"Information is the resolution of uncertainty" Whole Genome approaches to genetic diagnosis on the PICU.

Mark J Peters [1,2] MD PhD, Hywel J. Williams PhD[3]

1. Respiratory Critical Care and Anaesthesia Unit, UCL Great Ormond Street Institute of Child Health, London, United Kingdom.
2. Paediatric Intensive Care Unit, Great Ormond Street Hospital, London, United Kingdom.
3. Genetic and Genomic Medicine, Cardiff University, Cardiff, United Kingdom.

Corresponding author
mark.peters@ucl.ac.uk

Key words
Whole genome sequencing, critical illness, whole exome sequencing, diagnosis, rare disease

This work received no direct funding but was supported by the National Institute for Health Research Great Ormond Street Hospital Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the UK Department of Health.

Declaration of Interests

No conflicts of interest to declare.

Word count 1028

Sherwin B. Nuland, surgeon and bioethicist wrote "*To become comfortable with uncertainty is one of the primary goals in the training of a physician"*. (1) Whether we are comfortable with it or not, uncertainty is a feature of our work. We are rarely certain about the risk and benefits of therapy, we struggle to estimate prognosis and, often, diagnoses are elusive. We use clinical and laboratory information to try to reduce this uncertainty.

In this context is it worth remembering what '*information'* means formally. Claude Shannon, the Mathematician, Electrical Engineer and the father of Information Theory, defined it as '*a measure of the number of possible alternatives*.' (2) The 'bits' that we count in their billions in our computers and smartphones are choices between two states: '0 and 1' or 'true' and 'false'. The bit is therefore the lowest possible unit of information. In other words, there is no 'information' without an alternative.

Intensivists face a huge number of potential alternatives. Acquisition of information reduces this number of potential states and resolves some of this uncertainty. Inherited genetic disease is an example where the number of alternative states is both large and, as yet, not fully defined. The On-line Medelian Inheritance in Man (OMIM) database (https://www.omim.org) as of 15th June 2019 lists more than 7,000 phenotypes and 16,092 genes. Traditional approaches of testing for a focussed panel of genes based on phenotype, requires the expertise from our clinical genetics colleagues. Their expertise is information that reduces the number of alternatives. However, we often do not make diagnoses in our patients despite high levels of suspicion of a genetic basis for disease.

An alternative approach, is the more 'brute force' approach to collecting information of whole genome sequencing. (3-7)

In this issue of the journal the paper Wu et al describe such an approach to genetic diagnosis in undiagnosed critically-ill children and newborns in Taiwan. (8) They used whole exome sequencing in 40 children and newborns with suspected genetic disease. Some 50-100,000 variants were obtained. Eleven of the 40 (27.5%) were diagnosed with previously reported mutations and another 8 had strong pointers to a new diagnosis. Mean time to a diagnosis was 6.2 days (range 4.3-9). When describing these new diagnoses the authors state: *Many were [*the*] first example of these diseases recognized in Taiwan*.

This is the hallmark an extraordinary powerful diagnostic technique – to identify diseases that the clinical staff may not be considering. Humans are biased by personal experience more than formal probabilities. In Bayesian terms the 'prior probability' of these diseases as assessed by the team may have been close to zero. The post-test probability however is extremely high. We rarely have the benefit of such extreme modification of the odds of a disease by a single result. The second important observation in the paper is that the information obtained by this technique often had value. The diagnostic information reduced the alternatives states so far as to prompt a change in therapy or goals of care in 10 of the cases.

Before we offer up our clinical geneticists' colleagues' salaries to our institutions as possible cost-savings, we should consider the potential sources of bias in this approach. Individual patients are selected. The criteria applied by Wu et al: *suspicion of a genetic condition* in

critically ill children (or newborns at screening), are wide. The level of suspicion will vary between observers based on personality and experience. This might include a judgement of how abnormal a plasma ammonia level is in a sick newborn. This judgement will change with time– as will the performance of genetic testing at different points in the natural history. The background genetic diversity of the study population will also influence performance. Taiwanese results might not be reproduced in London, Philadelphia or Cape Town.

The choice of WES – examining the coding element of the genome only – rather than its big brother: whole genome sequencing (WGS) is a potential source of bias. A recent review concluded that *"from a clinical / technical point of view, WGS is the better*[approach]" in that it provides more complete coverage of the genome. (9)

Both techniques present a challenge of how to filter tens of thousands (WES) to millions of variants (WGS) down to the specific ones causing disease. This filtering cascade is subject to bias given our incomplete understanding of rare disease genetics.

Most known rare disease genes result from the identification of pathogenic coding variants. We can annotate and interpret these coding 'protein-altering' variants accurately. Analysis of large public genetic databases allows us to estimate these allele frequencies. Hence we can remove common variants since these are not compatible with the rare incidence of the disease. Combining just these two filters (protein-altering and rare frequency) removes >95% of the variants generated in most WGS studies.

Having parental data available permits further filtering on inheritance patterns. Finally, a panel is applied so that only variants within genes thought to be relevant to the phenotype are evaluated. This can be a powerful filter but also a potential major source of bias. Wu at al describe a text-mining approach to phenotyping with few details but including putting descriptive key words into pubmed. This prone to human selection and representation bias – where features of our own mental model of a genetic condition are more likely to be searched for. A more objective approach might be the use of artificial intelligence led scanning of patient's health-data to focus on a bespoke gene panel. This method has reported some success. (10)

Even with optimal WGS filtering, diagnostic rates are typically <50%. To improve this, we will have to devise additional methods to explore our data. The ways to do this are exhaustive but include annotation of the non-coding (98% of the total) genome, consideration of polygenic causations (11) and accurate identification of complex structural variants (12).

We have never had so much information available to us as intensivists. We can resolve more uncertainty than ever before. At the same time, we are better appreciating the huge scale of our residual uncertainty. Managing that uncertainty will remain one of our main roles for now at least.

References

1. Nuland SB: The Uncertain Art. Random House; 2008.

2. Shannon CE: **A mathematical theory of communication**. *The Bell System Technical Journal* 2014; 27:379–423

3. Kernohan KD, Hartley T, Naumenko S, et al.: Diagnostic clarity of exome sequencing following negative comprehensive panel testing in the neonatal intensive care unit. *Am J Med Genet A* 2018; 176:1688–1691

4. Sweeney NM, Nahas SA, Chowdhury S, et al.: The case for early use of rapid whole-genome sequencing in management of critically ill infants: late diagnosis of Coffin-Siris syndrome in an infant with left congenital diaphragmatic hernia, congenital heart disease, and recurrent infections. *Cold Spring Harb Mol Case Stud* 2018; 4:a002469

5. Smith EE, Souich du C, Dragojlovic N, et al.: Genetic counseling considerations with rapid genome-wide sequencing in a neonatal intensive care unit. *J Genet Couns* 2019; 28:263–272

6. French CE, Delon I, Dolling H, et al.: Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med* 2019; 45:627–636

7. Mestek-Boukhibar L, Clement E, Jones WD, et al.: Rapid Paediatric Sequencing (RaPS): comprehensive real-life workflow for rapid diagnosis of critically ill children. *J Med Genet* 2018; 55:721–728

8. Wu E-T: Critical trio exome benefits in-time decision making for pediatric patients with severe illnesses. *Pediatr Crit Care Med*

9. Meienberg J, Bruggmann R, Oexle K, et al.: Clinical sequencing: is WGS the better WES? *Hum Genet* 2016; 135:359–362

10. Clark MM, Hildreth A, Batalov S, et al.: Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Science Translational Medicine* 2019; 11:eaat6177

11. Niemi MEK, Martin HC, Rice DL, et al.: Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* 2018; 562:268-271

12. Sanchis-Juan A, Stephens J, French CE, et al.: Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing.*Genome Med* 2018; 10:95