Descriptively Probabilistic Relationship Between Mutated Primary Structure of Coagulation Factor IX and Clinical Severity of Hemophilia B

Shaomin Yan[1]
Guang Wu[2]

[1]*Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi, CN-530007, China,*

[2]*Computational Mutation Project, DreamSciTech Consulting, Shenzhen, Guangdong Province CN-518054, China*

KEY WORDS: Amino acid, Bayes' law, coagulation factor IX, cross-impact analysis, distribution probability, hemophilia B, mutation.

ABSTRACT

Hemophilia B is a recessive bleeding disorder resulting from mutations in the coagulation factor IX gene. As this disease is characterized by clinical and molecular heterogeneity, the building of relationship between its genotype and phenotype would be great helpful for better diagnosis, prognosis and treatment. We use a descriptively probabilistic method, cross-impact analysis, to couple the changed primary structure of mutant human coagulation factor IX with the severity of hemophilia B with the help of the amino-acid distribution probability as a quantitative measure for mutation. Then we use the Bayesian equation to calculate the probability that the severity of hemophilia can be defined under a mutation. A patient has larger than 0.5 chance of being defined as severity of hemophilia B when a new mutation is found in coagulation factor IX. In this way, we take the first step towards further modeling of genotype-phenotype relationship in human coagulation factor IX.

INTRODUCTION

The coagulation factor IX precursor contains coagulation factor IXa light chain and heavy chain. After activation of coagulation factor IX to factor IXa, this enzyme interacts with the active cofactor form of factor VIII, to form a complex on membrane surfaces. This complex converts factor X to factor Xa [1]. Thus, the coagulation factor IX is one of critical components of the blood coagulation pathways, and its deficiency causes hemophilia B [2].

Hemophilia B is a recessive bleeding disorder that results from mutations in the coagulation factor IX gene on the X chromosome [3-5]. It occurs in one of 30 000 live male births in all populations [6, 7]. Major acute and chronic complications are often secondary to recurrent bleeding [8]. The unpredictable, recurrent, spontaneous bleedings mainly appear in soft tissues and/or major joints. Recurrent bleeding in large joints usually leads to crippling arthropathies in a majority of severely affected patients.

The clinical severity of hemophilia B corresponds to the level of circulating coagulation factor IX. Severe hemophilia occurs in less than 1% of coagulation factor IX activity. With moderate hemophilia, 1 –

5% of coagulation factor IX activity, there is infrequent, spontaneous bleeding. The presence of at least 5% of coagulation factor IX seems to protect those with mild hemophilia against spontaneous bleeding [6, 8]. Each individual case of hemophilia is characterized by a series of unique parameters, emphasizing the variability and heterogeneity of this disease. These parameters include the mode of initial presentation, the baseline level of the clotting factor, and the presence or absence of a relevant family history [9].

Although the affected males are born to carrier females, up to 50% of cases appear de novo as a result of new mutations [10-12]. Approximately 1 000 unique mutations causing hemophilia B have been reported in humans [13-21]. Approximately 3% of hemophilia B patients have major deletions in the coagulation factor IX gene, half of which are complete [22].

As hemophilia B is characterized by clinical and molecular heterogeneity [23], it is important to find a way to connect the mutations and their clinical outcomes together, by which we could approach to predicting a possibly clinical outcome when a mutation is found. For clinical manifestation, it is easy to consider its appearance/non-appearance as an event with two options, but it is hard to present a mutation, which can occur at different position with different amino acid at coagulation factor IX, as an event with limited choices. Without limited choices, it means that the coagulation factor IX needs to be represented as a number, then any mutation would leads this number to change. In other words, we need to convert a 20-letter symbolized protein sequence into a numeric sequence in order to reach this above aim. Actually, there are currently several ways in doing so, for simplest example, we can use the physicochemical property of amino acid to replace each amino acid in a protein to get the numeric sequence, however the physicochemical property of amino acid is not subject to mutation.

Since 1999, our group has developed three approaches to doing this conversion (for reviews, see [24-26]), and our approaches are more suitable to study the mutations. In this study, we use our approach to building a descriptively probabilistic relationship between mutated primary structure of coagulation factor IX and clinical severity of hemophilia B.

MATERIALS AND METHODS

Data

The human coagulation factor IX precursor with total 145 mutations is obtained from UniProtKB/Swiss-Prot entry [27]. Of them, 141 are missense mutations, 1 insertion and 3 deletions.

Amino-acid distribution probability before and after mutation

The position of amino acid in a protein can be associated with probability, computed using $\frac{r!}{(q_0!\times q_1!\times\ldots\times q_n!)}\times\frac{r!}{(r_1!\times r_2!\times\ldots\times r_n!)}\times n^{-r}$ [28], where r is the number of amino acids, n is the number of partitions, rn is the number of amino acids in the n-th partition, qn is the number of partitions with the same number of amino acids, and ! is the factorial function.

For example, there are fourteen glutamines (Q) in normal human coagulation factor IX, positioned at 2, 57, 90, 96, 143, 167, 185, 216, 219, 237, 241, 292, 370 and 408. According to the equation above, we can imagine the coagulation factor IX as 14 partitions with equal length, each contains 33 (461/14 = 32.93) amino acids because the coagulation factor IX is composed of 461 amino acids. Then, fourteen Qs have the distribution patterns as those in the second column in Table 1, whose amino-acid distribution probability is $r_1 = 1$, $r_2 = 1$, $r_3 = 2$, $r_4 = 0$, $r_5 = 1$, $r_6 = 2$, $r_7 = 2$, $r8 = 2$, $r_9 = 1$, $r_{10} = 0$, $r_{11} = 0$, $r_{12} = 1$, $r_{13} = 1$, $r_{14} = 0$, *and* $q_0 = 4$, $q_1 = 6$, $q_2 = 4$, $q_3 = 0$, $q_4 = 0$, $q_5 = 0$, $q_6 = 0$, $q_7 = 0$, $q_8 = 0$, $q_9 = 0$, $q_{10} = 0$, $q_{11} = 0$, $q_{12} = 0$, $q_{13} = 0$, $q_{14} = 0$, then

$$\frac{1!}{4!\times6!\times4!\times0!\times0!\times0!\times0!\times0!\times0!\times0!\times0!\times0!\times0!\times0!\times0!}\times\frac{1!}{1!\times1!\times2!\times0!\times1!\times2!\times2!\times2!\times1!\times0!\times0!\times1!\times1!\times0!}\times14^{-14}$$

$$=\frac{87178291200}{24\times720\times24\times1\times1\times1\times1\times1\times1\times1\times1\times1\times1\times1\times1}\times\frac{87178291200}{1\times1\times2\times1\times1\times2\times2\times2\times1\times1\times1\times1\times1\times1}$$

$$\times\frac{1}{11112006825558016}=0.1031$$

Table 1. *Glutamines and histidines and their probability before and after mutation at position 167 in factor IX*

Partition	Before mutation		After mutation	
	Glutamine (Q)	Histidine (H)	Glutamine (Q)	Histidine (H)
I	1	1	1	1
II	1	0	1	0
III	2	0	2	0
IV	0	0	1	1
V	1	0	0	0
VI	2	2	2	0
VII	2	4	3	3
VIII	2	1	0	3
IX	1	2	1	1
X	0	1	0	2
XI	0	-	1	0
XII	1	-	1	-
XIII	1	-	0	-
XIV	0	-	-	-
Probability	0.1031	0.0286	0.1544	0.0539

Table 2*. Computation on cross-impact analysis in Fig. 1*

P(2) = 57/141 = 0.4043
$P(\bar{2})$ = 1 – P(2) = 1 – 0.4043 = 0.5957 = 84/141
$P(1\mid\bar{2})$ = 51/84 = 0.6071
$P(\bar{1}\mid\bar{2})$ = 1 – $P(1\mid\bar{2})$ = 1 – 0.6071 = 0.3929 = 33/84
P(1|2) = 31/57 = 0.5439
$P(\bar{1}\mid 2)$ = 1 –P(1|2) = 1 – 0.5439 = 0.4561 = 26/57
$P(1\bar{2})$ = $P(1\mid\bar{2})$ × $P(\bar{2})$ = 51/84 × 84/141 = 0.3617 = 51/141
$P(\bar{1}\bar{2})$ = $P(\bar{1}\mid\bar{2})$ × $P(\bar{2})$ = 33/84 × 84/141 = 0.2340 = 33/141
P(12) = P(1|2) ×P(2) = 31/57 × 57/141 = 0.2199 = 31/141
$P(\bar{1}2)$ $P(\bar{1}\mid 2)$ P(2) = 26/57 × 57/141 = 0.1844 = 26/141

Any point mutation leads an amino acid to change to another one, which certainly changes the distribution pattern of both original and mutated amino acids, thus the amino-acid distribution probabilities will be different for both original and mutated amino acids in the normal and mutant coagulation factor IX.

For example, there is a mutation at position 167 changing Q to histidine (H), then we have 13 Qs after mutation (column 4, Table 1), that is,

$r_1 = 1, r_2 = 1, r_3 = 2, r_4 = 1, r_5 = 0, r_6 = 2, r_7 = 3, r_8 = 0, r_9 = 1, r_{10} = 0, r_{11} = 1, r_{12} = 1, r_{13} = 0$, *and* $q_0 = 4, q_1 = 6, q_2 = 2, q_3 = 1, q_4 = 0, q_5 = 0, q_6 = 0, q_7 = 0, q_8 = 0, q_9 = 0, q_{10} = 0, q_{11} = 0, q_{12} = 0, q_{13} = 0$, then

$$\frac{13!}{4!\,6!\,2!\,1!\,0!\,0!\,2!\,3!\,0!\,1!\,1!\,0!\,0!} \times \frac{13!}{1!\,1!\,2!\,1!\,0!\,2!\,6!\,0!\,1!\,1!\,1!\,1!\,0!} \times 13^{-13}$$

$$= \frac{6227020800}{24 \times 720 \times 2 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1 \times 1} \times \frac{6227020800}{1 \times 1 \times 2 \times 1 \times 1 \times 2 \times 6 \times 1 \times 1 \times 1 \times 1 \times 1}$$

$$\times \frac{1}{302875106592253} = 0.1544$$

For the mutated amino acid, there are 10 Hs in normal coagulation factor IX and 11 Hs in the mutant. Their distribution probabilities are 0.0286 and 0.0539 before and after mutation, so the mutation increases the distribution probability of H.

Because this mutation increases the distribution probability of both the original and mutated amino acids, its overall effect obviously brings about an increment of the distribution probability in the mutant coagulation factor IX, (0.1544 - 0.1031) + (0.0539 - 0.0286) = 0.0766. Actually we have used this approach in many our previous studies [29-47].

In this manner, we have different numbers for different mutations in coagulation factor IX and their documented clinical manifestation, and we therefore can build a quantitative relationship between changed primary structure of coagulation factor IX and clinical severity of hemophilia B.

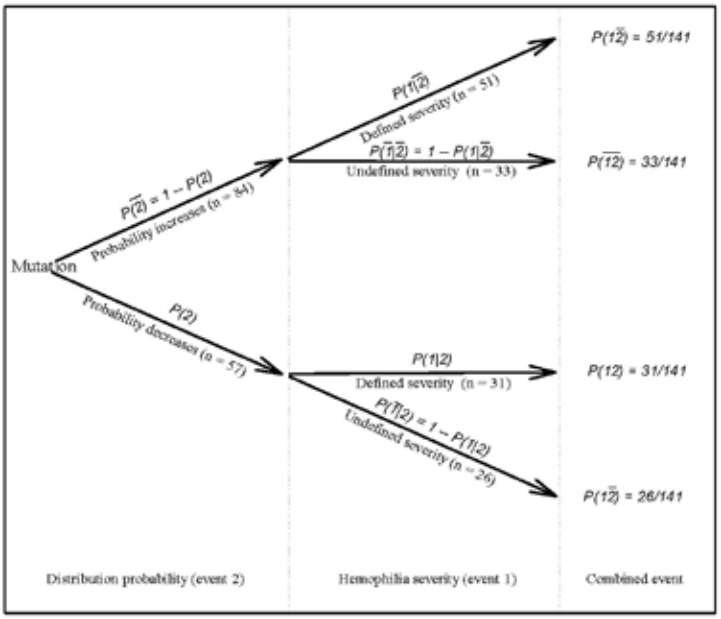
RESULTS AND DISCUSSION

Currently, 141 mutations are documented with hemophilia B, among which 82 are defined as severity. Thus, we can use the cross-impact analysis to build a quantitative relationship between the increase/decrease of distribution probability after mutations and the defined/undefined severity of hemophilia B, because the appearance/non-appearance is an event with two options, and the mutation effect on coagulation factor IX is also an event with two options as increased or decreased amino-acid distribution probability, while the cross-impact analysis is particularly suited for these [38, 48-53].

Figure 1 shows the cross-impact relationship between coagulation factor IX mutations and their hemophilia severity. At the level of amino-acid distribution probability, P(2) and $P(\overline{2})$ are the decreased and increased probabilities induced by mutations, and 57 and 84 mutations result in the distribution probability decreased and increased, respectively. At the level of hemophilia severity: (i)is the $P(\overline{1}|\overline{2})$ impact probability (conditional probability) that the hemophilia severity is defined under the condition of increased distribution probability, and 51 mutations have such an effect. (ii)is the impact probability that the hemophilia severity is not defined under the condition of increased distribution probability, and 33 mutations work in such a manner. (iii) P(1|2) is the impact probability that the hemophilia severity is defined under the condition of decreased distribution probability, and 31 mutations play such a role. (iv) $P(\overline{1}|2)$ is the impact probability that the hemophilia severity is not defined under the condition of decreased distribution probability, and 26 mutations fall into this category. At the level of combined events, we can see the combined results of mutations and disease severity.

Table 2 lists the computed probabilities with respect to Fig. 1,

Fig. 1*. Cross-impact relationship among coagulation factor IX mutations, hemophilia severity, and combined results.*



from which several interesting points can be found. (i) As $P(\bar{2})$ is larger than P(2), a mutation has a large chance of increasing the distribution probability in mutant coagulation factor IX. (ii) As is $P(\bar{1}|\bar{2})$ larger than $P(1|\bar{2})$, a mutation that increases the distribution probability has six tenths chance of being defined as the severity of hemophilia. (iii) As P(1|2) is slightly larger than $P(\bar{1}|2)$, a mutation that decreases the distribution probability has more than a half a chance of being defined as the severity of hemophilia.

From these probabilities, we can use the Bayes' law [54], $P(1|2)=P(2|1)\frac{P(1)}{P(2)}$, to determine the probability that the hemophilia severity defined under a mutation, which is P(1) in this equation. As P(2) and P (1|2) can be found in cross-impact analysis, while P (2|1) is the probability that the distribution probability decreases under the condition of hemophilia severity defined.

As P(1|2) = 31/57 = 0.5439 (Table 2), and P(2|1) = 31/(51 + 31) = 0.3780, $P(1)=\frac{P(1|2)}{P(2|1)}P(2)=\frac{0.5439\times 0.4043}{0.3780}=0.5817$ namely, the patient has larger than 0.5 chance of being defined severity of hemophilia B when a new mutation is found in coagulation factor IX.

In some sense, this study is somewhat similar to the currently popular analysis, genome wide association, the difference is that the single-nucleotide polymorphism is analyzed in genome wide association, while our association is a step ahead, because we have the probability that the occurrence of disease when a single-nucleotide polymorphism is found at protein level.

ACKNOWLEDGEMENTS

This study was partly supported by Guangxi Science Foundation No. 0991080 and 07109001A.

REFERENCES

1. Furie B, Furie BC. Molecular basis of blood coagulation. Cell 1988; 53:505.
2. Furie B, Furie BC. The molecular and cellular biology of blood coagulation. N Engl J Med 1992; 326:800.
3. Kurachi K, Davie EW. Isolation and characterization of a cDNA coding for human factor IX. Proc Natl Acad Sci USA 1982; 79:6461-6464.
4. Anson DS, Choo KH, Rees DJG, Giannelli F, Gould KG, Huddleston JA, Brownlee GG. The gene structure of human anti-haemophilic factor IX. EMBO J 1984; 3:1053-1060.
5. Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). Biochemistry 1985; 24:3736-3750.
6. Roberts H, Hoffman M. Hemophilia and related conditions—inherited deficiencies of prothrombin (factor II), factor V, and factor VII to XII. In: Beutler E, Lichtman M, Coller B, Kipps T (editors). Williams Hematology. 5th ed. New York; McGraw- Hill; 1995, pp. 1415-1416.
7. Soucie JM, Evatt B, Jackson D. Occurrence of hemophilia in the United States. The Hemophilia Surveillance System Project Investigators. Am J Hematol 1998; 59:288-294.
8. Roberts H, Lozier J. Clinical aspects and therapy for hemophilia B. In: Hoffman R, Benz E, Shattil S, et al, (editors). Hematology, Basic Principles and Practice. New York; Churchill-Livingstone; 1991, pp. 1325-1331.
9. Furie B, Limentani SA, Rosenfield CG. A Practical guide to the evaluation and treatment of hemophilia. Blood 1994; 84:3-9.
10. Sommer SS. Assessing the underlying pattern of human germline mutations: lessons from the factor IX gene. FASEB J 1992; 6:2767-2774.
11. Giannelli F, Green PM, High KA, Sommer S, Poon M-C, Ludwig M, Schwaab R, Reitsma PH, Goossens M, Yoshioka A, Brownlee GG. Haemophilia B: database of point mutations and short additions and deletions -- fourth edition, 1993. Nucleic Acids Res 1993; 21:3075-3087.
12. Martlew VJ. Peri-operative management of patients with coagulation disorders. Br J Anaesthesia 2000; 85:446-455.
13. David D, Rosa HAV, Pemberton S, Diniz MJ, Campos M, Lavinha J. Single-strand conformation polymorphism (SSCP) analysis of the molecular pathology of hemophilia B. Hum Mutat 1993; 2:355-361.
14. Aguilar-Martinez P, Romey M-C, Schved J-F, Gris J-C, Demaille J, Claustres M. Factor IX gene mutations causing haemophilia B: comparison of SSC screening versus systematic DNA sequencing and diagnostic applications. Hum Genet 1994;94:287-290.
15. Wulff K, Schroeder W, Wehnert M, Herrmann FH. Twenty-five novel mutations of the factor IX gene in haemophilia B. Hum Mutat 1995; 6:346-348.
16. Heit JA, Thorland EC, Ketterling RP, Lind TJ, Daniels TM, Zapata RE, Ordonez SM, Kasper CK, Sommer SS. Germline mutations in Peruvian patients with hemophilia B: pattern of mutation in Amerindians is similar to the putative endogenous germline pattern. Hum Mutat 1998; 11:372-376.
17. Montejo JM, Magallon M, Tizzano E, Solera J. Identification of twenty-one new mutations in the factor IX gene by SSCP analysis. Hum Mutat 1999; 13:160-165.
18. Wulff K, Bykowska K, Lopaciuk S, Herrmann FH. Molecular analysis of hemophilia B in Poland:

12 novel mutations of the factor IX gene. Acta Biochim Pol 1999; 46:721-726.
19. Vidal F, Farssac E, Altisent C, Puig L, Gallardo D. Factor IX gene sequencing by a simple and sensitive 15-hour procedure for haemophilia B diagnosis: identification of two novel mutations. Br J Haematol 2000; 111:549-551.
20. Espinos C, Casana P, Haya S, Cid AR, Aznar JA. Molecular analyses in hemophilia B families: identification of six new mutations in the factor IX gene. Haematologica 2003; 88:235-236.
21. Onay UV, Kavakli K, Kilinc Y, Gurgey A, Aktuglu G, Kemahli S, Ozbek U, Caglayan SH. Molecular pathology of haemophilia B in Turkish patients: identification of a large deletion and 33 independent point mutations. Br J Haematol 2003; 120:656-659.
22. Venceslá A, Barceló MJ, Baena M, Quintana M, Baiget M, Tizzano EF. Marker and real-time quantitative analyses to confirm hemophilia B carrier diagnosis of a complete deletion of the F9 gene. Haematologica 2007; 92:1583-1584.
23. Tsai KL, Clark LA, Murphy KE. Understanding hereditary diseases using the dog and human as companion model systems. Mamm Genome 2007; 18:444-451.
24. Wu G, Yan S. Randomness in the primary structure of protein: methods and implications. Mol Biol Today 2002; 3:55-69.
25. Wu G, Yan S. Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint. Acta Pharmacol Sin 2006; 27:513-526.
26. Wu G, Yan S. Lecture Notes on Computational Mutation. New York; Nova Science Publishers; 2008.
27. http://expasy.org/uniprot/P00740., Accession number P00740; updated to January 15, 2008; entry version 134.
28. Feller W. An Introduction to Probability Theory and Its Applications. 3rd ed, Vol, I. New York; Wiley; 1968, pp. 34-40.
29. Gao N, Yan S, Wu G. Pattern of positions sensitive to mutations in human haemoglobin □-chain. Protein Pept Lett 2006; 13:101-107.
30. Wu G, Yan S. Prediction of distributions of amino acids and amino acid pairs in human haemoglobin □-chain and its seven variants causing □-thalassemia from their occurrences according to the random mechanism. Comp Haematol Int 2000; 10:80-84.
31. Wu G, Yan S. Analysis of distributions of amino acids, amino acid pairs and triplets in human insulin precursor and four variants from their occurrences according to the random mechanism. J Biochem Mol Biol Biophys 2001; 5:293-300.
32. Wu G, Yan S. Analysis of distributions of amino acids and amino acid pairs in human tumor necrosis factor precursor and its eight variants according to random mechanism. J Mol Model 2001; 7:318-323.
33. Wu G, Yan S. Random analysis of presence and absence of two- and three-amino-acid sequences and distributions of amino acids, two- and three-amino-acid sequences in bovine p53 protein. Mol Biol Today 2002; 3:31-37.
34. Wu G, Yan S. Analysis of distributions of amino acids in the primary structure of apoptosis regulator Bcl-2 family according to the random mechanism. J Biochem Mol Biol Biophys 2002; 6:407-414.
35. Wu G, Yan S. Analysis of distributions of amino acids in the primary structure of tumor suppressor p53 family according to the random mechanism. J Mol Model 2002; 8:191-198.
36. Wu G, Yan S. Determination of sensitive positions to mutations in human p53 protein. Biochem Biophys Res Commun 2004; 321:313-319.
37. Wu G, Yan S. Searching of main cause leading to severe influenza A virus mutations and consequently to influenza pandemics/epidemics. Am J Infect Dis 2005; 1:116-123.
38. Wu G, Yan S. Prediction of mutation trend in hemagglutinins and neuraminidases from influenza A viruses by means of cross-impact analysis. Biochem Biophys Res Commun 2005; 326:475-482.
39. Wu G, Yan S. Timing of mutation in hemagglutinins from influenza A virus by means of amino-acid distribution rank and fast Fourier transform. Protein Pept Lett 2006; 13:143-148.
40. Wu G, Yan S. Prediction of possible mutations in H5N1 hemagglutinins of influenza A virus by means of logistic regression. Comp Clin Pathol 2006; 15:255-261.
41. Wu G, Yan S. Prediction of mutations in H5N1 hemagglutinins from influenza A virus. Protein Pept Lett 2006; 13:971-976.
42. Wu G, Yan S. Improvement of model for prediction of hemagglutinin mutations in H5N1 influenza viruses with distinguishing of arginine, leucine and serine. Protein Pept Lett 2007; 14:191-196.
43. Wu G, Yan S. Improvement of prediction of mutation positions in H5N1 hemagglutinins of influenza A virus using neural network with distinguishing of arginine, leucine and serine. Protein Pept Lett 2007; 14:465-470.
44. Wu G, Yan S. Prediction of mutations engineered by randomness in H5N1 neuraminidases from influenza A virus. Amino Acids 2007;34:81-90.
45. Wu G, Yan S. Prediction of mutations in H1 neuraminidases from North America influenza A virus engineered by internal randomness. Mol Divers 2007; 11: 131-140.
46. Wu G, Yan S. Prediction of mutations initiated by internal power in H3N2 hemagglutinins of influenza A virus from North America. Int J Pept Res Ther 2008; 14: 41-51.
47. Wu G, Yan S. Prediction of mutation in H3N2 hemagglutinins of influenza A virus from North America based on different datasets. Protein Pept Lett 2008; 15: 144-52.
48. Gordon TG, Hayward H. Initial experiments with the cross-impact matrix method of forecasting. Futures 1968; 1:100-116.
49. Gordon TG. Cross-impact matrices – an illustration of their use for policy analysis. Futures 1969;

2:527-531.
50. Enzer S. Delphi and cross-impact techniques: an effective combination for systematic futures analysis. Futures 1970; 3:48-61.
51. Enzer S. Cross-impact techniques in technology assessment. Futures 1970; 4:30-51.
52. Sage AP. Methodology for Large-Scale Systems. New York; McGraw-Hill; 1977, pp. 165-203.
53. Wu G. Application of cross-impact analysis to the relationship between aldehyde dehydrogenase 2 and flushing. Alcohol Alcohol 2000; 35:55-59.
54. Wikipedia. Bayes' theorem. en.wikipedia.org/wiki/Bayes'_theorem. 2008.