

Report on ICMS Workshop on Statistical Methods for Genetic Epidemiology Edinburgh May 7–11 2007

Scientific Organisers

1. Nuala A. Sheehan (Principal Organiser), Reader in Statistical Genetics, University of Leicester, UK. (nas11@le.ac.uk)
2. Elizabeth A. Thompson, Professor of Statistics, Biostatistics and Genome Sciences, University of Washington, Seattle, USA. (thompson@stat.washington.edu)
3. Max P. Baur, Director of Institute of Medical Biometry, Informatics and Epidemiology, University of Bonn, Germany. (Max.Baur@ukb.uni-bonn.de)

1 Deviations from the original proposal

There were no significant deviations from our original proposal to the ICMS.

2 Short description of the meeting

Genetic Epidemiology is such an active area of research at the moment that there is a serious shortage of expertise in the UK to meet the current demands. Highly sophisticated statistical methods are required to deal with the complexity that arises in many biological applications due to large datasets, indirect measurements, complex underlying biological processes and various combinations of all three. Moreover, recent developments in large scale bioscience are rapidly out-pacing current analytic capacity. Strong national and international collaborations are crucial if UK scientists are to stay up to date in this rapidly developing area. The aim of this workshop was to gather together a sufficiently expert group with interests ranging across diverse areas of genetic epidemiological application in order to focus on the computational and methodological issues that are *common* to family studies, genetic association studies and the analyses of complex traits. Participants thus had expertise in Markov chain Monte Carlo methods, graphical modelling, statistical computation and in the management and analysis of large applied population studies. One objective was to assess existing methods of analysis of large and complex datasets, to consider appropriate adaptation of these methods and to investigate the potential for new methods which address

common problematic features of all these datasets. Another was to actively encourage young researchers into this field by providing the opportunity to become acquainted with the main research issues in the area and to make vital contacts with the leading people and their groups in an informal small group setting.

3 Comprehensive report of the workshop

The workshop ran during a full week beginning at lunchtime on Monday and ending at lunchtime on Friday. The aim was to keep it small and focused but it was apparent from the many requests sent to the organisers and to the ICMS that this was a very popular topic and we could have had a much larger meeting. That, however, was not the purpose here as high-level discussion is more easily facilitated in smaller groups. However, given the amount of interest it generated, it might be worthwhile considering a larger and more open follow-up meeting in the future. There were 34 participants in total of whom 17 were from the UK, 6 from the US and 11 from mainland Europe reflecting the national and international activity in this area. We had 19 men and 15 women, which is not unusual in this field, and 9 of the 34 were junior researchers. The junior research places were advertised and selection was based on a combination of research interests and the desire to prioritise groups that were not already represented among the senior participants.

The format of the workshop comprised expository talks by the key participants outlining the main challenges to current methodology, in terms of problems, limitations and computational issues, research talks on new methods currently under development and themed discussion sessions chaired by leading researchers in the relevant areas. Most of the senior participants are well-known figures in genetic epidemiology and statistical genetics. We deliberately included 5 senior people from the related areas of traditional epidemiology, bioinformatics, causal reasoning, population biobanking and medical statistics in order to lend a fresh perspective and to broaden our expertise base. All these individuals integrated well giving wonderful presentations and contributing keenly to all the discussion sessions. A full list of participants and their e-mail addresses together with the workshop schedule is available on the ICMS website. Many of the speakers have also contributed the slides for their presentations.

3.1 Workshop Highlights

The workshop began with two overview talks that were very well received. David Clayton (Cambridge University) brought us all up to date with current topical events by describing the experiences of the Wellcome Trust Case-Control Consortium with genome-wide association studies. (The first paper from the consortium was published in *Nature* in June 2007.) The different kinds of bias that arise with such data were introduced and the general problem of defining the power of a study when very little is known about the

effect sizes of interest. Replication of results across different studies, rather than citation of p-values, is all-important. Although it is sometimes claimed that “the more data we get, the less we need statistics”, it was pointed out that other problems, such as copy number variation, will present many more modelling challenges than the diallelic single nucleotide polymorphisms (SNPs) of genome-scan analysis. Duncan Thomas (University of Southern California, Los Angeles) followed up with an overview of the problems in inferring complex biological pathways where there sometimes appears to be an enormous gulf between the knowledge we have of the underlying biology of a problem and the data we have available to us. There were 19 talks on current research around topics as diverse as multifactorial traits, relationship estimation, graphical models and causality, linkage analysis, quantitative traits, DNA structure and function and population stratification. (See ICMS website for further details.).

The three 90-minute discussion sessions were a feature of this workshop and worked well because of the size of the group and the willingness of the participants to engage. The first of these led by Elizabeth Thompson (University of Washington, Seattle), Cornelia Van Duijn (Erasmus University, Rotterdam) and Simon Heath (Centre National de Génotypage, Paris) was a lively interchange of ideas on the role of pedigrees and family data in the current climate of genome wide association and population biobank studies. The question “Is linkage dead or just under-nourished?” was where the panellists began. The overall conclusion seemed to be that linkage *and* association are both required for the analysis of multifactorial traits but this discussion was a theme that ran through the entire meeting. The genome wide association discussion was led by David Clayton (Cambridge), Duncan Thomas (UCLA), Peter Holmans (University of Wales, Cardiff) and Heike Bickeböllner (University of Göttingen) and provided an opportunity for participants to explore some of the issues brought up in David Clayton’s opening talk. Issues of multiple testing, heterogeneity, selection of cases and controls, and potential use of family data in association studies (following on from the previous day’s discussion) were debated long into the tea break. The final discussion session on data integration was led by Robert Elston (Case Western Reserve University, Cleveland), John Whittaker (London School of Hygiene and Tropical Medicine) and Wally Gilks (University of Leeds). This was more exploratory in nature as the need to incorporate more and more different forms of information was discussed. Integrating good knowledge on physical maps and biological function with genetic data will undoubtedly lead to much more complicated modelling but summary statistics or data “disintegration” will then be important for interpretation and communication.

On Wednesday morning, junior researchers were invited to suggest topics that they wished to have discussed in more detail. These topics were organised under the following eight broad headings:

1. variance components models for quantitative traits
2. family-based analyses using FBAT etc.
3. genetic imprinting
4. future issues for genome-wide association
5. population versus family data

6. rare variants and haplotypes
7. relationship estimation
8. Mendelian randomisation

and were scheduled, four at a time in parallel, for each of the two 90-minute small group sessions in the programme. Each group was led by at least two senior researchers and groups varied widely in scope ranging from small focused technical discussions to larger and more open discussions of current work in the area. The aim of this exercise was to provide an opportunity to find out about the key people and current state-of-the-art in these areas in a more informal setting and to establish groups of researchers with common interests. The feedback from these sessions was presented and discussed more generally by the whole group in the workshop “Wrap-up” session on Friday morning and it was noted that many beneficial discussions took place and good contacts were made. It was also agreed that Wally Gilks’ presentation on chromosome territories was totally new to most people in the audience and posed many questions about gene-gene interactions, for example, that had not been previously considered. If different areas of the genome, generally assumed to be completely independent, really can be transcribed at the same time, many current modelling assumptions should be completely revised.

3.2 Involvement of participants

One of the main successes of this workshop was in getting such a prestigious group of senior people to come and stay for the full week. Besides creating a great feeling of continuity, this enabled important discussions to carry on and develop over several days and gave everyone a chance to meet everyone else and to join in these discussions as they progressed. The enthusiasm and energy that people brought to the meeting and their willingness to engage in open and frank discussions on many issues was a vital factor in the workshop’s success. We felt that getting everyone together in the evenings was important, particularly for involving the large proportion of young researchers, although this did result in a very intense schedule. The general consensus, however, was that the event, although exhausting and highly concentrated, was extremely rewarding and that more meetings of this type would be beneficial to the scientific community.

3.3 Impact of the workshop: collaborations and new work

The questionnaire feedback and personal comments we received from the participants implied that people were generally very supportive of the idea of having a small group together in one place for such a long time. Some went so far as to say that it was the best meeting they had been at in years. Much more networking and collaborative discussions can take place in such a setting than in a general conference. The ICMS headquarters in India Street provided a perfect environment for this to happen and everybody commented on the

smooth organisation of the meeting and excellent quality of the conference facilities, meals etc. Several participants have mentioned specific collaborations that they intend to pursue as a result of this meeting while others have consolidated old ones with fresh input and new ideas. Some people commented that even if specific collaborations did not ensue, they now know who the main players in this area are and will know whom to approach in the future. Many requests for a “follow-up” meeting were also received.

3.4 General comments and feedback

There was one comment that the ratio of senior scientists to post-docs was “intimidating” at times and a suggestion that more post-docs and less senior people should have been invited. This aspect was considered when planning the workshop and the organisers felt that it would be risky to have substantially more than a quarter of the participants as juniors for such an event since only the senior people could be relied on to lead and sustain the discussions which were a main feature of this meeting. A different kind of meeting with more emphasis on formal talks could have accommodated more junior participants. Perhaps the above comment indicates that we need more small meetings like this where junior people actually get used to circulating with the big names in the area and use the occasion to learn from them rather than feel intimidated. Most of our post-docs had never been to anything like this before and most of them really seemed to relish it. It was also noted that the poster sessions were not always very well attended but that it was difficult to see how this could have been organised otherwise. This is a good point. Some of the post-docs prepared posters which were on display for three full days in a ground floor room of the building. The scheduled poster sessions were all late in the day (17 : 30 – 18 : 30) but, in principle, people could see them anytime during lunch or coffee breaks as well. In practice, the schedule was so tight that participants tended to take a break before joining the group again for dinner in the evening. We would wish to address this issue were we to organise another event like this.

4 Acknowledgements

We would like to express our gratitude to the ICMS and its funders for the financial support and excellent local organisation for this workshop. From the questionnaire responses, it is safe to say that all the participants join us in thanking the ICMS for giving us such an opportunity.

Nuala Sheehan (with Max Baur and Elizabeth Thompson)