

Programming and statistics in the academics environment

Arne Ring

*Principal Statistician
Leicester Clinical Trials Unit
University of Leicester, UK*

*Professor of Statistics
University of the Free State
Bloemfontein, South Africa*

- Established 2009
 - Fully registered (UKCRC) since Oct 2013
- Support of all aspects of clinical trials
 - Grant applications, protocols
 - Trial management, data base development
 - Randomisation, analysis and reporting
- Currently about 45 active trials
 - Statistical support of about 30 trials at different stages
- 5 Statisticians at different levels
 - 3 more to be hired within next 6 months
 - No “statistical programmer”

Academic trial units

UK: Clinical trial unit

- About 50 CTUs
- Must fulfil registration criteria
 - Core funding of senior posts
 - Local quality system



Germany: Koordinierungszentrum

- About 20 KKSs
- National network
 - Initial funding from government
 - National quality system



Overview

- Personal background
 - Industry – Oxford – Leicester
- Current situation in UK and Germany
 - CTUs + KKS
 - Recent adverts programmer – compared to stats posts
- Professionalism from IT vs. stats
- (Future) role of data scientist

Personal background

- Industry statistician and statistics team leader
 - Boehringer Ingelheim
 - Internal and external collaborations with statistical programmers
- Diabetes Trials Unit Oxford
 - Primarily industry trials with academic support
 - Originally only (senior) statisticians for support and analysis
 - Collaboration with programming CRO + hiring internal programmer
- Leicester Clinical Trials Unit
 - Primarily academic trials (non-CTIMP)
 - Statisticians at various levels

Academic biostatistical support

- General departments (UK/Germany)
 - Health Sciences / Public Health / Statistics
 - Medical Departments
- Dedicated organisations for trials and studies
 - Clinical trial units
 - External academic institutions
(often with some epidemiological focus)

Academic biostatistical support

- In the UK, statisticians are hired at 4 academic grades
 - a. Research Assistant
 - b. Teaching Fellow / Research Associate / Medical Statistician
 - c. Lecturer / Research Fellow / Senior Statistician
 - d. Senior Lecturer / Senior Research Fellow / Principal Statistician
- In Germany, there is only one level (equivalent to b-c)
- Statistical junior roles focus on implementation of analyses
 - MSc in Statistics generally essential requirement
 - Often little background in computer science
 - Analyses (and their planning) typically less than 50% of overall capacity
 - Guided by senior oversight
 - Generally tenure track to senior positions available

Academic biostatistical support

- In general, there is little professional programming in UK/Germany academia
 - Typically all analysis requests are handled by statisticians
 - Not aware of any „statistical programmers“ at general departments
- Dedicated organizations for trials and studies
 - Some organizations have dedicated statistical programming support
 - In particular if CTIMP or industry trials are supported
- In UK decreased use of SAS in academia
 - STATA is more popular
 - Manual transfer of results to publications is often no burden

Main focus of analysis support

- Statistical methodology
 - Implementation of standard or advanced statistical methods
- Output generation
 - Provision of statistical displays in format understandable by general scientists
- Data preparation for statistical analysis
 - Irregularities in data highlighted and queried
 - Data structure adheres to requirements in statistical procedures
- Data preparation for regulatory reporting / data transparency
 - CDISC transformation
- Documentation and validation of analyses
 - Analysis plan and technical appendices to specify implementation steps

Main focus of academic analysis support

- Statistical methodology
 - Implementation of standard or advanced statistical methods
- Output generation
 - Provision of statistical displays in format understandable by general scientists
- Data preparation for statistical analysis
 - Irregularities in data highlighted and queried
 - Data structure adheres to requirements in statistical procedures
- Data preparation for regulatory reporting / data transparency
 - CDISC transformation
- Documentation and validation of analyses
 - Analysis plan and technical appendices to specify implementation steps

Less priority in academic analysis support

- Structured, modularised programming
 - Single program from data base to output (depending on complexity)
 - Code might often not be very efficient
- Data preparation for statistical analysis
 - Typically no harmonised approach to structure data for efficient programming
 - “Harmonised” data structure may be used (e.g. for scientific meta-analyses), but often not fully documented
- Professional output generation
 - Standard output from statistical software often rather awkward
 - Professional tables for presentations/publications are “handcrafted” and filled manually
- Data preparation for regulatory reporting
 - CDISC transformation

Statistical programming in industry

- Statistical methodology
 - Implementation of **standard** or advanced statistical methods
- Output generation
 - **Provision of statistical displays in format understandable by general scientists**
- Data preparation for statistical analysis
 - Irregularities in data highlighted and queried
 - **Data structure adheres to requirements in statistical procedures**
- Data preparation for regulatory reporting / data transparency
 - **CDISC transformation**
- Documentation and validation of analyses
 - Analysis plan and **technical appendices to specify implementation steps**

Statistical programming in industry

- Statistical methodology
 - Implementation of **standard** or **advanced statistical methods**
- Output generation
 - **Provision of statistical displays in format understandable by general scientists**
- Data preparation for statistical analysis
 - **Irregularities in data highlighted and queried**
 - **Data structure adheres to requirements in statistical procedures**
- Data preparation for regulatory reporting / data transparency
 - **CDISC transformation**
- Documentation and validation of analyses
 - Analysis plan and **technical appendices to specify implementation steps**

Ongoing / Future development

- Professional statistical programming
 - Modularised programming
 - Professional output
- Delivery of analyses of clinical trials
 - Implementation of methods to provide trial results
- Medical background of data
 - Familiarisation with data source and potential issues that could affect programs



Statistical
programmer



Medical
statistician

Ongoing / Future development

- Professional statistical programming
 - Modularised programming
 - Professional output
- Delivery of analyses of clinical trials
 - Implementation of methods to provide trial results
- Medical background of data
 - Familiarisation with data source and potential issues that could affect programs



Data
Scientist

Not necessarily in a single person, but in the team

Further Thoughts

- Own programming support would have been difficult without industry background and intensive collaboration with programmers
 - Several contributions with programmers to German statistical/programming conferences
 - Professional structure of derived data sets implemented
 - Efficiency valued during analysis programming
 - Did not always know before how good they were
 - Online-support by SAS community was extremely helpful

Adverts - last month

Am Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS GmbH ist in der Abteilung „Klinische Epidemiologie“ (Leitung Prof. Dr. med. Edeltraut Garbe) in der Fachgruppe „Arzneimittelerisikoforschung“ zum nächstmöglichen Zeitpunkt - unter Vorbehalt der Stellenfreigabe - die Stelle eines/einer

SAS-Programmierers/in

Entgeltgruppe bis zu 10 TVL je nach Qualifikation
39,2 Stunden/Woche, Teilzeitbeschäftigung ist möglich

vorerst befristet auf 2 Jahre zu besetzen. Eine Verlängerung der Stelle ist vorgesehen.

Aufgaben:

- Data Management
- Erstellung, Validierung und Ausführung von SAS Programmen
- Erstellung und Validierung von SAS Macros
- Generierung von Reports, Tabellen, Grafiken und Listings
- Statistische Auswertung epidemiologischer Studien

Anforderungen:

- Fachhochschulabschluss Mathematik/Statistik/Informatik bzw. Abschluss eines FH-Studiums in der Medizinischen Dokumentation oder Ausbildung zum Medizinischen Dokumentar oder vergleichbare Abschluss oder einschlägige Berufserfahrung in der statistischen Programmierung
- Fundierte praktische SAS-Kenntnisse
- Vorteilhaft sind Erfahrungen bei der Erstellung von Abfragen für relationale Datenbanken mit SQL und Erfahrung in der SAS-Programmierung mit großen Datenmengen

- 36,000 – 51,500 EUR
- TVL 13: 42,000 – 61,000 EUR

Statistical Programmer

Queen Mary, University of London - Wolfson Institute of Preventive Medicine

Location:	City Of London	Placed on:	8th October 2014
Salary:	£31,735 to £41,553 per annum inclusive of London Allowance	Closes:	6th November 2014
Hours:	Full Time	Job Ref:	QMUL4735
Contract:	Contract / Temporary		

The Centre for Cancer Prevention is at the forefront of research internationally into the prevention and control of cancer. We have particularly active programmes of research studying breast, cervical, colorectal and prostate cancer prevention, cancer screening and statistical methods for clinical trials and epidemiology.

Applications are invited for Statistical Programmer to provide statistical support to centre scientists to ensure data from studies is analysed accurately. This will involve writing statistical programmes to carry out non-standard statistical analyses, including disease progressing estimation.

	Requirements	Essential/Desirable
Qualifications	Degree in statistics, computer science or related quantitative field Post-graduate qualification in related discipline	Essential Desirable
Experience	Experience of working in a biomedical research environment Experience of working with large data sets Experience of statistical programming Experience writing programmes in STATA, SAS or R Experience of programming in generic languages such as C++	Desirable Essential Essential Essential Desirable
Knowledge, skills & abilities	Strong analytical and interpretative skills Programming skills Scientific knowledge and knowledge of medical terminology IT literate with a good working knowledge of statistical, mathematical and database packages Strong writing skills Ability to interpret the scientific literature and incorporate into project	Essential Essential Desirable Essential Desirable Desirable

People Management (*supervisory responsibility, human relations skills*)

May be expected to line manage other statistical or data management staff.

Development pathways

- Statistician
 - From academia with higher experience to industry
 - Professional statistical training often required to get into industry
- Statistical Programmer / Data Scientist
 - From Industry with higher experience to academia
 - Programmer with computer science background enriches teams
 - But will probably be a minority
 - Should have strong communication skills
 - Be able to provide training on efficient programming

Further thoughts

- How do academic institutions become attractive to programmers/data scientists?
 - Could fill a central post in analysis support
 - Opportunity to broaden background on medical data
 - Demonstrate communication + leadership skills

Ongoing / Future development

- Professional statistical programming
 - Modularised programming
 - Professional output
- Delivery of analyses of clinical trials
 - Implementation of methods to provide trial results
- Medical background of data
 - Familiarisation with data source and potential issues that could affect programs



Data
Scientist

Not necessarily in a single person, but in the team

Overview

- Academic trial centres – such as Clinical Trial Units (CTU) in the UK – are quite variable with regards to their statistical and programming support, depending on the nature of trials (e.g. CTIMP or not) and the funding sources (public or industry). Some larger CTUs have developed distinct roles like the pharmaceutical industry, with some form of specialism between programmers and statisticians who collaborate to efficiently plan and deliver the analysis, often beyond the actual trial objectives.
- Many other CTUs focus on support by statisticians only. In this case, there is typically a single medical statistician who plans and implements the trial analysis, supervised by a more senior statistician. All steps – analysis plan and display templates, derived data sets, statistical procedures, output generation and the discussions with the investigators are performed by this statistician. Experience in programming is gained as part of the job (and often not a major component during the selection process for a new post), which implies a somewhat lower level of proficiency.
- Based on my experience in or with industry and academia, I believe that some kind of specialisation would increase the efficiency of reporting trials. In contrast to industry, the number of programmers and data scientists might not become large, but dedicated professional support by senior programmers might increase the overall level of structured programming.
- I would like to discuss, how such roles could be developed to become attractive, which requirements would be involved and which opportunities would arise.