NCS Non-Clinical Statistics Conference
Padua, IT / 23 – 25 September 2026

**TITLE: Identification of cell type aggregates in spatial transcriptomics data**

**SPEAKER:** Koen Van den Berge[1]

**COAUTHORS:** Ewoud De Troyer[1], Sofía Prieto León[2], Helena Geys[1,2], Olivier Thas[2]

[1] Statistics and Decision Sciences, Johnson and Johnson Innovative Medicine, Beerse, Belgium; [2] Data Science Institute, Hasselt University, Hasselt, Belgium

**ABSTRACT:** Contemporary spatial transcriptomics technology allows for high-throughput measurement of gene expression in tissues, at subcellular resolution, while preserving cellular organization. This effectively allows researchers to study gene expression in cells within their original context, hence also unraveling tissue organization at the level of cell states and cell types. While the potential is undeniable, answering complex questions involving the spatial organization of cells requires a new set of high-dimensional statistical analysis paradigms capable of incorporating spatial information.

A commonly reoccurring question in the analysis of spatial transcriptomics datasets concerns the grouping of cell types or marker genes in space, i.e., whether the distribution of a cell type in the tissue is more clustered between conditions; and if so, to what degree and where. For example, in inflammatory bowel disease, it is hypothesized that B-cell aggregates impair healing of inflamed epithelial tissue, and identification of these aggregates is of interest. Standard analysis practice to identify such aggregates adopts methodology originally developed in the field of ecology, where data is typically sparser and of lower dimensionality, including the Getis-Ord test statistics [Getis & Ord (1992)].

We will show that commonly adopted approaches suffer from technical artifacts, and that the typically assumed asymptotic null distribution of the widely used Getis-Ord test statistic is inappropriate in the context of spatial transcriptomics. In terms of interpretation, we show that the test statistic does not distinguish between association in space and spatial autocorrelation, and clarify how and when these different null hypotheses can be assessed. For statistical inference, we also discuss the use of permutation approaches.

Finally, we demonstrate the methodology to detect immune cell infiltrates in inflammatory bowel disease, and how these change upon treatment.

**BRIEF SPEAKER BIO:** Koen Van den Berge holds Master's degrees in Biology and Statistics, and has earned a PhD in statistical genomics from Ghent University in 2019. He performed postdoctoral research at the University of California, Berkeley, and Ghent University, where he developed statistical methods for normalization and interpretation of high-throughput sequencing datasets. Since 2021, he works as statistician at Johnson & Johnson, where he supports the analysis of 'omics data in the context of drug discovery projects across therapeutic areas.