, R. (2014). Towards personalized medicine: leveraging patient similarity and drug similarity analytics. *AMIA Jt. Summits Transl. Sci. Proc.* 2014, 132–136.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Brown. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.