

François COLLIN, PhD

Statistical Consultant

Katowice, Poland

☎ 0048 787 906 750

✉ fca.collin@gmail.com

📄 FCACollin.github.io/Latarnia



*Statistician with experience in biostatistics
– Medical sciences, genetics, plant physiology, geostatistics –*

Experience

Today **Senior R programmer**, *FCA Collin for UCB*, Remote/international team, involved in post-hoc analysis using R, 1.5 years (ongoing).

I did:

- Increase awareness of the developer mindset for statistical programming (e.g. distributed version control, git branching workflow, CI/CD, Agile-inspired tool-set, automated documentation).
- Develop new statistical programming outputs using R: improved graphics, reactive outputs (Shiny App), interactive outputs, videos, automated reporting and scripts.
- Develop prototypes: visualization catalog, GxP-supporting packages (e.g. standardization of validation with R, standard annotations for static outputs).
- Propose a concept: *analysis as a package* with automatic testing and automatic extensive documentation.
- Enhance connections between statistical programmers, IT developers and end-users with emphasize on automation and reproducibility.
- Support general R coding upskill (e.g. training, code-review, internal technical presentations).
- Initial role: request-based statistical programming.

Dec 2020 **Senior R programmer**, *FCA Collin for Roche*, Remote/international team (from
Apr 2020 Warsaw to San Francisco) following **Agile process**, involved in R package development from design to execution, 9 months.

- Programming R-base suite for statistical analysis of clinical data.
- From clinical standards STDM/ADaM and analysis to Tables, Listings and Graphs (TLG).
- R for visualisation: base graphs, ggplot2, code new grid elements, Shiny Apps for interactive applications and dashboards, interactive html/js graphs.
- R for statistics: function development for standard outputs (e.g. mixed model for repeated measurements, Cox PH regressions, binary response analysis).
- R in a project: package development with continuous integration and development with Jenkins (CI/CD), git-branching workflow, team convention and good practices for robust code.
- Large autonomy granted, proactive attitude to create value was encouraged: new ideas, identification of risks/opportunities, generate new designs and prototypes.

- Today **Independent consultant**, *Francois Collin*, Katowice Poland, 2.5 years (ongoing).
- Jan 2020 **Expertise** academic observational studies, academic real world evidence, genetics biomarker identification, predictive algorithms, publication support, reporting automation.
- **Portfolio**: <https://fcacollin.github.io/Latarnia>
 - **Recent misc. assignments**:
 - (2021) Real-World Evidence - Multiple Sclerosis (MS): analysis of Polish mortality data - MS as primary cause of death.
 - (2021) Real-World Evidence - Multiple Sclerosis (MS): Associations between Immunoglobulin against S or P SARS-COV-2 proteins in a pool of RRMS patients.
 - (2021) Clinical study - Investigate the relations between bariatric surgery, endocannabinoid systems and diabetes. Scope: SAP and execution for lab test / vital signs data and efficacy evaluation.
 - (2020) Clinical study - micro-RNA and diabetes. Scope: SAP for Next Generation Sequencing data and execution covering differential expression, classification, patient profiling, building biomolecular networks and research of genetic candidate biomarkers.
 - (2020) Real-World Evidence - Multiple Sclerosis (MS): analysis of public electronic health records for the evaluation of MS therapies in Poland from 2014 to 2018. Two recent developments: the JCV seroconversion for patient receiving natalizumab, the disease progression independent of relapse activity.
- Dec 2019 **Human Genetics/Expert in Sciences and Technology**, *the Medical University of Białystok (Poland)*, Centre for bioinformatics and data analysis, 1.5 years.
- Aug 2018 Research support for data analysis, machine-learning models and publication.
- **Lung cancer - Next Generation Sequencing (NGS)** for prognosis, **biomarker**, transcriptome for the prediction of type and stage of tumor implemented in an interactive and reactive prototype.
 - **Multiple Sclerosis - Observational study** for post-registration therapy evaluation.
 - **Diabetes - NGS/Clinical study**: research of therapeutic targets via analysis of RNA-seq data.
 - **Teaching**: 8-hour practical for **RNA-seq data analysis**, PhD student co-supervisor.
 - **Pharmacogenomics**: PGx-based contraindication prediction.
 - **Personalised Medicine**: statistics, product development and prototyping for Image.me.
- Jul 2018 **Service Analyst**, *Sopra Steria (Katowice, Poland)*, ITIL working environment.
- Mar 2018 Analysis/resolution of L1 IT support (general assistance, customer contact).
- Mar 2014 **R&D Researcher/Stat. modeler**, *Arvalis (Orsay, France)*, 1.5 years.
- Sept 2012 Improvement of market-leader Farmstar: satellite-image based management of wheat crops.

Education

- Jan 2018 **PhD, Dual Awarded Degree**, *The University of Nottingham/AgroParisTech (UK/France)*, 4 years.
- **Statistics**: Tolerance of wheat crops to STB. Study, protocol design, Statistical Analysis Plans.
 - **Project management**: 1 meta-analysis + 3 experiments during 3 years in 2 countries with 3 institutes. Outcome: international conferences, peer-review articles, English-written thesis.
- Sep 2012 **Master of Applied statistics**, *Agrocampus Ouest (Rennes, France)*, 2 years.
- Sep 2010 Content: 1.5-year learning program for applied statistics + 6-month graduation internship.

Skills

Methods **Statistics** from usual inferential techniques to state-of-the-art predictive/machine-learning algorithms, exploratory strategies, structure/unstructured data. **Statistical programming** (10 years experience). Relational **database**. Automated **reports and dashboards**.

Code/Soft.

- Senior programmer: **R**, **Shiny**, **rmarkdown**, **markdown** for statistics, reactive prototypes, automated reporting, reproducible science and R package.
- Daily user of **git**, **github** (branching workflow), CI/CD (Jenkins).
- Good knowledge of **SQL (PostgreSQL)**, **LaTeX**, bash, html/css, HPC and much more.

Languages **Native French; Fluent English; Polish** (in progress).

Scientific productions (ORCID 0000-0003-0524-5755)

- 2021 **Article (peer reviewed)**, *M.O. Bancak, F. Collin, P. Gate, D. Gouache, P. Bancal*, Towards a global characterization of winter wheat cultivars behavior in response to stressful environments during grain-filling, *Eur. J. Agron.*, 133, <https://doi.org/10.1016/j.eja.2021.126421>.
- 2021 **Article (peer reviewed)**, *K. Kapica-Topczewska, F. Collin, J. Tarasiuk, A. Czarnowska, M. Chorąży, A. Mirończuk, J. Kochanowicz, A. Kułakowska*, Assessment of Disability Progression Independent of Relapse and Brain MRI Activity in Patients with Multiple Sclerosis in Poland, *J. Clin. Med.*, 10(4) 868, <https://doi.org/10.3390/jcm10040868>.
- 2020 **Article (peer reviewed)**, *K. Kapica-Topczewska, F. Collin, J. Tarasiuk, A. Czarnowska, M. Chorąży, A. Mirończuk, J. Kochanowicz, A. Kułakowska*, JCV status, seroconversion rate and the risk of progressive multifocal leukoencephalopathy in Polish JCV seronegative patients with relapsing-remitting multiple sclerosis, *European Neurology*, 83(5), <https://doi.org/10.1159/000510849>.
- 2020 **Article (peer reviewed)**, *K. Kapica-Topczewska, F. Collin, J. Tarasiuk, M. Chorąży, A. Czarnowska, M. Kwaśniewski, W. Broła, H. Bartosik-Psujek, M. Adamczyk-Sowa, J. Kochanowicz, A. Kułakowska*, Clinical and epidemiological characteristics of multiple sclerosis patients receiving disease-modifying treatment in Poland, *Neurol Neurochir Pol*, 54:161–168, <https://doi.org/10.5603/PJNNS.a2020.0020>.
- 2020 **Article (peer reviewed)**, *M. Niemira, F. Collin, A. Szalkowska, A. Bielska, K. Chwiałkowska, J. Reszec, J. Niklinski, M. Kwasniewski, A. Kretowski*, Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (WGCNA), *MDPI*, 12:37, <https://doi.org/10.3390/cancers12010037>.
- 2020 **Article (peer reviewed)**, *F. Padilla-Martínez, F. Collin, M. Kwasniewski, K. Adam*, Systematic Review of Polygenic Risk Scores for Type 1 and Type 2 Diabetes, *MDPI*, 21:1703, <https://doi.org/10.3390/ijms21051703>.

- 2019 **Article (peer reviewed)**, K. Kapica-Topczewska, J. Tarasiuk, **F. Collin**, W. Broła, M. Chorąży, A. Czarnowska, M. Kwaśniewski, H. Bartosik-Psujek, M. Adamczyk-Sowa, J. Kochanowicz, A. Kułakowska, The effectiveness of interferon beta versus glatiramer acetate and natalizumab versus fingolimod in a Polish real-world population, PLOS ONE, 14:1–12, <https://doi.org/10.1371/journal.pone.0223863>.
- 2019 **Conference: ESHG 19, Gothenburg (Sweden)**, **F. Collin**, M. Niemira, A. J. Krętowski, J. Nikliński, M. Kwaśniewski, Poster: User-friendly and machine learning-empowered platform for classification of NSCLC based on RNA-seq profiling, https://fcacollin.github.io/Latarnia/doc/poster_190603.pdf, Awarded fellowship for European Countries.
- 2018 **Ph.D Manuscript**, **F. Collin**, The tolerance of wheat (*Triticum aestivum* L.) to Septoria tritici blotch, University of Nottingham/AgroParisTech, <https://pastel.archives-ouvertes.fr/tel-02443529/document>.
- 2018 **Article (peer reviewed)**, **F. Collin**, P. Bancal, J. Spink, P. Kock Appelgren, J. Smith, N.D. Paveley, M.O. Bancal, M.J. Foulkes, Wheat lines exhibiting variation in tolerance of Septoria tritici blotch differentiated by grain source limitation, Field Crops Research, 217:1–10, <https://doi.org/10.1016/j.fcr.2017.11.022>.
- 2018 **Conference: Phloème 2018 - 1ère biennales de l'innovation céréalière, Paris (France)**, **F. Collin**, P. Bancal, M.J. Foulkes, M.O. Bancal, La tolérance du blé à la septoriose, awarded "Promising thesis".
- 2016 **Conference: ESA 14, Edinburgh (UK)**, **F. Collin**, P. Bancal, J. Foulkes, M.O. Bancal, Poster: A statistical analysis of GxE contribution to leaf senescence during grain filling in wheat, https://fcacollin.github.io/Latarnia/doc/Collin2016_ESA_poster.pdf.
- 2016 **Conference: 9th ISSDC, Paris**, **F. Collin**, D. Gouache, M.O. Bancal, P. Bancal, Poster: Tolerance of wheat to Septoria tritici blotch: genetic vs environmental variations of key traits.
- 2015 **Article (peer reviewed)**, P. Bancal, M.O. Bancal, **F. Collin**, D. Gouache, Identifying traits leading to tolerance of wheat to Septoria tritici blotch, Field Crops Research, 180:179–185, <https://doi.org/10.1016/j.fcr.2015.05.006>.
- 2012 **Master thesis**, **F. Collin**, Modelling the Nitrogen Use Efficiency of an inorganic nitrogen fertilizer application on winter wheat, <https://dumas.ccsd.cnrs.fr/dumas-00741001/en/>.
- 2012 **Conference: R meeting, Bordeaux**, G. Bessigneul, **F. Collin**, M. Gauthier, M. Gérard, S. Lê, Lightning talk: A method for ecological data mining, bordeaux.inria.fr/recueil_resumes_R2012.pdf.
- 2011 **Misc: Sensometric school project, Rennes**, **F. Collin**, M. Gauthier, Sensorial analysis methods: the 'tablecloth' method, sensominer.free.fr/desc-napping.html.
- 2011 **Article (peer reviewed)**, F. Montfort, S. Poggi, S. Molière, **F. Collin**, E. Lemarchand, D.J. Bailey, Opportunities to reduce *Rhizoctonia solani* expression on carrots by biofumigation with indian mustard, Acta Horticulturae (ISHS), 917:149–157, <https://doi.org/10.17660/actahortic.2011.917.19>.