Planning epidemiologic and molecular studies for a Portuguese cohort of Multiple Myeloma

Gisela Ferreira, MD[1,2]; Vera Afreixo, PhD[3,4*]; Raquel M Silva, PhD[2]

[1]Haematology Service, Centro Hospitalar do Baixo Vouga, EPE, Aveiro, Portugal
[2]Department of Medical Sciences & iBiMED, University of Aveiro, Portugal
[3]Department of Mathematics & CIDMA, University of Aveiro, Portugal
[4]IEETA, University of Aveiro, Portugal

*Corresponding author: Raquel M Silva, raquelsilva@ua.pt

Keywords: Epidemiology, Multiple Myeloma, Sample size

Introduction

Multiple Myeloma (MM) is the second most common hematologic malignancy. It is a chronic plasma cell neoplasm that typically occurs in older adults (over 60 years old) and has significant morbidity caused by end-organ destruction. MM is preceded by a pre-malignant and asymptomatic condition named Monoclonal Gammopathy of Undetermined Significance (MGUS). Defining the causes and molecular mechanisms underlying MGUS progression to MM is essential for better prevention strategies, motivating the project here proposed.

Our main question is to evaluate the impact of environmental factors in the development of MM. A case control retrospective study is planned and will be developed using CHBV patient and control samples from the hematology service, where it is estimated that about 20 new patients (MM) per year are identified. The controls will be age-matched patients followed in Immuno-Hemotherapy consultation excluding malignant hematologic conditions.

To submit the study protocol to the different ethics committees, it is necessary to include the study design and it is usual to identify the sample size, often using statistical techniques to estimate it. The objective of the present work is to discuss this part of the planning process.

Methods

The main issue associated with planning is to assess the effect of environmental toxicity on the development of MM, and it will be assumed that this effect can be quantified by combining personal information with environmental data. From the review

of the literature, no information was found to identify the variable capable of measuring the environmental toxicity to which each individual was exposed, and of course the effect of this toxicity on the occurrence of MM has not yet been estimated. To measure the environmental toxicity for each individual, a measure of the exposure / exposure time to different toxicants will be created.

Considering that we anticipate a large number of individuals available in the control group, and to reduce the bias associated with possible confounding factors, we intend to perform a propensity score matching, and work with more homogeneous groups (patient vs. control). In this context a balanced planning will be chosen (the number of individuals in the groups will be similar).

Results

If we assume that the effect of environmental toxicity is moderate (since there is no previous information) and that we intend to test the parametric hypotheses at a significance level of 5% and a power of 80%, we would need about 65 individuals in each group (n1+n2=65+65=130). For the reality of CHBV this is a very large number. Then, how should we proceed?

Can we discuss other alternative numbers for the size of each group? From previous experience we have the perception that we could prepare a suitable scenario for what can be collected during the study period. Is this approach methodologically correct? When calculating the sample size we can have many options from the definition of the main hypothesis, the type of statistical test used to evaluate (e.g. parametric vs non-parametric, one- vs two-tailed), fixed effect size, Type I and Type II errors associated to the hypothesis tests. In Table1, some values of the sample size are registered for balanced scenario with the t-test for comparison of two independent groups, and we can observe that depending on the chosen options the suggested size can vary from 26 to 651, per group.

We could also take advantage of the ease of collecting controls to advocate unbalanced planning that globally requires more of the overall size of the sample (n1 + n2) but would allow for a smaller patient size.

If we chose a non-parametric approach with the same options discussed above, we could increase the size of the estimated sample (Table2).

Discussion and Conclusions

If the main research questions identified were different, e.g.: i) To evaluate the effect of environmental toxicity on different hematological pathologies such as MM, MDS (Myelodysplastic Syndromes) vs control; ii) Assess the effect of toxicity on disease

severity; iii) To evaluate the association of genetic variants and environmental toxicity with the occurrence of the pathology with / without interaction between the possible factors; we might have to increase the sample size even further, as these hypotheses suggest, for example, the comparison of more than two groups (parametric vs non-parametric), generalized linear models (e.g., logistic regression).

For unbiased effect size calculations, several statistical parameters that will be only obtained after have to be inferred "a priori". Discussing these questions is important for appropriate project planning and for better study design.

Table 1: Sample size estimation using G*Power 3.1.9.2.

significance level	power	effect size (d)	test type	tail(s)	n2/n1	n1 patients	n2 control
0.05	0.95	0.2	t-test	two	1	651	651
0.05	0.8	0.2	t-test	two	1	394	394
0.05	0.95	0.5	t-test	two	1	105	105
0.05	**0.8**	**0.5**	**t-test**	**two**	**1**	**64**	**64**
0.05	0.95	0.6	t-test	two	1	74	74
0.05	0.8	0.6	t-test	two	1	45	45
0.05	0.95	0.7	t-test	two	1	55	55
0.05	0.8	0.7	t-test	two	1	34	34
0.05	0.95	0.8	t-test	two	1	42	42
0.05	0.8	0.8	t-test	two	1	26	26

Table 2: Sample size estimation using G*Power 3.1.9.2. Parent distribution, min ARE; test type, Wilcoxon-Mann-Whitney test.

significance level	power	effect size (d)	tail(s)	n2/n1	n1 patients	n2 control
0.05	0.95	0.2	two	1	754	754
0.05	0.8	0.2	two	1	456	456
0.05	0.95	0.5	two	1	122	122
0.05	**0.8**	**0.5**	**two**	**1**	**74**	**74**
0.05	0.95	0.6	two	1	85	85
0.05	0.8	0.6	two	1	52	52
0.05	0.95	0.7	two	1	63	63
0.05	0.8	0.7	two	1	39	39
0.05	0.95	0.8	two	1	49	49
0.05	0.8	0.8	two	1	30	30

Acknowledgments

RMS is funded by national funds (OE), through FCT – Fundação para a Ciência e a Tecnologia, I.P., in the scope of the framework contract foreseen in the numbers 4, 5 and 6 of the article 23, of the Decree-Law 57/2016, of August 29, changed by Law 57/2017, of July 19. This work is partially funded by FEDER (Programa Operacional Factores de Competitividade-COMPETE) and by National Funds through the FCT, in the context of the projects UID/BIM/04501/2013 and POCI-01-0145-FEDER-007628 to iBiMED, and UID/MAT/04106/2019 to CIDMA.