Comments concerning the long range planning exercise for the statistical sciences

Jerry Lawless, Mary Thompson and Grace Yi

The following remarks deal with a number of issues in the discussion topics circulated by the LRP steering committee. They were prepared by the three of us, but with input and endorsement from University of Waterloo colleagues Bovas Abraham, Steve Brown, Shoja Chenouri, Cecilia Cotton, Joel Dubin, Adam Kolkiewicz, Christiane Lemieux, David Matthews, David Saunders and Changbao Wu.

Science and Statistical Science

A key feature of current and emerging science in every area is the need to analyze more and more observational (as opposed to experimental) data, much of which has a complex structure. Such data require that statistical methods play a crucial role. In addition, statistical analysis and algorithms underlie decision making based on data. Fields of major current activity include genomics and genetics, medicine, public health, economics and finance, climatology, environmental science, marketing and manufacturing. Examples of the impact of statistical research in medicine and genetics can be seen, for example, in the algorithms used to "call" genotypes based on DNA analysis; in the methodology that has been developed to deal with the simultaneous testing of thousands or hundreds of thousands of genetic association hypotheses in genome wide association studies; and in the predictive models that have been developed for "personalizing" risks of disease and the impacts of interventions. Many other contributions are related to technology, for example medical imaging and diagnostic procedures and the use of customer profiles to target marketing and advertising on the internet. Statistical scientists in Canada and elsewhere are faced with tremendous and exciting challenges; they have already been responding for many years but the pace of science and technological development is accelerating. There is an urgent need for more statistical scientists and for a concomitant rise in funding to support research and the training of highly qualified personnel (HQP).

Statistics is a rapidly growing discipline and abundant evidence shows that statistics now stands in a prominent place. The numbers of PhDs produced in statistics has been growing steadily over the past forty years. For instance, according to the US National Research Council, the number of statistics and probability Ph.D.'s granted in 1981 was almost double the 1970 level. In contrast, Ph.D. production in 1980 was 72% of the 1970 level for engineering, 69% for chemistry, and 60% for physics and astronomy. Ph.D. production in statistics and probability over that period was roughly half that of all core mathematics and 50% greater than the output in applied mathematics (Moore and Olkin, 1984). In a more recent article Crank (2010) examines PhD production in statistics in the USA over 2000-2007. Different data sources give somewhat different absolute numbers, but all indicate a 30-40 percent increase in the annual number of PhDs over that period. Similarly, memberships in the Statistical Society of Canada (SSC) and the American Statistical Association (ASA) have each grown by about 50 per cent over the past 25 years. Currently, the SSC has about 900 members and the ASA about 18,000. In spite of this, there remains a tremendous demand for new statistical scientists.

Research Funding

NSERC Discovery Grants are a crucial component of statistical research within Canadian science, and are the main source of support for core innovation in statistical methods and theory. They are also crucial for training; it is essential that tomorrow's statistical scientists be afforded the opportunity to delve deeply into methods and theory. NSERC is the key provider of support for this activity and failure to continue and in fact expand this will lead to new generations of statistical scientists whose ability to innovate is underdeveloped. Moreover, it will not be possible to attract and support the numbers of students that the private and public sectors require. Training of HQP is discussed below, but we note that historical and current levels of NSERC support to individual researchers fail to provide adequate funding for Canadian or international graduate students. With 70 percent of researchers receiving less than $20,000, the majority of researchers can at best provide full research support for a single student.

Statistics is a data-oriented science and as noted, modern data arising from technological and scientific fields are commonly of high dimension and complex structure, and sophisticated computational procedures are involved in their analysis. Statistical research is similarly highly computer-dependent, and there is a need within research teams for statistical programmers and persons proficient at data management. Such assistants increase tremendously the productivity of statistics researchers and it is crucial that there be funding to employ such individuals as technical assistants for individuals or groups of researchers. Technical assistants have been the norm for many years in computer and information science, and in sciences like biology and chemistry; their addition to statistical research teams is long overdue. We note that research trainees at the Masters and Doctoral levels cannot fulfill this role as full time students; they have neither the time nor, in most cases, the requisite specialist skills.

We recognize that research support is available from other granting agencies for some statistical work within specific scientific fields, for example, genomics, medicine and public health. The statistical research component of projects supported by other Canadian agencies is typically small, because the demands from the science for rapid analysis of data do not allow for due consideration of new methodology, let alone theory that ties together methodology, illuminates its properties and suggests new directions. It is vitally important that statistical scientists in Canada collaborate with scientists in other areas, and that they draw inspiration and motivation for new research in statistics from this. The other areas cannot be expected to provide major support for innovation in statistical science, however; this role must be fulfilled by NSERC.

A similar remark applies to NSERC collaborative research grants programs. Most collaborative grants provide support not for fundamental research in statistical theory and methods, but rather for the application of existing methodology or the refinement of such methodology for a specific area of science or technology. Other models for funding collaborative research which exist in some countries seem outside NSERC’s purview. In the US, for example, faculty and students receive more research support from grants or contracts in other scientific fields, biomedical sciences and public health being the most prominent. Research support for these areas is much greater in the US than in Canada, as is recognition of the need for expert statistical collaboration and the provision of grants that support methodological development in statistics. Moreover, methodology grants in biostatistics are much more widely available

from US agencies like the National Institutes of Health (NIH) than from CIHR. In Canada, NSERC is a crucial source of support for researchers in biostatistics and more broadly, for methodological innovations in statistics that are motivated by research in areas supported by other granting agencies.

Institutes

The institutes have not to this point played a major role in research for the statistical sciences. Their main contributions are through support for specialized workshops and symposia, and through quite small amounts of postdoctoral fellowship support. This is arguably as it should be, since most areas of statistical science have quite different emphases and needs than most areas of mathematics. Natural collaboration for most statisticians is not with mathematicians but with scientists in other fields. It is important that NSERC support for statistical science not be confounded with or closely linked to support for the institutes. Statistical scientists make up about 35 percent of the current NSERC Discovery Grant holders in the mathematical and statistical sciences (for 2009-10, 314 out of 961). The portion of the institutes' research activity that is associated with statistical science is much lower; in most cases it is 10 percent or less.

An area that deserves study for institutes in general, whether they are related to statistical science or not, is a mechanism for assuring that any research grants provided by institutes to individual researchers or groups have transparent and rigorous peer review processes. The rigors of the Discovery Grants peer review process should not be waived in the review of proposals for support made to an institute.

Training of HQP

We have noted the crucial role that NSERC support plays for the training of graduate students in statistical science. There is an urgent need for an increase in the number of highly qualified individuals. There are two main factors: increasing demands from scientific, technological, commercial and government groups for HQP; and the age structure of many statistics departments and groups, for which recent and impending retirements are creating a major need for personnel.

A reasonable and effective structure for training HQP should be targeted to meet varying demands. HQP in a research team may variously include PhD students, Master's students, senior undergraduate students, post-doctoral fellows, and technical assistants. To maximize innovation in statistical research, support for HQP training should grow at the same pace as graduate programs and faculty numbers at Canadian universities. Opportunities for enhanced graduate training should be coordinated in time and location and funded on a continuing basis.

Some graduate students get partial support from collaborative grants and from arrangements with researchers and groups in other scientific fields. We view this as important; in addition to receiving advanced training in statistics, students and postdoctoral fellows should be exposed to other areas of science, and to the statistical needs in these areas. Internships whereby a graduate student spends two to four months with a university, scientific, government or industrial partner are extremely valuable for the student and for faculty seeking to develop collaborations. We note that MITACS has been a supporter of internships, and as MITACS will now focus on a broader constituency, NSERC should

consider how it might supplement this program for the mathematical and statistical sciences. In conjunction with improved funding for research assistantships supported by Discovery Grants, these would enhance training and promote innovative research by graduate students.

Statistical programmers represent an important component of statistical research teams, as noted above. They help to speed up methodology development but are also important for dissemination of newly developed methods through software. The training of such highly qualified personnel is essential and might typically be done within Master's programs, supplemented by specialist training in computing. Finally, we note that postdoctoral fellows are of increasing importance in the training of new statistical scientists in many areas. Development of the kinds of expertise needed to make substantial contributions in areas such as genomics or medical imaging begins with a firm grounding in doctoral research but also requires a period of intensive learning and collaboration with scientists which is facilitated by postdoctoral funding. It is important that NSERC support aspects of such specialized training that involve the development of new statistical theory and methods.

Finally, in recognizing the need to attract more students to statistics, we note the importance of having researchers in undergraduate classrooms. The attraction of Canadians into statistical science has decreased markedly over the past 30 years and while it is the responsibility of universities and professional societies to address this, NSERC should be cautious about moves that foster further separation between research and teaching.

International

This is not a major factor for research in statistical science, assuming that Discovery Grants continue to support travel to conferences as well as occasional field trips and visiting scientists. Statisticians do not in most cases rely on major infrastructure or laboratories for fundamental work in theory and methodology, so visits to foreign centres is not a main issue. International co-operation among research institutes and centres is helpful in planning and coordinating workshops and symposia, but this is reasonably straightforward.

Statistical Science and NSERC Support

We have indicated above the importance of NSERC support for research in statistical methodology and theory. The Discovery Grants program is crucial to the fundamental research and training enterprise in statistical science. It is also highly important for excellent and innovative collaborative research, because it is only through well trained personnel that the challenges arising from new scientific research and technology can be met, along with the need to provide software platforms for implementing methodology.

References

K. Crank (2010). Counting Statisticians: How Many of Us Are There? *Amstat News, May 2010, 26-27.*

D. S. Moore and I. Olkin (1984). Academic Statistics: Growth, Change and Federal Support. *The American Statistician, 38, 1-7.*