Statistics of statisticians

Critical masses for research groups

If you work in a group of researchers, how big should your group be? Too small, and you have no one to bounce ideas off – and you may face extinction by funding cuts. Big groups bring fertile interactions and better-quality work. So is bigger always better? And does it depend on what you are researching? **Ralph Kenna** and **Bertrand Berche** have found out…

Do big groups do better research than small ones? Funders believe so – but is it true?

The notion of *critical mass* in research has been around for a long time[1] but it has recently become much more important. Policy makers, university managers – and, let's face it, the people who allocate funds – want research to be efficient; they want value for their money. They want the best research possible out of the researchers they have or pay for. So they look for indicators of the quality of academic research – and the simpler those indicators, in their view, the better. One simple indicator, so it has been claimed, is the size of the research group – the number of people involved in it. So this has become an issue to funders and to managers, and notions associated with critical mass have become increasingly important. It is perhaps surprising, therefore, that until very recently critical mass has lacked proper definition and measurement, and even understanding[2, 3].

Group thinking can bring quality results… © iStockphoto.com/LifesizeImages

The old notion of critical mass was of some sort of threshold group or department size, below which research quality tends to be poor and above which research standards start to improve. The idea has been extended to, and perhaps beyond, its logical conclusion: that "the bigger the group, the better", and that "benefit continues to accrue through increasing scale". And this belief has had consequences. It has brought calls from some lobbyists to concentrate resources into a small number of elite research institutions, where the groups will be big and the research correspondingly better. However, despite analyses based on counting citations of research papers, no evidence for such a threshold has ever been found[4]. Is the idea actually true?

To test it, we need a mathematical model for the relationship between research quality and group quantity. Happily, such a model has recently been developed[1]. It is a model that belongs, with many others, to the new and popular discipline known as sociophysics, in which physical principles are applied to social phenomena, and it uses ideas imported from statistical mechanics. The new model turns the old idea of critical mass on its head. Rather than a minimum group size required for quality research, critical mass emerges as an upper limit, above which research quality either tends not to improve or the rate of improvement starts to level out. This levelling out of research quality is due to communication limits and is known as the *Ringelmann effect* in sociology[5]. Ringelmann published his work in 1913, so the idea is approaching its hundredth birthday. Only now is it being tested mathematically in this context.

Typically for sociophysics models, the new theory has a very simple basis. One could assume that the strength of a research group or department is a simple sum of the strengths of the individuals comprising it. If the mean strength of the individuals in a group is a, and if the group has N members, then the strength of the group would be aN, a linear function of group size. But our assumption is a naïve one. Research groups are complex systems, in which interactions between individuals play crucial roles – so we need a better model which takes account of the interactions. Suppose an interaction between researchers gives an added effect, and the mean strength of that effect is b for each interaction. In a group of N people there are $N(N - 1)/2$ possible two-way communication links; then the group strength becomes $aN + bN(N - 1)/2$. This has an N^2 term in it, so the strength of the group as a whole is

… but a one-man band can seek inspiration unhindered. © iStockphoto.com/Peter Booth

In apes, chimpanzees and man there is a limit to the number of individuals who can communicate with each other

quadratically related to the number of people in it. If we then define quality as strength per head, we end up with a linear relation between quality and quantity. As an equation, average quality $\langle s \rangle$ can be written, when N is below the critical mass, as

$$\langle s \rangle = a_1 + b_1 N \qquad (1)$$

However, within a given discipline, one may expect there to be a limit to the number of colleagues with whom one can meaningfully communicate. In social groups, this limit is known as the *Dunbar number*[6]. Dunbar found it in chimpanzees and apes as well as in human beings. For research groups (who of course do not want to be confused with Dunbar's other primates) it is called the *upper critical mass*[1]. Its size turns out to depend on the discipline involved – chemistry, physics, archaeology and mathematics, for example, all have different upper critical masses. Once the group exceeds this size it tends to fragment into subgroups. The quadratic term associated with interactions between individuals is no more, as scientists cannot interact meaningfully across the entire department. Instead there is a quadratic term associated with the number of interactions between the various subgroups. However, this tends to be weaker than the interactions between individuals, decreasing as the upper critical mass increases[1]. Beyond the upper critical mass the dependency of quality on quantity is much reduced. The model thus predicts an average or expected piecewise linear relationship between quality s and quantity N in research,

$$\langle s \rangle = a_2 + b_2 N \qquad (2)$$

where now N is above the critical mass. The a_2 and b_2 here are different from the a and b in equation (1); the line has a much shallower slope.

Since these equations only predict the average dependencies of research quality on group size, and since they take account of average interactions between agents, it is common to speak of *mean field theory* (borrowing parlance from statistical physics).

Theory is all very well; but this is a theory which it is possible to test. To test it, empirical data are required. Happily, the UK's Research Assessment Exercise (RAE) provides plenty.

For those not familiar with it, the RAE takes place about every 5 years. It is an assessment of the quality of university research which the government uses to decide how funds should be allocated between different universities, institutions and departments. Not

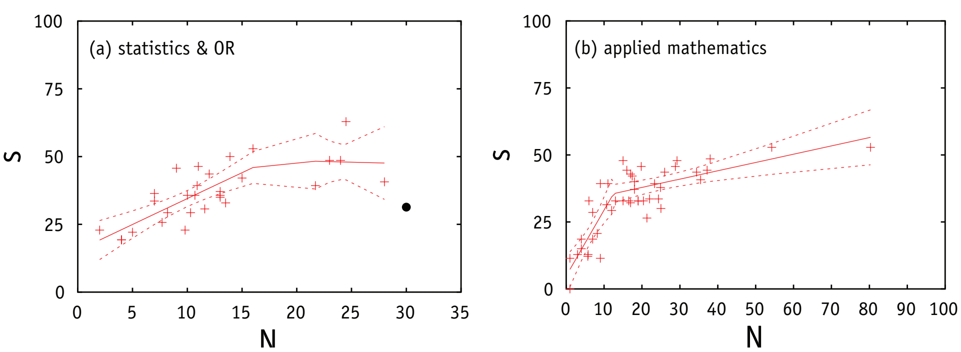


Figure 1. Quality as a function of quantity for research groups in (a) statistics and operational research and (b) applied mathematics, with best fits to the expected behaviour discussed in the text. (The dashed curves represent 95% confidence intervals for these fits.) The black dot in (a) represents a joint submission by Edinburgh and Heriot-Watt universities and is considered an outlier. The coefficients of determination are R^2 = 0.60 and 0.74 and the data pass the Kolmogorov–Smirnov normality test. The breakpoint or upper critical mass is N_c = 17 ± 6 for statistics and operational research, compared to N_c = 13 ± 2 for applied mathematics

surprisingly, it is taken very seriously indeed by university managers and by researchers themselves.

The most recent RAE, in 2008, sought to measure the quality of research coming from various groups in 67 academic disciplines in universities and research institutes across the country. Research was scrutinised to determine the proportions carried out at each of five levels:

A policy of concentrating resources into a few larger groups is unjustified

- 4*: World-leading research;
- 3*: Internationally excellent research;
- 2*: Research that is internationally recognised;
- 1*: research recognised at a national level;
- unclassified research.

Following the RAE, a formula is used to determine how funding is distributed to higher education institutes for the subsequent years. The original formula used by the Higher Education Funding Council for England valued 4* and 3* research seven and three times more highly than 2* research and allocated no funding to lower-quality research.

So how large should a group of (say) statisticians be before communication problems start to set in? One of the assessment divisions of the 2008 RAE was the Statistics and Operational Research unit. It received 30 submissions, comprising 389 individuals in groups of sizes ranging from $N = 2$ to $N = 30$. The mean group size was 13. It is useful to compare it to the Applied Mathematics unit, because of the similarities between the disciplines. Plots of quality of research, as assessed by the RAE, against number of people in the research groups are given in Figure 1. There are clear correlations between quality and quantity in each case, and each of them fits our theory very nicely. For each discipline we see, as predicted, two straight lines, of differing slopes. The point at which they intersect is N_c, the upper critical mass. Below the critical mass,

Table 1. Upper critical mass estimates for a selection of academic disciplines. When there are more than N_c researchers in a group, quality improvement per head starts to decline

Research discipline	N_c
Pure mathematics	≤4
Foreign languages	6 ± 1
History of art, performing arts, communication and music	9 ± 2
Agricultural sciences	10 ± 3
Economics/econometrics	11 ± 3
Applied mathematics	13 ± 2
Architecture/planning	14 ± 3
Sociology	14 ± 4
Archaeology	17 ± 3
Statistics and operational research	17 ± 6
Nursing	18 ± 5
Philosophy/theology	19 ± 3
Biology	21 ± 4
History	25 ± 5
Politics/international relations	25 ± 5
Physics	25 ± 5
Art and design	25 ± 8
Education	29 ± 5
Geography/environment	30 ± 3
Law	31 ± 4
English (in the UK)	32 ± 3
Chemistry	36 ± 13
Medical sciences	41 ± 8
Business/management	48 ± 8

quality of research increases rapidly with the size of the group; above the critical mass the increase is much slower.

And, as we can see, the upper critical mass for statisticians is 17 (plus or minus 6). This is the best size for any statistical research group. For applied mathematicians it is 13 (plus or minus 2). Once a research group has attained this size, it does not really help much (nor does it hurt) to add more and more researchers. In other words, a continual policy of concentrating resources into fewer, larger groups is unjustified.

It is, of course, possible to fit the data to other mathematical formulae. But the fits described above have the advantage of being sourced in a microscopic, agent-based model, permitting an interpretation of cause and effect. A short list of estimates for the upper critical masses of a selection of science disciplines is given in Table 1. (See Harrison[1] and Kenna and Berche[2, 7] for more extensive lists and for fits to other functional forms.)

It is perhaps worth noting that the table agrees, to some extent, with preconceived notions of the way that different academic disciplines behave. Pure mathematicians have the smallest "best group size"; and they are typically pictured as loners, sometimes antisocial ones. Perhaps less expected is the position of foreign language research, which also seems to work best in very small groups. The biggest "best group size" belongs to the medical sciences, and the complexity of much medical research also makes this expected. The number of co-authors of medical papers is also typically large, which is anecdotal confirmation of the result.

The RAE and its successor, the Research Excellence Framework, which will replace it in 2014, are of enormous importance to the entire academic research community in the UK. Under the Research Excellence Framework statistics and operational research will be considered as part of the overall mathematical

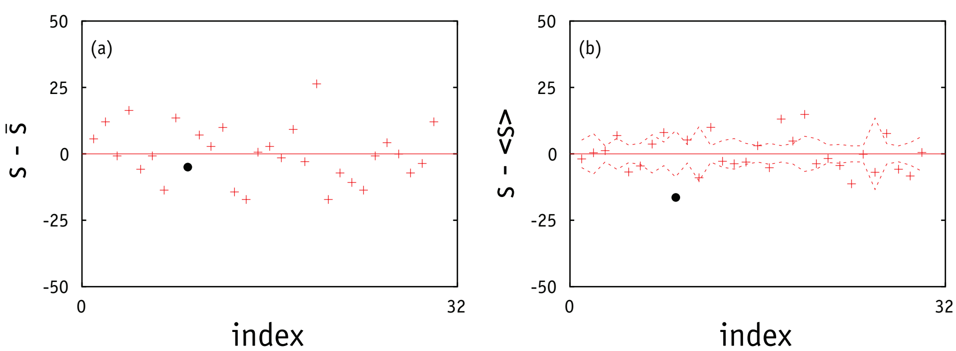
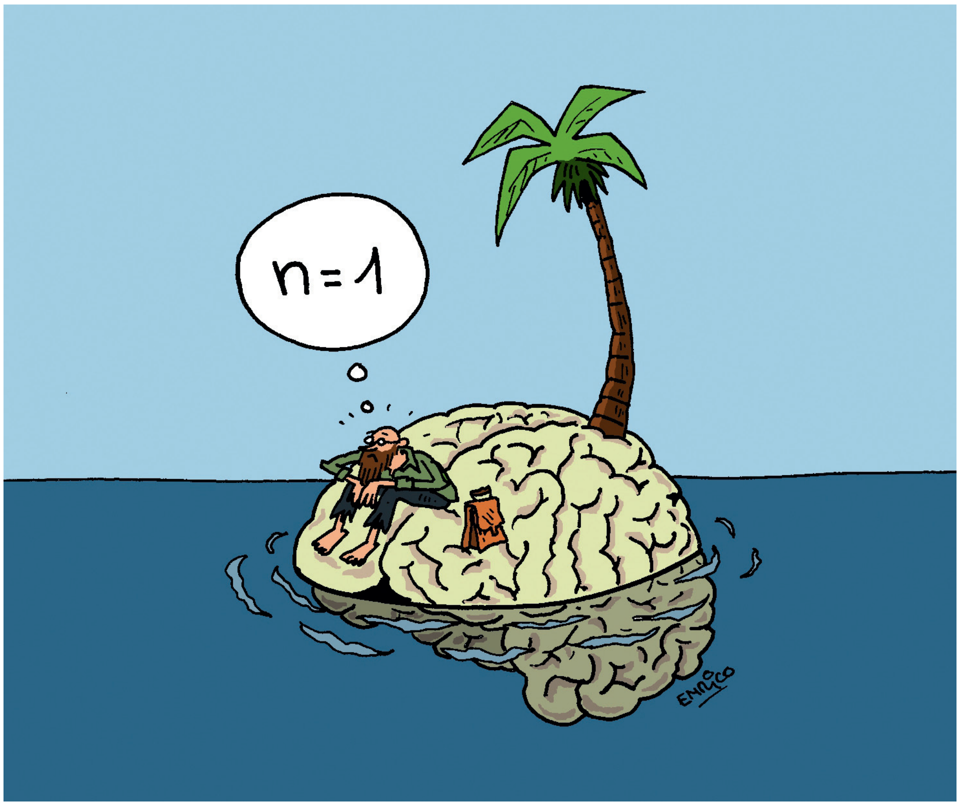


Figure 2. (a) Quality measurements normalised to the overall mean for statistics and operational research groups at RAE 2008. (b) The same data normalised to the expectation ‹s›, taking size into account. Here distance along the x-axis corresponds to alphabetical listing of the universities to which the groups belong



sciences unit of assessment, which also includes pure and applied mathematics. Thus RAE 2008 provided the only opportunity to measure critical mass for statistics and operational research as a single discipline, distinct from their mathematics cousins.

After the RAE, universities and research groups are ranked in the media according to performance. The ranking, though, is done without taking size of the group into account. Figure 2(a) illustrates such a ranking; the different research groups are ranged above or below an average irrespective of how big each group is. But as we have seen, it is clear that such plots (and rankings) do not compare like with like: they do not take size into account. It is perhaps more sensible – and fairer – to take the expected performance related to their size into account as well; this is shown in Figure 2(b). The data is more tightly distributed; the variation caused by different group size has been eliminated. The tighter distribution of Figure 2(b) (where the range is 26.1 and standard deviation is 6.7) illustrates the superiority of the underlying size-based model over that of Figure 2(a) (with range 43.6 and standard deviation 10.7).

By considering how the strength of research groups might be expected to improve by either adding more researchers or transferring researchers between groups, one can show that there is also a *lower critical mass*, which more closely corresponds to the traditional notion, although it is not a threshold[1]. This may be interpreted as the minimum size a group or department should achieve for it to be viable in the long term. The lower critical mass turns out to be pleasantly simple: it is half the value of the upper critical mass. Groups whose size is smaller than this are vulnerable and should seek to achieve the lower critical mass to avoid extinction. Therefore, statistics and operational research groups should strive to achieve a size of at least $N_c/2 \approx 9$ members, and should be happy if they have over $N_c \approx 17$ staff. But beyond that, it does not much matter how many staff the group has – more staff beyond the critical number will tend only to increase the quantity of research, but not its quality.

So what lessons can be drawn from this type of analysis? It is worth reiterating that there is no threshold group size beyond which research quality significantly improves. On the contrary, there is a measurable upper critical mass, beyond which the Ringelmann effect kicks in. Secondly, having established that a community of researchers is greater than the sum of its parts, it is clear that facilitation of communication should form an important management policy in academia. For example, while modern managerial experiments such as distance working or "hotdesking" may be reasonably employed in certain industries, these would have a negative effect for researchers. Instead, for them, putting individuals' office spaces close to each other to facilitate spontaneous two-way interaction between as many people as possible is important. Researchers have always known this to be the case. Critical mass analyses along the lines presented here may help them take this message to university managers and to policy makers.

In short, then, bigger is indeed better, but only up to a point. Give us sufficient staff and help us communicate with each other. But there is no need to take this policy to extremes.

References

1. Harrison M. (2009) Does high-quality research require "critical mass"? In D. Pontikakis, D. Kyriakou and R. van Baval (eds), *The Question of R&D Specialisation: Perspectives and Policy Implications*, pp. 53–54. Luxembourg: JRC Scientific and Technical Reports for European Commission.
2. Kenna, R. and Berche, B. (2010) The extensive nature of group quality. *Europhysics Letters*, **90**, 58002.
3. Kenna, R. and Berche, B. (2010) The critical mass and the dependency of research quality on group size. *Scientometrics*, **86**, 527–540.
4. Evidence (2010) *The Future of the UK University Research Base*. Report for Universities UK. Available at `http://www.universitiesuk.ac.uk/Publications/Documents/UUK-FutureOfResearch-LiteratureReview.pdf`
5. Ringelmann, M. (1913) Recherches sur les moteurs animés: travails de l'homme. *Annales del'Institut National Agronomique*, **2**, 2–39.
6. Dunbar, R. I. M. (1992) Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, **22**, 469–493.
7. Kenna, R. and Berche, B. (2011) Statistics of statisticians: Critical mass of statistics and operational research groups in the UK. Accepted for publication in *International Journal of Modern Physics*. (An earlier version is available at `http://arxiv.org/abs/1102.4914`)

Professor Ralph Kenna is Deputy Director of the Applied Mathematics Research Centre at Coventry University. Professor Bertrand Berche is head of the Department of Physics at the Statistical Physics Group, Institut Jean Lamour, at Nancy University.