

Clinical and Clinical-Genomic Data Mining for Diagnosis and Prognosis  
Barry Robson, IBM Research

Clinical and clinical-genomic data mining involves analysis of some of the most challenging high-dimensional data. The number of combinations of associations and other correlations between clinical, lifestyle and genomic features is trans-astronomical and requires special emerging techniques. A special theory, Zeta Theory, has been developed in which the expected value of information measures can be written in terms of the Riemann Zeta Function, items on records are represented by primes, and records are coded as products of these primes or as lists of partial products in equilibrium. In consequence of this approach, a direct mapping of all statements, theorems, functions, and formulae in prime number theory to records and data mining has been noted and has suggested several useful algorithms. By these means, the program CliniMiner (“FANO”) has been developed, and clinical data as well as administrative data and epidemiologic records can be submitted to useful analysis for non-intuitive but informative insights, with careful attention to HIPPA requirements. Recently the order of a million records from several collaborators, each with hundreds of parameters, has been analyzed with interesting results.

---

Barry Robson BSc (Hons), PhD, DSc  
IBM Distinguished Engineer,  
Strategic Advisor,  
Chair Professional Interest Communities (Computational Biology and Medicine),  
Computational Biology Center,  
T. J. Watson Research Center,  
1101 Kitchawan Road, Yorktown Heights, NY 10598  
1-914-945-1030 (office); 1-914-945-2079 (assistant)  
<http://www.research.ibm.com/people/r/robson/>  
[robsonb@us.ibm.com](mailto:robsonb@us.ibm.com)