Network modeling. To build the BN we used a popular Bayesian approach that was developed by Cooper and Herskovitz[2] and is implemented in the program Bayesware Discoverer (www.bayesware.com). The program searches for the most probable network of dependency given the data. To find such a network, the program explores a space of different network models, scores each model by its posterior probability conditional on the available data, and returns the model with maximum posterior probability. This probability is computed by Bayes' theorem as
$p(M \mid D) \propto p(D \mid M)\, p(M)$,
where $p(D \mid M)$ is the probability that the observed data are generated from the network model *M*, and $p(M)$ is the prior probability encoding knowledge about the model *M* before seeing any data. We assumed that all models were equally likely a priori, so that $p(M)$ is uniform and the posterior probability $p(M \mid D)$ becomes proportional to $p(D \mid M)$, a quantity known as *marginal* likelihood. The marginal likelihood averages the likelihood functions for different parameters values and it is calculated as
$p(D \mid M) = \int p(D \mid \theta)\, p(\theta) d\theta$
where $p(D \mid \theta)$ is the traditional likelihood function and *p(θ)* is the parameter prior density. The set of marginal and conditional independences represented by a network M imply that, for categorical data in which *p(θ)* follows a Dirichlet distribution and with complete data, the integral $p(D \mid M) = \int p(D \mid \theta)\, p(\theta) d\theta$ has a closed form solution[3] that is computed in product form as:
$p(D \mid M) = \prod_i p(D_i \mid M_i)$
where M_i is the model describing the dependency of the ith variable on its parent nodes – those node with directed arcs pointing to the ith variables-- and D_i are the observed data of the i*th* variable[3]. Details of the calculations are in [4].
The factorization of the marginal likelihood implies that a model can be learned locally, by selecting the most probable set of parents for each variable, and then joining these local structures into a complete network, in a procedure that closely resembles standard path analysis. This modularity property allows us to assess, locally, the strength of local associations represented by rival models. This comparison is based on the *Bayes factor* that measures the odds of a model M_i versus a model $\tilde{M}_i$ by the ratio of their posterior probabilities $p(M_i \mid D_i) / p(\tilde{M}_i \mid D_i)$ or, equivalently, by the ratio of their marginal likelihoods $\rho = p(D_i \mid M_i) / p(D_i \mid \tilde{M}_i)$. Given a fixed structure for all the other associations, the posterior probability $p(D_i \mid M_i)$ is $p(D_i \mid M_i) = \rho / (1 + \rho)$ and a large Bayes factor ρ implies that the probability $p(D_i \mid M_i)$ is close to 1, meaning that there is very strong evidence for the associations described by the model M_i versus the alternative model $\tilde{M}_i$. Note that, when we explore different dependency models for the ith variable, the posterior probability of each model depends on the same data D_i.

Even with this factorization, the search space is very large and to reduce computations, we used a bottom-up search strategy known as the K2 algorithm[2].The space of candidate models was reduced by first limiting attention to diagnostic rather than prognostic models, in which we modeled the dependency of SCD complications and

laboratory variables on death. We also ordered the variables according to their variance, so that less variable nodes could only be dependent on more variable nodes. Simulations results we have carried out suggest that this heuristic leads to better networks with largest marginal likelihood. As in traditional regression models, in which the outcome (death) is dependent on the covariates, this inverted dependency structure can represent the association of independent as well as interacting covariates with the outcome of interest [3]. However, this structure is also able to capture more complex models of dependency [5] because, in this model, the marginal likelihood measuring the association of each covariate with the outcome is functionally independent of the association of other covariates with the outcome. In contrast, in regression structures, the presence of an association between a covariate and the outcome affects the marginal likelihood measuring the association between the phenotype and other covariates, reducing the set of regressors that can be detected as associated with the variable of interest.

The BN induced by this search procedure was quantified by the conditional probability distribution of each node given the parents nodes. The conditional probabilities were estimated as

$$p(x_{ik} \mid \pi_{ij}) = \frac{\alpha_{ijk} + n_{ijk}}{\alpha_{ij} + n_{ij}}$$

where x_{ik} represents the state of the child node, π_{ij} represents a combination of states of the parents nodes, n_{ijk} is the sample frequency of (x_{ik} , π_{ij}) and n_{ij} is the sample frequency of π_{ij} . The parameters α_{ijk} and $\alpha_{ij} = \sum_k \alpha_{ijk}$ encode the prior distribution with the constrain $\sum_j \alpha_{ij} = \alpha$ for all j, as suggested in[3]. We chose $\alpha = 16$ by sensitivity analysis[3].

The network highlights the variables that are sufficient to compute the score: these are the variables that make the risk of death independent of all the other variables in the network and appear in red in Figure 1. These variables are the “Markov blanket” of the node death as defined in [3].

1. West, M.S., et al., *Laboratory profile of sickle cell disease: a cross-sectional analysis. The Cooperative Study of Sickle Cell Disease.* J Clin Epidemiol, 1992. 45(8): p. 893-909.

2. Cooper, G.F. and G.F. Herskovitz, *A Bayesian method for the induction of probabilistic networks from data.* Mach Learn, 1992. 9: p. 309-347.

3. Cowell, R.G., et al., *Probabilistic Networks and Expert Systems.* 1999, New York: Springer Verlag.

4. Sebastiani, P., M. Abad, and M.F. Ramoni, *Bayesian networks for genomic analysis*, in *Genomic Signal Processing and Statistics*, E. Dougherty, et al., Editors. 2005. p. 281-320.

5. Hand, D.J., H. Mannila, and P. Smyth, *Principles of Data Mining*. 2001, Cambridge, MA: MIT Press.