Wolfson School of Mechanical, Electrical and Manufacturing Engineering

Investigating UK Biobank blood metrics variation to inform cell therapy manufacturing process control

Jamie Thurman-Newell, Jon Petzing, David Williams

Loughborough University

Research Aims

To investigate whether UK Biobank blood metric data shows variation in White Blood Cell count that can act as a healthy population benchmark for Stem Cell Therapy manufacturing process control

Cell therapies are becoming increasingly viable on a clinical basis, to the extent where upscale to centralised or distributed manufacture is possible. Manufacturing processes and control paradigms have been traditionally predicated on minimal variation of the product and the capability to control the manufacturing process using product limits. Automation within traditional manufacture that has adhered to this philosophy has led to in-process, statistical process control. Whilst there is a requirement to translate and scale up the production of cell therapies, the manufacturing processes need to understand and accommodate the issues associated with, and context of, using cells as products.

Conventional manufacturing process control would normally be defined as typically 10% to 30% of the defined process limits i.e. less than 0.1 orders of magnitude. Existing work has already defined significant variation of starting materials in Haematopoetic Stem Cell (HSC) therapy exemplars [1, 2, 3] that may be 3 to 6 orders of magnitude variation. But there is now a need to examine a baseline healthy population for variation in blood metrics to further inform potential manufacturing process control paradigms.

Analysis Methodology

UK Biobank [4] was petitioned for access to 502,664 patient records (Table 1). Access to blood metrics and associated data tags were obtained, with emphasis on; age, health state, gender, weight, centre location. These a were prioritised to align with existing analyses completed on multi site and single site clinical centre data [1, 2, 3].

At the point that this investigation was carried out, UK Biobank did not directly measure cell markers that specifically identified stem cells such as HSCs. CD34+ cell counts would have been appropriate to directly compare the relatively healthy population of the UK Biobank to the relatively unwell populations in multi and single centre clinical analysis [1, 2, 3]. UK Biobank does however record red blood cell counts, WBC counts and the specific subsets of WBCs such as leukocytes and monocytes. Biological variation was therefore explored as a function of WBCs rather than and as a surrogate to HSCs.

The comma-separated variation output was processed in IBM's SPSS 22.0 statistics package. Distributions of cell metrics were analysed for normality. The minimum, maximum and range data were defined along with Standard Deviation and Variance.

These biological measurements were found to be non-parametric and consequently non-parametric statistical tools were used to compare median and distribution of the data. The Pearson product-moment correlation co-efficient was used to measure linear correlation. Independent samples median tests, independent Mann-Whitney tests (for two independent groups, such as gender) and independent sample Kruskal-Wallis (for multiple independent groups) tests were used to measure whether the median or the distribution of population differs significantly between groups. Example results for variation as a function of age, weight, and UK Assessment Centre location are shown in Figures 1, 2 and 3 respectively.

		Valid n	Minimum	Maximum	Mean	Median
Unique ID		502,664	-	-	-	-
Gender	Male	229,182	-	-	-	-
	Female	273,467	-	-	-	-
Birth Year		502,649	1934	1971	1952	1950
Age		502,649	45	82	64.46	66.00
Weight (kg)		499,874	30.0	197.7	78.1	76.4
BMI		499,543	12.12	74.68	27.43	26.74

Table 1: Summary of Participants within UK Biobank [4]

Figure 1: The spread of WBC count per litre measured per individual participant, against age in years

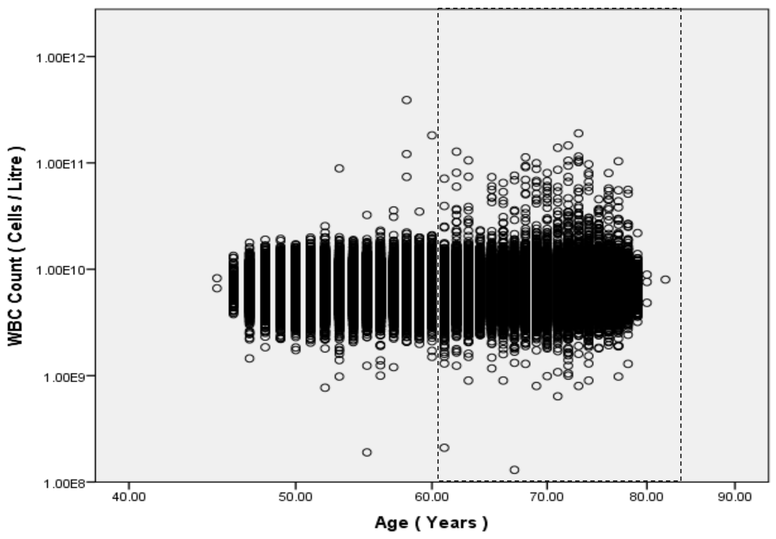


Figure 2: Calculated WBC count / litre per individual as a function of weight in kilograms

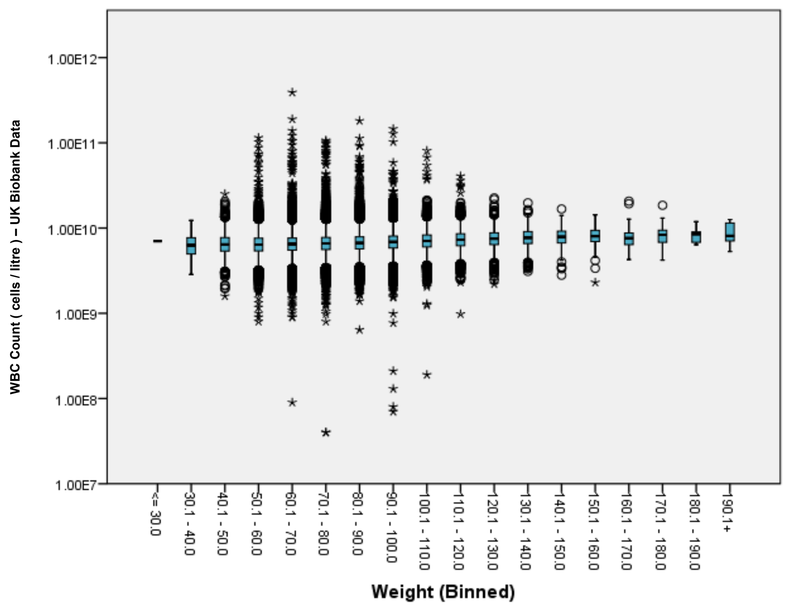
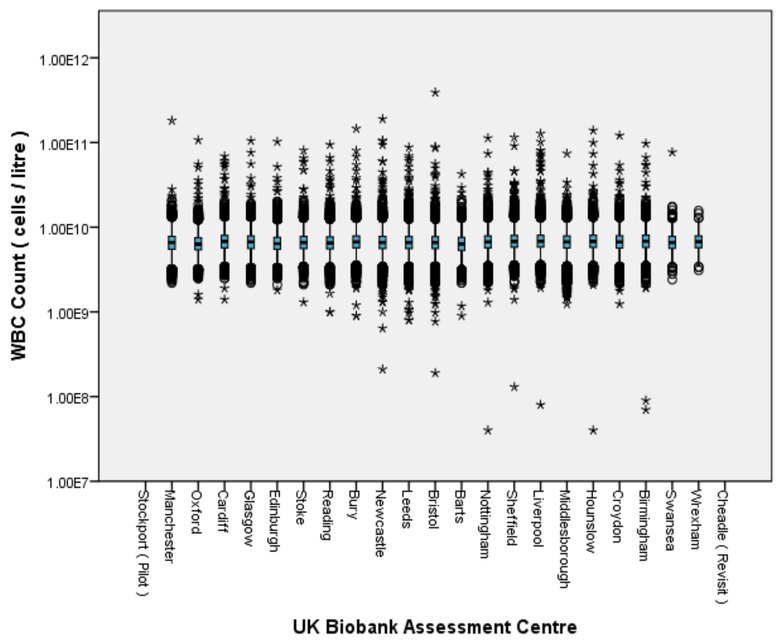


Figure 3: WBC count / litre per individual as a function of location of UK Assessment Centre



Key Results

An important assumption that has been made is that the UK Biobank (500,000+ records) is a representative sample of the UK general population. The original data gathering studies were carried out on a volunteer-basis that may have influenced the population – people of a mindset that were more inclined to volunteer, and may reflect a particularly health state, lifestyle or character.

The variation in WBC count can be between 7.00 x 10^7 cells/litre and 3.90 x 10^{11} cells/litre – a variation of up to four orders of magnitude. Both the TNC count and WBC counts within the UK Biobank had similar ranges (range; 3.90 x 10^{11} and 3.86 x 10^{11} cells / litre), similar distributions around the mean (standard deviation; 2.12 x 10^9 cells / litre and 2.10 x 10^9 cells / litre), and similar data spread (variance; 4.51 x 10^{18} cells / litre and 4.42 x 10^{18} cells / litre). This analysis aligns with the lower end of variation seen in the multi site and single site clinical centre analysis [1, 2, 3].

The population characteristics examined in this research (age (Figure 1), disease state, gender, weight (Figure 2) and, geographical location (Figure 3)) all appear to influence the biological variation encountered but there is an implication this relates to a multiplicative effect based on multiple, interacting variables, rather than the result of a handful of key metrics. However, there are several potentially influential metrics that were not examined which may play a key role e.g. alcohol consumption, diet, concurrent medication and smoking.

A notable observation from this analysis is the considerable number of data 'outliers'. These outliers represent a state of health, keyed uniquely to be the individual, and cannot be removed to smoothen a statistical distribution. Understanding these outliers, and what they represent, is critical to understanding the behaviour of the starting material and the process; it will influence the way in which cellular therapies are manufactured and monitored, and act as a guide as to whether biological variation should be accommodated or should be controlled.

Conclusions

- White Blood Cell (WBC) count can vary up to 4 orders of magnitude (7.00 x 107 cells/litre and 3.90 x 1011 cells/litre)
- Gender appears to impact the amount of variation in WBC and TNC counts
- Increasing age shows a trend towards increasing variation in WBC and TNC counts
- Increasing weight shows a trend towards decreasing variation in WBC and TNC counts
- UK Assessment Centre, analogous of geographical area, appears to have a statistically significant effect on WBC counts
- Significant outliers exists in the data that represent a state of health, keyed uniquely to be the individual, and cannot be removed to smoothen a statistical distribution.

REFERENCES

1. Thurman-Newell JA, Petzing J, Williams D. Quantification of biological variation in blood-based therapies: A summary of a meta-analysis to inform manufacturing in the clinic, Vox Sanguinis. 109(4), 394-402 (2015).
2. Thurman-Newell JA, Petzing J, Williams D. A meta-analysis of biological variation in blood-based therapy as a precursor to biomanufacturing. Cytotherapy. 18, 686-694 (2016).
3. Thurman-Newell JA, Petzing JN, Williams DJ. Analysis of biological variation in blood-based therapy as a precursor to biomanufacturing, MIT Center for Biomedical Innovation – 2015 BioMAN Summit, Boston, USA, 18th – 19th November 2015.
4. UK BioBank, http://www.ukbiobank.ac.uk

ACKNOWLEDGEMENTS
This research is kindly supported by the the UK EPSRC/MRC Doctoral Training Centre for Regenerative Medicine at Loughborough University (EP/L105072/1), and UK Biobank for access to UK biometric datasets.

CONTACT INFORMATION
Wolfson School
Loughborough University
Leicestershire LE11 3TU UK
E-mail: R.Grant@lboro.ac.uk