University of British Columbia
Department of Statistics

Towards a long range plan for statistics

Jim Zidek, February 2011

GENERAL CONSIDERATIONS. The radical change in NSERC's mission, from the supporting of basic research to the RPP – Research Partnership Program will have a much more profound effect on the direction of the discipline of statistics than anything long range planners can achieve.

The RPP.... NSERC's new mission, which apparently developed without consulting Canada's research community, enshrines the trinity of the Research Partnership Program. Thus 60 % of NSERC's $1bi budget is allocated to, *innovation* (productive use of new knowledge in all sectors of the economy and society- $3.25bi), *people* (highly skilled, well educated and capable of lifelong learning - $3.02bi), and 40% to *discovery* (competitive research in science and engineering - $4.00bi). The overall goal of this experiment of stimulating industrial R&D seems laudable. But it is being created at the cost of the Discovery Grants (DGs) program. So while NSERC's overall budget has grown, the DG program has in 1978 dollars, has declined, so that average DG has declined from a peak of around $15k p.a in the mid 1980s to just over $10k today. NSERC's stated plan is to reduce funding for this program further in the coming years.

The issues here are about substance, not money. Its curiosity driven research that produces the innovations (e.g. the laser) that most benefit the economy and society in the long run. Less well recognized is the vital role DGs play in education by keeping educators informed and up-to-date on developments in their discipline through conferences, workshops and international collaboration. This translates not only into research papers, but as well into textbooks, up-to-date curriculum, and above all, the inspired young minds that turn into the *people* the RPP needs to work properly. .

By design apparently, the RPP will manipulate both industry and academic researchers into a targeted research alliances with emphasis on industrial partnership programs such as the Collaborative Research and Development (CRD) program – good programs in which SFU and UBC statistical scientists are currently engaged. But with that change in direction, comes the new emphasis on training HQP and at the expense of time for high level research.

Other top-down changes. Lots of other changes seem to have been imposed on the research community including the merger of statistics and mathematics into a single funding envelope despite the divergence between those disciplines that has been taking place since the 1970s so that they have little in common today. In reality statistics now has stronger links to fields such as medicine, biology, computer science (more generally, information science), and geoscientific than with mathematics. The new grant adjudication also seems to have imposed by bureaucrats without consultation. The intended result seems to be dynamically changing and unpredictable grant renewal outcomes and the success rate on renewals has dropped dramatically in the most competition. No indication is given as to why the new system is thought to produce better science. But it sure will increase the number of HQPs for the RPP, even if we have to drag them in off the street!

In summary, long range planning seems likely to have at most a marginal impact on the direction of the discipline, compared to the changes already imposed from above.

SPECIFIC ISSUES.

Science.

Highlights.

*New directions in statistics....*Since the middle of the last century, mathematical statistics in North America has evolved into statistical science, which has thrived through a virtuous circle of interdisciplinary research: subject areas have benefited from statistical advice and statisticians have found new theoretical directions. The whole sub-discipline biostatistics was born. More recently have come bioinformatics and statistical computing with their research challenges such as: the problem of synthesizing the classical, inferential foundations on which statistics has been built with the algorithmic approaches of computer science; dealing with high dimensional as well as massive data sets; data compression without loss of statistical information; non-traditional data forms such as text, images and the continuous data records; the large p vs n data problems with vastly more parameters than data records. A new frontier involves the unification of physical and statistical approaches to modelling for synergistic effect – the former provides physical knowledge through things like differential equations for dynamic systems while the latter can quantify uncertainty in such systems. This frontier includes the design of computer experiments for numerical models and the fusion of simulated data from those models and measurements. Although this frontier would seem a natural place for collaboration of applied mathematicians and statisticians, in fact, that has not taken place. Instead its the physical scientists with whom the strongest links have forged. This echos the difficulty experienced by the Statistical and Applied Mathematical Sciences Institute (SAMSI) in creating this bridge and of UBC's Institute of Applied Mathematics and Statistics, formed in the earlier 1970s. As a founding member of the latter, I can attest to its vibrancy as well as its success in enabling graduate study in applied math and statistics outside the tight strictures algebra – topology – analysis requirements of mathematica in the days when all were housed under one roof.

Loss of the core... The great success of statistical science has led to a worrying trend away from the core. Increasingly little attention is paid to basic theory and foundations. The decline is exacerbated by NSERC's emphasis on HQP training. Students naturally gravitate to the supervisors working in areas seen more likely to lead to jobs, 70% of which for statistics PhDs are found in government or the private sector rather than academia. But then, responding to the now very demanding training mandate in DG program, researchers gravitate toward toward trendy areas such as biostatistics and finance.

Changes. The interconnected global community means is making national boundaries increasingly irrelevant, in particular for scientific inquiry and this is leading to increasingly to multi-national research collaborations. Some of these are driven by global scale problems such as climate change. The upshot is the need for national infrastructures to fund, rationalize and coordinate these collaborations and the associated exchange programs for scientists as well as technologists and HQP.

Innovation is producing the new “sputnik moments” and a vital need for knowledge based economies. Thus for example, China has begun investing heavily in R&D and aggressively going after talented researchers in other countries. International competition for those top researchers includes generous research support.

The consequences of climate change are seemingly been felt with an emerging interest in mitigation and risk management. The recently established Pacific Institute for Climate Studies is a manifestation of that trend. Rising food costs, forthcoming water shortages and rising ocean levels will be spawning a lot of research opportunities.

Opportunities for statistics... Statistical scientists will see increasing areas of research opportunity generated by concern about climate change. Work has already begun on the uncertainty quantification for computer simulation models, in particular weather and climate models. Inferring historical climate patterns by paleo – climate data analysis is another active research area. In Canada, statistical scientists and their subject area counterparts have been investigating the risks to Forests of the changing climate. Finally UBC statistical scientists have collaborated with scientists from Agriculture and Agro-Food Canada (AAFC) in an agroclimate risk management study. The work began with one year of funding from the National Institute of Complex Data Systems until its funding renewal application failure. The AAFC provided complementary funding. The research involved the modelling of crop yields as a function of climate variables as well as the phenology of plant development with the ultimate goal of linking these models to downscaled regional climate models. The outside interest groups for these projects included farmers, insurance companies and investors. Attention is now turning to monitoring programs where networks of micro-sensors are deployed to measure soil temperature and moisture. A session at the International Statistical Institute's Dublin meeting in August 2011 has been organized on that topic.

Cross – disciplinary work... This topic has been addressed above under "Highlights". Briefly statistical scientists have not found a basis for collaboration with mathematical scientists. However research collaborations have been established with a large number of other research communities.

Research Funding.

Infrastructure... The Statistical Consulting and Research Lab (SCARL) in my Department, which was established in the 1980's under NSERC's Infrastructure Grants Program continues to be my main source of research support. It provides and excellent distributed computing system with high end processors that has served me and my graduate students well. Under contract with FPInnovations it has also provided for a number of years, it has provided research support for that off campus wood products research lab. That laid the foundation for a successful NSERC – CRD application by UBC and SFU statistical scientists who now do collaborative research with wood scientists at FPInnovations. Paid by the grant, SCARL provides a programmer analysis as a technology transfer agent who has access to the proprietary data on which discoveries can be tested. SCARL also hires statistics graduate student assistants from time-to-time as needed to handle the incoming workload on a fee basis. Finally the Lab is providing a software develop to help develop statistical packages as a way of disseminated research results.

Such infrastructure seems essential in universities to assist researchers if NSERC's RPP experiment has any hope of reaching its goals. That success cannot be achieved in the long run through ad hoc arrangements set up on a partnership-by-partnership basis by individual investigators.

Also needed are much improved cyber services at the department level. These enable regular research group meetings, mine currently including regular meetings with a collaborator in the UK and a UBC grad student. As well such a facility could be used by cyber courses - PIMS allowed us to use its facilities for two experimental graduate level cyber courses involving four Pacific NW institutions on specialized topics. Those at UBC are extremely expensive and are not always available.

Activities... Interdisciplinary activities, will likely be most important for my research, including workshops with subject area and industry partners. BIRS provides an ideal environment for enabling such activities to take place. Conferences will continue to play a key role. Finally my collaborative research benefits from field trips and personal interaction with my co-investigators.

Research opportunities to benefit... A policy change would go a lot further than any specific research opportunity to increase research capacity of Canadian researchers in general, relative to their

US counterparts, namely the elimination of NSERC's policy that makes teaching release ineligible. At my university courses now require a lot more time nowadays than before. At the other extreme, to deal with budget cuts successive UBC administrations have been downloading to departments, administration that used to be handled centrally. Fortunately I was able to retire and now am able to give all my time to research and associated activities.

The Institutes:

Relevance and long term impact... MITACS is one of NSERC's great success stories, although its being retired due to NCE rules. Chances of success of the RPP would be enhanced by funding some version of it under a new program to preserve it as an infrastructure on which to build a bridge between the industry and the constituencies MITACS represents, including statistics. It has become a hugely valuable asset to the community.

Both PIMS and BIRS have also been a substantial benefit to the statistics community. Having recently spent 3 years on the BIRS review panel, I can attest to the generally high quality and remarkable diversity of applications for the 48 weekly slots in any given year. We had about 140 applications in the last round and about 2/3 of these could have been funded had space been available. Few of these were from statisticians although quite a number came from quantitative but not non-math science areas. Our panel was asked to fairly balance the distribution of them to ensure that these peripheral areas got represented subject to being of acceptable quality. PIMS has awarded me and my three co-PIs a 2 year collaborative research grant that enabled us to hold about a dozen workshops, conferences and (two) summer schools. Some infrastructural arrangements were provided to help us organize these events. It also provided some postdoctoral and GRA funding for research on climate. Finally it allowed us to use its cyber facilities to give to specialized graduate courses in statistics to the four institutions involved: SFU, UBC-Okanagan, UBC-Vancouver and the U Washington from which the co-PIs came. One of those courses was given by an agro-scientist involved in our NICDS program.

The NICDS is the institute that is missing from the list and should not be. It was created as a result of an earlier NSERC assessment of the discipline and was exemplary for requiring external partner support. A great deal was accomplished on a very modest budget and it remains a mystery as to why it was not refunded by NSERC, especially since it was so well reviewed. It had a great deal of relevance to us, since it enabled the launch of our agroclimate risk management project mentioned above. Moreover, it would have fit very well into the RPP program.

Overall its hard to comment on the long term impact of the institutes on statistics since I don't have to the information needed to make that assessment. However, they do provide valuable facilities and opportunities for multi-institutional programs that would be very hard to organize with their help. For example, our agro-scientist instructor referred to above, was able to use cyber facilities at the U of Lethbridge, because the course was being given as a part of our PIMS sponsored collaborative research group program. It could not have been given otherwise.

New activities... IT technology is making it feasible to have multi-institutional graduate programs and we aspire to have such program in environmetrics. PIMS would provide a framework in which to launch that program in the West (so that time zone issues do not arise). This would mean that top notch statistical scientists could give specialized courses with sufficient enrolments to make them feasible.

Training. The biggest obstacle to training for most statistics faculty is a lack of time. Training capacity of tightly circumscribed by teaching duties over 8 months of the year, and growing administrative duties on the other over the entire year. Training itself is very difficult and time consuming except in the case of that remarkably rare gifted grad student. Faculty have to split their time between training and doing their own research, which should be at a higher level.

Structures... Training as mandated by the Discovery Grants rules under current adjudication processes works against researcher's at smaller or undergraduate only institutions. Here a structure is needed to facilitate co-supervisory roles for such researchers and the cyber infrastructures described above could be used for research group meetings. The students would be expected to their co-supervisors for periods of time when feasible with expenses covered by NSERC grants.

Pipelines... The UGRA is a valuable program but difficult to implement because of the lack of time as discussed above. But this is the group of people who will be become the advanced trainee's of tomorrow for the RPP.

International.

Support for international colleagues... The UK's Leverhulme Trust

http://www.leverhulme.ac.uk/funding/RF/RF.cfm

provides exemplary support for a range of investigators, including visiting fellowships. Such a program would well be copied here to enable the international collaborations described above. However, given the declining level of the Discovery Grants program, the funding such a program could not come out of that budget.

Structures to support... The cyber facilities mentioned above would certainly assist international collaboration by enabling research group meetings.