Women in Science Without Borders

South African Journal *of* Science

Time to say…

It is often time to say something significant, and when we do, spatial terms come to the fore: 'This is not a borderline case' or 'Welcome to new horizons' or 'Here's where we draw the line'.

'This is not a borderline case'. Women in Science Without Borders (WiSWB) is a relatively new initiative conceived to empower women in science, whether young or older, by encouraging research excellence and impact and challenging gender inequality. It is hoped that WiSWB will attract attention and help to address concerns of women already working, or studying, in the basic and applied sciences. In 2018, WiSWB and the University of Johannesburg worked together to host the 2nd International WiSWB indaba (conference) with the support of the South African Council for Scientific and Industrial Research (CSIR).

As the SAJS does not publish special issues or conference proceedings, an agreement was reached that would result in the best conference papers, having gone through the normal processes of Associate Editor and double-blind review, being published in this themed issue, which has been sponsored by the DST-NRF Centre of Excellence in Mathematical and Statistical Sciences. Six papers from the WiSWB indaba – three peer-reviewed Research Articles and three Commentaries – comprise this Women in Science themed issue. Four papers reflect very helpfully on the vexatious issue of the persistent under-representation of women in the natural/basic sciences (with the possible exception of the biosciences). But career challenges, double and equally demanding roles, the need to measure under-representation in, for example, mathematics or computational science, all receive considered attention. Placing the issue of gender into a broader context, one of the papers ends with this observation: '…it is recommended that the African Union champions the setting up of a taskforce to coordinate the collection of a detailed research profile of researchers and research outputs. This initiative will assist in meaningful consolidation of R&D data to understand the status quo, and to identify locations for interventions as appropriate, to provide tangible inputs to help make the African Renaissance a reality.'

'Welcome to new horizons'. The second part of this Leader also addresses a matter of gender and – just as significantly – a key change in the editorial team. For the first time in its 115 years, the *South African Journal of Science* will have a woman scholar serving as its Editor-in-Chief. It gives me great pleasure to introduce Professor Jane Carruthers, who will start as Editor-in-Chief from April, and to welcome her warmly to her new position and to the Journal.

Jane is an environmental historian and an Emeritus Professor of the University of South Africa, where she pioneered her discipline. Her book *The Kruger National Park: A Social and Political History* has become a standard reference worldwide. Her research areas include environmental justice and the history of national parks, the origins of wildlife management and game ranching, land restitution issues, landscape design and ecology, heritage, cartography, and transnational history. Jane is widely published in books and scholarly journals. Her latest book is *National Park Science: A Century of Research in South Africa*. Jane is a member of many editorial boards and has been awarded a number of prizes. She is a Fellow of the Royal Society of South Africa and a Member of the Academy of Science of South Africa, past president of the Southern African Historical Society, and of the International Consortium of Environmental History Organizations. For her work as the initial chair of the Academic Advisory Board of the Rachel Carson Center for Environment and Society, she was named an honorary Carson fellow

in December 2014 and a publication honouring Jane can be found on their site here. The publication is a notable testimony to her scholarship and contributions to environmental history (and other histories) in South Africa, and also a valuable read. Very best wishes for the years ahead, Jane.

'Here's where we draw the line'. This is my 38th Leader and the last I shall write. That line drawn, it is interesting to note the focus of my first Leader, written in January 2013, and what might have changed since then.

The Leader was titled 'Education, training and innovation in the National Development Plan – 2030' (NDP) and considered some of the challenges that education and science faced, at that time, in relation to helping achieve the aims of the newly released Plan. At the beginning of 2013, 2030 was about 18 years away; it is now closer to 10 years away. So what has changed over the past 6–7 years?

In 2013, President Zuma was of the view that the NDP would result in, amongst other things, a society in which '…all will have water, electricity, good health, libraries, good schools, roads, good hospitals and clinics, safety and security, recreational facilities, a growing economy and jobs'. The reality is that almost none of these admirable aims has come close to being realised.

In some instances, things have become worse – starting with the sad reality that very few people even talk about the NDP any longer. It is true that more people have access to piped water than previously, but electricity is often not delivered – even to those who are part of the distribution system. Unemployment remains high, and the economy is not growing in pace with the national need.

The essential drivers of the changes envisioned in the NDP – primary, secondary and tertiary education – have not improved. In South African schools, students perform very close to, or at the bottom of, international math, science and language tests. The real Grade-12 pass rate is not the 78.2% announced for 2018 but somewhere between 37.5% and 40% because about half of Grade-1 entrants never reach Grade 12. Worse, perhaps, is the reality that about 3.2 million young people in the 18–24 age cohort are still not in education, employment or training. Valid historical and related reasons can be (and are) cited to explain these conditions, but former President Zuma and those linked to him also have much for which to account.

Reasons for hope remain, though. We still have 10 years in which to move towards the goals of the NDP. Universities are graduating thousands of students each year, most of whom are able to contribute to development. The possibility of honest government is closer to becoming reality than it has been for some time. Now is not the time at which to abandon hope.

There is one further '*line to be drawn*' – that which links together the people who have worked with and supported me for almost 6.5 years: the Executive Officers of ASSAf, Roseanne Diab and Himla Soodyall; the members of the Editorial Advisory Board and its Chair Johann Mouton; Susan Veldsman, Linda Fick, Nadine van der Merwe and Sbonga Dlamini have been constant and selfless colleagues and friends – as have other colleagues in the Scholarly Publishing Unit of ASSAf. The Journal's Associate Editors have been key to maintaining the Journal's scientific integrity – and a pleasure with whom to work. And then, of course, there are the Journal's readers who even read Leaders. No Editor-in-Chief could survive without the consideration of such an interdependent community – of which it has been an honour to be a part.

AUTHOR:
Julien Benoit[1] iD

AFFILIATION:
[1]Evolutionary Studies Institute,
School of Geosciences, University
of the Witwatersrand, Johannesburg,
South Africa

CORRESPONDENCE TO:
Julien Benoit

EMAIL:
julien.benoit@wits.ac.za

Parental care or opportunism in South African Triassic cynodonts?

In a paper published in *Nature* in 2018, Hoffman and Rowe[1] describe the discovery of an adult tritylodontid cynodont, *Kayentatherium,* from the Jurassic of the Kayenta Formation (Arizona, USA), accompanied by at least 38 perinatal juveniles, all at the same very early stage of development. Such a high number of juveniles in one clutch is found only in a handful of oviparous reptiles, and never in viviparous or ovoviviparous species[1], suggesting that these cynodonts laid eggs. As tritylodontids are amongst the closest relatives to Mammaliaformes, and sometimes even reconstructed as their sister clade[2], this textbook changing discovery implies that all non-mammaliaform synapsids had an essentially reptilian-like reproductive biology[1].

Hoffman and Rowe's[1] discovery forces a reappraisal of the interpretation of the rich fossil record of parental care in South African non-mammaliaform cynodonts[3].

First, the implications of Hoffman and Rowe's[1] discovery for the evolution of parental care and lactation need to be discussed. With at least 38 hatchlings to feed, it can be safely assumed that lactation was not the main source of nutrients in *Kayentatherium* neonates (nor, by extension, in all more basal non-mammaliaform synapsids). Instead, perinatal juveniles of *Kayentatherium* already had functional teeth despite their early age, which (1) impedes lactation[4] and (2) strongly suggests that they were self-sufficient at birth and capable of foraging nearby vegetation[5]. Extensive parental care was thus most likely not essential to newborn survival in *Kayentatherium*. In contrast, early Mammaliaformes are characterised by (1) a diphyodont dentition, which indicates lactation[4], and (2) a remarkably smaller body mass compared to their non-mammaliaform cynodont ancestors, which imposes a physical limitation to the size of the eggs and, incidentally, the yolk they contain, thus enforcing parental care and lactation[6].
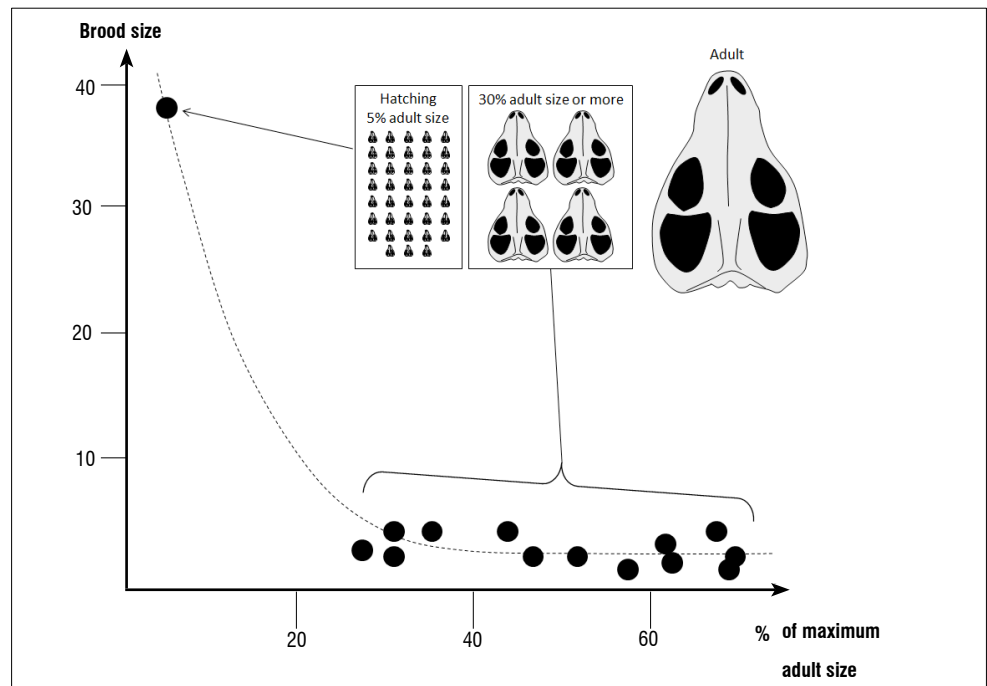
Sources: Data from Hoffman and Rowe[1] and Jasinoski and Abdala[3]

Figure 1: The drop in non-mammaliaform cynodont demographics.

The presence of an adult *Kayentatherium* next to perinatal juveniles indeed suggests that the eggs were incubated. This observation makes this specimen a landmark for any discussion about the origin of endothermy in mammalian ancestry, but more importantly for this discussion, it also suggests that a mammary gland could have produced a milk-like secretion as in modern monotremes that use their mammary glands to moisturise their leathery eggs to prevent desiccation.[7] Therefore, despite the fact that *Kayentatherium* likely did not suckle its young, the possibility that this taxon possessed an ancestral mammary gland that secreted a moisturising milk-like fluid cannot be excluded, thus not completely invalidating hypotheses that similar structures were already present in tritylodontids.[8]

The high number of juveniles accompanying the *Kayentatherium* described by Hoffman and Rowe[1] sharply contrasts with the rest of the fossil record of non-mammaliaform cynodont adult–juvenile associations, primarily documented in the South African Karoo genera *Galesaurus* and *Thrinaxodon*[3]. Before Hoffman and Rowe's[1] discovery, the maximum number of juvenile non-mammaliaform cynodonts that were found associated with an adult was four[3]. Similarly, an adult–juvenile association is documented for *Heleosaurus*, a more basal pelycosaur synapsid, for which four juveniles were found together alongside an adult.[9] The *Kayentatherium* juveniles are substantially smaller, and most likely less mature, than any other juvenile non-mammaliaform cynodonts previously recovered in a parent and offspring association. The *Kayentatherium* juvenile skull size is about 5% of the maximum

adult size for the taxon[1], whereas the skull size of the otherwise youngest non-mammaliaform cynodont recorded alongside an adult is 31% of the maximum adult size[3]. A burrow cast found with three juvenile *Langbergia* (but no adult) contains an individual with a skull 27% of the maximum adult size for this taxon.[3] Juvenile *Heleosaurus* are 39–54% the size of the associated adult.[9] All of these specimens are in a very advanced growth stage compared to the *Kayentatherium* perinates described by Hoffman and Rowe[1]. If these demographics are representative of the general condition in non-mammaliaform cynodonts, it means that from the time interval during which juveniles were growing from 5% to about 30% adult size, some 90% of the litter was lost. This high level of early juvenile mortality would further support the absence of lactation and parental care in non-mammaliaform cynodonts. In this case, far from evidencing parental care, the South African fossil record would actually support that non-mammaliaform cynodonts were not extensively taking care of their young beyond incubation.

A revised interpretation

Given that Mammaliaformes is a clade deeply nested within Cynodontia, so far the null hypothesis has historically been to consider the association of juveniles and adults in non-mammaliaform cynodonts as evidence for parental care. However, amongst the 12 specimens that show aggregation of many individuals (one in *Kayentatherium*, one in *Langbergia*, eight in *Thrinaxodon* and two in *Galesaurus*), only 4 actually include an adult associated with juveniles, whereas 3 show interspecific associations.[3,10] Thus, there is nearly as much evidence for parental care in non-mammaliaform cynodonts as there is evidence for interspecific shelter-sharing in the Triassic, advocating a case for the frequent co-occurrence of adults and juveniles actually being the result of opportunistic shelter-sharing (as some mammals do today[11]) as opposed to parental care. For the smallest individuals of any age or species, aggregation might have been advantageous as it provides protection against predators and extreme temperature changes in the dry and highly seasonal conditions of the Triassic.[3,10,12,13] The smallest individuals would be tolerated by the largest ones as they emit more body heat as a result of their high surface to volume ratio, thereby aiding in heat conservation during periods of hibernation or aestivation. Following this hypothesis, these fossils could instead suggest simple mutualistic behaviour rather than elaborate parental care.

Considering the facts presented above, a new interpretation of the aggregation of non-mammaliaform cynodonts is proposed here. Instead of parental care, these associations could instead actually offer evidence that opportunistic inter- and intraspecific shelter-sharing was tolerated as it provided a selective advantage to all participating parties.

Acknowledgements

References

1. Hoffman EA, Rowe TB. Jurassic stem-mammal perinates and the origin of mammalian reproduction and growth. Nature. 2018;561:104–108. https://doi.org/10.1038/s41586-018-0441-3

2. Ruta M, Botha-Brink J, Mitchell SA, Benton MJ. The radiation of cynodonts and the ground plan of mammalian morphological diversity. Proc Biol Sci. 2013;280(1769), Art. #20131865, 10 pages. https://doi.org/10.1098/rspb.2013.1865

3. Jasinoski SC, Abdala F. Aggregations and parental care in the Early Triassic basal cynodonts *Galesaurus planiceps* and *Thrinaxodon liorhinus*. PeerJ. 2017;5, e2875, 35 pages. https://doi.org/10.7717/peerj.2875

4. Luo Z-X, Kielan-Jaworowska Z, Cifelli RL. Evolution of dental replacement in mammals. Bull Carnegie Mus Nat Hist. 2004;36:159–175.

5. Koteja P. Energy assimilation, parental care and the evolution of endothermy. Proc R Soc Lond B. 2000;267:479–484. https://doi.org/10.1098/rspb.2000.1025

6. Hopson JA. Endothermy, small size, and the origin of mammalian reproduction. Am Nat. 1973;446–452. https://doi.org/10.1086/282846

7. Oftedal OT. The origin of lactation as a water source for parchment-shelled eggs. J Mammary Gland Biol Neoplasia. 2002;7:253–266. https://doi.org/10.1023/A:1022848632125

8. Benoit J, Manger PR, Rubidge BS. Palaeoneurological clues to the evolution of defining mammalian soft tissue traits. Sci Rep. 2016;6, Art. #25604, 10 pages. https://doi.org/10.1038/srep25604

9. Botha-Brink J, Modesto SP. A mixed-age classed 'pelycosaur' aggregation from South Africa: Earliest evidence of parental care in amniotes? Proc Biol Sci. 2007;274(1627):2829–2834. https://doi.org/10.1098/rspb.2007.0803

10. Fernandez V, Abdala F, Carlson KJ, Cook DC, Rubidge BS, Yates A, et al. Synchrotron reveals early Triassic odd couple: Injured amphibian and aestivating therapsid share burrow. PLoS ONE. 2013;8(6), e64978, 7 pages. https://doi.org/10.1371/journal.pone.0064978

11. Reichman O, Smith SC. Burrows and burrowing behavior by mammals. Curr Mammal. 1990;2:197–244.

12. Botha-Brink J. Burrowing in *Lystrosaurus*: Preadaptation to a postextinction environment? J Vert Paleontol. 2017;37(5), e1365080. https://doi.org/10.1080/02724634.2017.1365080

13. Botha-Brink J, Huttenlocker A, Angielczyk KD, Codron D, Ruta M. Breeding young as a survival strategy during earth's greatest mass extinction. Sci Rep. 2016;6, Art. #24053, 9 pages. https://doi.org/10.1038/srep24053

AUTHOR:
J. Francis Thackeray[1] iD

AFFILIATION:
[1]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Francis Thackeray

EMAIL:
mrsples@global.co.za

Alpha and sigma taxonomy of *Lystrosaurus murrayi* and *L. declivis*, Triassic dicynodonts (Therapsida) from the Karoo Basin, South Africa

Permo-Triassic dicynodont (Therapsid) fossils attributed to the genus *Lystrosaurus*[1] are known from the South African Karoo Basin, including two Triassic taxa referred to as *L. (Dicynodon) murrayi*[2] and *L. (Ptychognathus) declivis*[3]. Distinction of the two taxa has been recognised inter alia by Brink[4], Cluver and King[5], Botha and Smith[6], Grine et al.[7] and Botha-Brink et al.[8] However, in this instance, the appropriateness of alpha taxonomy may be questioned, recognising that boundaries between these species of *Lystrosaurus* are not necessarily clear, especially for specimens close to the time of divergence postdating the Permo-Triassic boundary, 252 million years ago. The question has arisen as to whether only one species occurred.[9] Thackeray[10] proposed the following hypotheses:

- H_T – that one species existed and

- H_{BB} – that there were two taxa, clearly distinct at the species level.

Both hypotheses need to be considered in the light of a statement by Camp[11]: 'traditional data were unable to distinguish between *L. murrayi* and *L. declivis* without overlapping measurements in regions where the geometric morphometric analyses indicated they differed most'. Irrespective of which statistical methods are used for taxonomic purposes, it is essential to have some kind of definition of a species that can be applied in palaeontological contexts, to address opposing views held by 'alpha taxonomists' who assume clear boundaries between species[12], and 'sigma taxonomists'[13-15] who do not.

Appeals have been made for the use of a statistical (probabilistic) definition of a species, using morphometric analyses whereby measurements of any Specimen A are compared (through least squares linear regression) with those of any Specimen B, and vice versa.[14-16] The standard error of the m coefficient (sem) in equations of the form $y = mx + c$ is calculated to reflect the degree of morphological similarity or dissimilarity when pairs of crania are compared. Notably, for large samples of conspecific modern taxa (used for reference purposes), there is a normal distribution of log-transformed sem values.[17]

In the case of conspecifics, a mean log sem value of -1.61 ± 0.1 has been calculated from anatomical measurements obtained from more than 70 modern taxa.[14,17] As it has been replicated from independent sets of data, for extinct and extant taxa[14, 18], it can be considered to reflect the typical degree of morphological variability in vertebrate taxa across evolutionary time and geographical space. As such, a mean log sem value (T) of -1.61 ± 0.1 has been proposed as a probabilistic definition of a species applicable in palaeontological contexts.[14] This morphometric approach has the potential to address taxonomic issues regarding Triassic specimens of *Lystrosaurus*. Examples are given here for selected specimens.

Ideally, complete and well-preserved holotypes for *L. murrayi* and *L. declivis* should be used for taxonomic purposes. Unfortunately, the holotypes for these two species described respectively by Huxley[2] and Owen[3], and curated at the Natural History Museum in London, are not sufficiently complete to allow detailed morphometric analyses based on many cranial dimensions. On the basis of morphology alone, Broom[19] questioned whether the (incomplete) holotype of *L. declivis*[3] was the same species represented by the (incomplete) holotype of *L. murrayi*[2]. In this study, cranial measurements of other (well-preserved, more complete) specimens attributed to either *L. murrayi* or *L. declivis* are analysed with a view to identifying potential neotypes on the basis of log sem statistics.

Materials

The specimens in this study are well-preserved crania of *Lystrosaurus*, selected on the basis of a relatively high number of measurable dimensions among the specimens classified by Camp[11] who indicates whether there is consensus between her identifications and those of others. These specimens are:

- NMNH 23349 attributed to *L. murrayi* (consensus identification), curated at the National Museum of Natural History, Smithsonian Institution, Washington.

- BMNH 36221 attributed to *L. declivis* (consensus identification), curated at the Natural History Museum in London.

- NMNH 12690, the identification of which is uncertain (*L. murrayi/declivis*), curated at the National Museum of Natural History, Smithsonian Institution, Washington.

These specimens are sufficiently well preserved to measure about 30 dimensions based on anatomical landmarks.[11] The 20 variables common to all three specimens are listed in Table 1.

Cranial dimensions used in this study. Reference numbers (with definitions) are those given by Camp[11]

Reference number	Definition
1	Basal cranial length
2	Anterior skull length
3	Posterior skull length
4	Pineal foramen length
5/6	Arc/chord length of frontal
7/8	Arc/chord length of nasal
9/10	Arc/chord length of premaxilla
12	Orbit height
13	Anterior-posterior snout length
14	Anterior-posterior length between anterior borders of orbit and nares
15	Dorsoventral height between ventral margins of orbit and nares
16	Dorsoventral snout height
17	Dorsoventral height of caniniform process
18	Minimum interorbital width
19	Maximum width across lateral margins of prefrontals
20	Minimum internarial width
21	Maximum width across lateral margins of caniniform processes
22	Maximum width across anterior quadrate foramina
23	Maximum anterior-posterior length of temporal fossae
24	Maximum width of temporal fossae
25	Minimum width of braincase
27	Maximum width across lateral margins of squamosals
39	Minimum medio-lateral width between tusks
42	Anterior-posterior length of interpterygoid fossa
43	Anterior-posterior length from interpterygoid fossa to occipital condyles
45	Anterior-posterior length from tip of snout to posterior margin of quadrate
46	Minimum medio-lateral width of pterygoids

Methods

Pairwise comparisons of cranial measurements are made using least squares linear regression analysis to calculate log sem values associated with specimens attributed to *L. murrayi* (a 'consensus' identification for NMNH 23349) and to *L. declivis* (a 'consensus' identification for BMNH 36221). Log sem values are also calculated for comparisons with NMNH 12690, the identification of which is not certain.

Results

Log sem values for pairwise comparisons are as follows, in increasing order:

- NMNH 12690 (*x*-axis) versus NMNH 23349 (*y*-axis)

  Log sem = **-1.474**

- NMNH 23349 (*x*-axis) versus NMNH 12690 (*y*-axis)

  Log sem = **-1.419**

- NMNH 12690 (*x*-axis) versus BMNH 36221 (*y*-axis)

  Log sem = **-1.104**

- NMNH 23349 (*x*-axis) versus BMNH 36221 (*y*-axis)

  Log sem = **-1.017**

- BMNH 36221 (*x*-axis) versus NMNH 23349 (*y*-axis)

  Log sem = **-1.014**

- BMNH 36221 (*x*-axis) versus NMNH 12690 (*y*-axis)

  Log sem = **-1.008**

Discussion and conclusions

Relatively low log sem values (ranging between -1.474 and -1.419) are obtained when NMNH 23349 (attributed to *L. murrayi*, 'consensus identification') is compared against NMNH 12690 (with uncertain status). In relation to the probabilistic definition of a species (T=-1.61±0.1), the two specimens may be conspecific because the mean log sem value of -1.446 falls within the 95% confidence limits for a species. However, this log sem value is near the upper limit, such that in terms of degrees of variation, it is relatively close to being different at a species level. This proximity is important in the context of a view that there may not be a distinct boundary between *L. murrayi* and *L. declivis*.

On the basis of an impressive sample of almost 200 crania of Triassic *Lystrosaurus* studied by Botha-Brink et al.[8], it has been demonstrated that the mean basal skull length of specimens attributed to *L. murrayi* is not significantly different from the mean basal skull length dimension of specimens attributed to *L. declivis*[10]. The mean values (and the associated standard deviations) are almost identical, as if one species was being sampled from the same kind of populations. In addition to the morphometric data presented here, based on as many as 20 variables per specimen, this result is relevant to the possibility that no distinct boundary exists between the two taxa at a species level, in the context of both temporal and spatial variability.

Do at least some crania represent different taxa at the level of a species? High log sem values (ranging between -1.020 and -1.014) are in fact obtained when BMNH 36221 (attributed to *L. declivis*, 'consensus' identification) is compared against NMNH 23349 (*L. murrayi*, 'consensus' identification). In relation to the probabilistic definition of a species (T = -1.61±0.1), these particular well-preserved specimens have a high probability of being different at a species level as the mean log sem value of -1.017 falls undoubtedly well outside the upper 95% confidence limit for a species. As such, these results for at least two particular specimens serve to refute the hypothesis (H$_T$) that all specimens attributed to either *L. declivis* or to *L. murrayi* represent only one species.[10] It could be concluded that *L. murrayi* is indeed represented by NMNH 23349, and that *L. declivis* is represented by BMNH 36221.

However, not all comparisons may necessarily reflect differences at a species level. In the case of closely related taxa, log sem values can serve to quantify *degrees* of similarity if not distinguish differences in *kind*. For instance, the relatively low mean log sem value of -1.446 obtained for the comparison between NMNH 12690 (previously identified uncertainly as *Lystrosaurus murray/declivis*) and NMNH 23349 indicates that the former specimen is relatively closer to *L. murrayi* (represented by NMNH 23349) than it is to *L. declivis* (represented by BMNH 36221), as a higher log sem value is obtained when the latter is compared to NMNH 12690.

Specimens such as NMNH 12690, whose identification has been uncertain (*L. murray/declivis*), may be those which fall within a spectrum of variability associated with two species which diverged close to the time of the Permo-Triassic boundary, 252 million years ago, but which may have hybridised. Hybridisation is recognised as a phenomenon that is much more common than previously thought.[15]

Because NMNH 23349 and BMNH 36221 are recognised here as being distinct at a species level for Triassic specimens of *Lystrosaurus*, on the basis of a probabilistic definition of a species (T=-1.61±0.1), and because they are sufficiently complete to allow many measurements per specimen, they are recommended as potential neotypes for *L. murrayi* and *L. declivis*, respectively.

When and if neotypes are formally described, they can be used as reference specimens to assess the *probability* that other Triassic fossils of *Lystrosaurus* do or do not belong to *L. murrayi* or *L. declivis*. The examples given above relate to sigma taxonomy[13-15], where sigma is the Greek letter Σ (S for spectrum), as opposed to alpha taxonomy[12] which assumes clear boundaries between species.

Acknowledgements

References

1. Cope ED. On the skull of the dicynodont Reptilia, *Lystrosaurus frontosus* from Cape Colony. Proc Am Philos Soc. 1870;11:419.

2. Huxley TH. On a new species of dicynodont (*D. murrayi*) from near Colesburg, South Africa; and on the structure of the skull in dicynodonts. Q J Geol Soc Lond. 1859;15:649–659. https://doi.org/10.1144/GSL.JGS.1859.015.01-02.63

3. Owen R. On some reptilian fossils from South Africa. Q J Geol Soc Lond. 1860;16:49–54. https://doi.org/10.1144/GSL.JGS.1860.016.01-02.07

4. Brink AA. Illustrated bibliographical catalogue of the Synapsida. Handbook 10, Part I. Pretoria: Government Printer; 1982.

5. Cluver MA, King GM. A reassessment of the relationships of Permian Dicynodontia (Reptilia: Therapsida) and a new classification of dicynodonts. Ann S Afr Mus. 1983;91:195–273.

6. Botha J, Smith RMH. *Lystrosaurus* species composition across the Permian-Triassic boundary in the Karoo Basin of South Africa. Lethaia. 2007;40:125–137. https://doi.org/10.1111/j.1502-3931.2007.00011.x

7. Grine FE, Forster CA, Cluver MA, Geogri JA. Cranial variability, ontogeny, and taxonomy of *Lystrosaurus* from the Karoo Basin of South Africa. In: Carrano MT, Gaudin TJ, Blob RW, Wi JR, editors. Amniote paleobiology: Perspectives on the evolution of mammals, birds, and reptiles. Chicago, IL: University of Chicago Press; 2006. p. 432–503.

8. Botha-Brink J, Codron D, Huttenlocker AK, Angielczyk KD, Rutas M. Breeding young as a survival strategy during earth's greatest mass extinction. Sci Rep. 2016;6, Art. #24053, 9 pages. https://doi.org/10.1038/srep24053

9. Thackeray JF, Durand JF, Meyer L. Morphometric analysis of South African dicynodonts attributed to *Lystrosaurus murrayi* (Huxley, 1859) and *L. declivis* (Owen, 1860): Probabilities of conspecificity. Ann Transv Mus. 1998;36:413–420.

10. Thackeray JF. Do specimens attributed to *Lystrosaurus murrayi* and *L. declivis* (Triassic Therapsida) represent one species? S Afr J Sci. 2018;114(3/4), Art. #a0258, 2 pages. https://doi.org/10.17159/sajs.2018/a0258

11. Camp JA. Morphological variation and disparity in *Lystrosaurus* (Therapsida: Dicynodontia) [MS thesis]. Iowa City, IA: University of Iowa; 2010. https://doi.org/10.17077/etd.5mqepvmv

12. Mayr E, Linsley EG, Usinger RL. Methods and principles of systematic zoology. New York: McGraw-Hill; 1953.

13. Thackeray JF. Sigma taxonomy in relation to palaeoanthropology and the lack of clear boundaries between species. Proc Eur Soc Stud Hum Evol. 2015;4:220.

14. Thackeray JF, Dykes S. Morphometric analyses of hominoid crania, probabilities of conspecificity and an approximation of a biological species constant. HOMO. 2016;67(1):1–10. https://doi.org/10.1016/j.jchb.2015.09.003

15. Thackeray JF, Schrein CM. A probabilistic definition of a species, fuzzy boundaries and 'sigma taxonomy'. S Afr J Sci. 2017;113(5/6), Art #0206, 2 pages. https://doi.org/10.17159/sajs.2017/a0206

16. Thackeray JF, Odes E. Morphometric analysis of early Pleistocene African hominin crania in the context of a statistical (probabilistic) definition of a species. Antiquity. 2013;87(335). http://antiquity.ac.uk/projgall/thackeray335/

17. Thackeray JF. Approximation of a biological species constant? S Afr J Sci. 2007;103:489.

18. Thackeray JF. Human evolution from an African perspective, with reference to Charles Darwin and a probabilistic definition of a species (sigma taxonomy). Bulletin du Musée d'Anthropologie Préhistorique de Monaco. 2016;suppl.6:21–25.

19. Broom R. The mammal-like reptiles of South Africa and the origin of mammals. London: Witherby; 1932. p. 244.

AUTHOR:
Tomoko L. Kitagawa[1]

AFFILIATION:
[1] Mathematical Institute, University of Oxford, Oxford, UK

CORRESPONDENCE TO:
Tomoko L. Kitagawa

EMAIL:
tomoko.kitagawa@maths.ox.ac.uk

In search of a place in history for mathematics: A lecture series in South Africa

This Commentary was invited by the *South African Journal of Science* and emanates from a seminar presented by Dr Tomoko Kitagawa entitled 'The History of Mathematics: An Interdisciplinary Work in Humanities' given at the Department of Historical and Heritage Studies of the University of Pretoria on 20 September 2018. Dr Kitagawa received her PhD from Princeton University and went on to teach history at Harvard University. Prior to her appointment at Harvard, she also worked for the Ministry of Foreign Affairs of Japan. She is an author of five books in Japanese, including a national bestseller, and was also selected as one of the 100 most influential people in Japan in 2012 and one of the 100 most amazing women in Japan in 2015.

The history of mathematics entails how mathematical knowledge has developed over time. Not every university offers courses on this topic, and it could be more common to read about the history of math for pleasure. Why do academics need to pay attention to the history of math, and how do we conduct research on such an interdisciplinary subject?

History of mathematics

When mathematics started to become an independent domain of study in the mid-17th century, learned individuals were interested in looking for 'truth' (i.e. universal patterns that do not vary based on time and place) and began exchanging their ideas in writing. Some gathered various data independently to examine whether mathematical rules always hold true, and when someone claimed to have found such a pattern, others worked on proofs to evaluate it. This cycle of creating and proving led to the consolidation of a universal language that explains such universal truths.

The story of mathematics tells the process of finding truths at a basic level, i.e. the history of math is the history of fundamentals. However, the investigative process varies from one case to another and also differs from culture to culture. Let us first examine five approaches in studying the history of mathematics.

Five approaches to the study of the history of mathematics

1. Macroanalysis
2. Compare and contrast (regional analysis)
3. Accumulation of overall knowledge
4. Institutional analysis
5. Biographies and genealogies

The common method of presenting the history of math entails macroanalysis that examines significant discoveries and developments chronologically. University textbooks often take this approach[1-3], usually beginning in ancient times to create a historical narrative from several different parts of the world. However, as the timeline reaches the Age of Enlightenment, the focus is more exclusively on European mathematicians. Because the significance and consequences of 'scientific revolution' were undeniably large, it created a common impression that modern math originated in Europe and spread to other parts of the world.

This macro approach covers the history of math on a global scale, but it risks constructing a Eurocentric view. As George Gheverghese Joseph points out:

> The standard treatment of the history of non-European mathematics exhibited a deep-rooted historiographical bias in the selection and interpretation of facts, and that mathematical activity outside Europe has as a consequence been ignored, devalued or distorted.[4(p.3)]

The second approach, the compare-and-contrast method, considers regional variants, starting by studying specific geographical places, then comparing those mathematical traditions to those outside these cultural boundaries. Historical, comparative studies on different places were particularly useful in the 1980s and 1990s, when the notions of 'West' and 'East' increasingly were relevant, mainly because of ongoing geopolitical issues. Consequently, historiography followed this trend of seeing the world as West and East, with scholars often comparing the uniqueness of cultural spheres, again frequently focusing on the European tradition. As a result, studies revealed the 'divides' between the West and East.

The third approach focuses on the accumulation of knowledge, and often describes technical developments in math. Represented well in the history of mathematical constants – e.g. pi (π) or the proof for Fermat's Last Theorem – this approach examines a specific problem, method, formula or proof, and demonstrates the process of solving it.[5-7] However, this approach assumes that knowledge is an ever-expanding, ever-evolving entity. What if the assumption were wrong? What if the basic understanding that we take for granted were wrong? (Knowledge that had been taken for granted has been shaken to its core more than once in history, e.g. the Copernican Revolution.)

The fourth approach focuses on a specific academic institution and examines its internal development and dynamics. Analysing the educational system and its influence on mathematical activities, this approach highlights the importance of institutions among professional mathematicians.[8,9]

Finally, the fifth approach concerns biographies. It is a popular research theme to recount great mathematicians' life stories[10] and explore the various scholars associated with eminent mathematicians.

Global perspectives

As for the global history of math, which approach should one take? Because global history deals with a wider area, we need to combine some elements from the above and illustrate a detailed picture of the international exchange of knowledge that has not been done before. For example, my ongoing research examines: (1) the development of dimensions associated with mathematical problems; (2) an understanding of the universe based on the movement of stars; and (3) syncretic visions of mathematical logic that cultural interactions facilitate. Observing math in the context of the 17th century, I began to see in many places that the imagination triggered the expansion of existing mathematical knowledge worldwide.

Employing a bird's-eye view to survey broader regions, I have traced how ideas move across national borders. Cross-cultural perspectives, which often have helped blend works from both the humanities and sciences, are vital and effective. They also create a clear contrast against the study of material culture. Studies on 'things' and 'knowledge' are, by their nature, different. For example, in studying the global movements of material goods, Paula Findlen eloquently wrote about the concept of 'global':

> *The global can lie within objects, interwoven into the very fabric of a thing, shaped by craft, knowledge and materials that migrate over time from one place to another. The global is also about the misconception of objects, including efforts to simulate, but not entirely replicate, something observed elsewhere. Things – and the ability to make them – travel, but they do not remain the same.*[11(p.244)]

Can we say the same about mathematical knowledge? Let us examine an example that conveys the sense of global knowledge currents, which, essentially, have remained the same.

The example comes from a mathematical textbook published in 1627 in Kyoto, Japan. This book, *Treatise for the Ages* (*Jinkōki*), often has been called a primer[12], presenting mathematical basics, e.g. numerical tables, abacus techniques and arithmetic operations. Although not a good source for studying technical advancements in math, *Treatise for the Ages* cited questions from European and Chinese math textbooks, so this book is a unique historical artefact that shows how exchanges of ideas had been occurring in the 17th century:

Question: There is 1 *to* [a unit of volume, which is equivalent to 18 litres] of oil. Two people are going to have one half each. There are a 3 *shō* [one tenth of 1 *to*] and a 7 *shō* measuring cups available for them to divide 1 *to* into 5 *shō* each. Find a way to share the oil equally using these two measuring cups.

This problem was not in Chinese math textbooks, but Niccolo Fontana Tartaglia (1449–1557) presented it in his book, using red wine instead of oil. Reflecting Kyoto's commercial culture, which included many oil merchants, the original author chose oil as the relevant liquid, but the problem, method and answer worked just the same with wine.

It turned out that mathematical knowledge was brought to Japan via several different routes. Two major routes were used to make the long-distance voyage from Europe to Japan: one was the major trade route, from Europe south to the tip of South Africa, then east to India, South Asia and on to Japan, and the other entailed sailing west from Europe to the southern tip of Latin America and crossing the Pacific Ocean. Thus, this mathematical knowledge did not come straight from Europe to East Asia, but was carried by those who stopped at ports in South Africa, India and Southeast Asia. Without travelling through the southern hemisphere, sailors on either route were unable to transport books to Japan.

How and where can we see Europeans' impact on South Africa? Both cultures likely were unable to communicate well enough to discuss scientific knowledge because of differences in language or logic. For example, the numerical system is a good reference point on differences in mathematical thought. In South Africa, the principal counting system was said to be two or 20[13], as opposed to Europeans' 10-based (i.e. the decimal system) or 60-based (i.e. clocks' hours, minutes and seconds) systems. Thus, differences in their intellectual cultures must have been rather ubiquitous.

What about exchange rates? When foreign sailors arrived at the harbour of South Africa, would they not be gathering supplies to continue their journey to the East? Considering the history of slavery and colonisation, reciprocity cannot be expected, but basic trade activity must have occurred, e.g. negotiating arrangements for bartering or exchanging goods for currency, using units and numbers. Math, especially when used in everyday conversation, reflected the ways in which people coded certain information – a context in which people examined knowledge from foreign cultures in detail. The study of global travel, including maritime and colonial history, enriches the history of mathematics greatly.

When taken together, humanities and science reveal the blind spot in existing scholarship. If we had focused solely on the political, cultural and social aspects of global travel, we would miss the scientific aspects of interactions. If we focused only on mathematical techniques, we would not perceive the interactions of knowledge among different cultures. History and math must be discussed together to demonstrate how we see the dynamics of intellectual currents, pushing and pulling mathematical ideas across cultural and national boundaries.

My interdisciplinary project reflects the effect of globalisation on academic studies. As my research progressed, I encountered another problem concerning the timeline. The historical narrative usually shows a linear progression, most frequently marked by years, but the adaptation of multiple timelines often is necessary for a project on global history. For example, sailing from Amsterdam to Kyoto in the 17th century took at least 2 years. This simple fact makes it apparent that we must deal with multiple timelines when writing a global history.

Furthermore, several communication pathways exist between Europe and East Asia. Because the transmission of ideas took a few different routes, several conversations happened concurrently over time. While a history that progresses with time is simple and clear in its presentation, we sometimes need to break away from perceptions of historical time as being the one and only method for structuring a historical narrative. With a single timeline, we ignore time's complexity. Thus, my research on global history aims to reconsider the meaning of time.

What is next?

Overcoming ethno-mathematics is an urgent task. In the context of South Africa, the nation has a richly diverse population, and scholars of both history and math could offer more insights into this intellectual world. The study of math's global history is like a lightbulb: when turned on, it illuminates surrounding areas that had been hidden in the dark.

Acknowledgements

References

1. Katz VJ. A history of mathematics. 3rd edition. New York: Pearson; 2014.

2. Boyer CB, Merzbach UC. A history of mathematics. 3rd edition. Hoboken, NJ: John Wiley & Sons; 2011.

3. Stillwell J. Mathematics and its history. New York: Springer; 2010. https://doi.org/10.1007/978-1-4419-6053-5

4. Joseph GG. The crest of the peacock: Non-European roots of mathematics. 3rd ed. Princeton: Princeton University Press; 2011.

5. Beckmann P. A history of pi. New York: St. Martin's Press; 1971.

6. Singh S. Fermat's last theorem: The story of a riddle that confounded the world's greatest minds for 358 years. London: Harper Perennial; 2005.

7. Szpiro GG. Poincare's prize: The hundred-year quest to solve one of math's greatest puzzles. London: Plume; 2008.

8. Fauvel J, Flood R, Wilson R, editors. Oxford figures: 800 Years of the mathematical sciences. Oxford: Oxford University Press; 1999.

9. Nadis S, Shing-tung Y. A history in sum: 150 Years of mathematics at Harvard (1825-1975). Cambridge, MA: Harvard University Press; 2013. https://doi.org/10.4159/harvard.9780674726550

10. Stewart I. Significant figures: Lives and works of trailblazing mathematicians. London: Profile Books; 2017.

11. Findlen P. Afterword. In: Gerritsen A, Giorgio R, editors. The global lives of things: The material culture of connections in the early modern world. New York: Routledge; 2016.

12. Horiuchi A. Japanese mathematics in the Edo Period (1600-1868): A study of the works of Seki Takakazu (?-1708) and Takebe Katahiro (1664-1739). Translated by Wimmer-Zagier S. Basel: Birkhäuser; 2013.

13. Huylebrouck D. Mathematics in (central) Africa before colonization. In: Anthropologica et Præhistorica. 2006;117:135–162.

AUTHORS:
Willem Coetzer[1]
Michelle Hamer[2,3]

AFFILIATIONS:
[1]South African Institute for Aquatic Biodiversity, Grahamstown, South Africa

[2]South African National Biodiversity Institute, Pretoria, South Africa

[3]School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa

CORRESPONDENCE TO:
Willem Coetzer

EMAIL:
W.Coetzer@saiab.ac.za

Managing South African biodiversity research data: Meeting the challenges of rapidly developing information technology

New developments in the funding requirements of biodiversity science as well as rapidly developing information technology warrant a sharper focus on the way in which biodiversity data are managed. We propose that an opportunity presents itself to develop a specific set of informatics skills among a new class of data analysts in the biodiversity science community. Our consideration of capacity development specifically emphasises the need for conceptual rigour, compliance with technical data standards and the culture of data publication or data sharing.

There is a pressing need for data stewardship skills and positions in the South African biodiversity science community. We describe previous and current initiatives that may help to provide the context of, and develop skills and capacity for, effective management or stewardship of biodiversity research data. The overlapping competencies of data stewardship, data curation and data preservation include:

> processes and activities related to the organization and integration of data collected from various sources, annotation of the data, and publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse and preservation.[1]

The role of the data steward can be distilled into the fundamental principles of findability, accessibility, interoperability and reusability (the FAIR principles).[1]

From where do biodiversity data originate?

During the last 5 years, a significant component of funding for South African biodiversity science has been channelled through the Foundational Biodiversity Information Programme (FBIP).[2] The FBIP is funded by the South African Department of Science and Technology (DST) and administered by the National Research Foundation and the South African National Biodiversity Institute (SANBI). The FBIP recognises the importance of biodiversity, not only in the narrower sense of a particular discipline of scientific research (e.g. taxonomy, systematics or ecology), but also in the broader context of the relevance of biodiversity to society. Four large, collaborative FBIP projects have been funded. These projects focus on marine biodiversity (the Seakeys Project), the effect of habitat fragmentation on the faunal diversity of Eastern Cape forests, filling gaps in biodiversity information to support decisions about the exploitation of shale gas in the Karoo (the Biogaps Project), and camera trapping of mammals to assess the status of species and populations inside and outside protected areas (the Snapshot Safari Project). In 2016, 20 smaller FBIP projects were undertaken to investigate a variety of subjects, including bat monitoring in the Kruger National Park, bryozoan e-taxonomy, and a number of applied projects, e.g. the use of polychaetes as bait, and a survey of earthworms and their use in vermicomposting. The FBIP explicitly requires researchers to generate and submit research data characterised as species occurrences, species attributes or population abundance records, or develop tools or generate data that facilitate the identification of species, including through molecular techniques (e.g. taxonomic keys or DNA barcodes). Physical specimens may or may not be preserved in the execution of these research projects. Resultant occurrence records may be associated with high-quality still images, videos or sound recordings.

The community of natural history collections (more appropriately referred to as natural science collections, or NSC) naturally intersects with the community of biodiversity researchers funded by the FBIP. Recent developments among South African NSC museums, including increased funding, promise to improve the conditions, operations and utilisation of South African NSC (see below). Much has been written about the use of NSC or NSC data.[3,4] Such uses include estimating the spread of invasive alien or pest species; evaluating the abundance, conservation status and distribution of threatened species[5-7]; or projecting the ecosystem impacts of urban development, e.g. changes in ecosystem services such as pollination[8].

Properly and efficiently managing the biodiversity research data described above presents technical and organisational challenges arising from the rapid development of technology. Researchers' or technicians' data management skills do not always match the increasingly stringent requirements to organise and store data from the broad and diverse array of biodiversity projects conducted by the South African research community. These requirements are both technical and administrative (e.g. that data should be available for others to use). How can we improve biodiversity data management, integration and utilisation (e.g. how should students collaborate with their supervisors to share data, especially when they are not on the same campus)? Where should the data be stored and who should be responsible for data storage and long-term data preservation? Which data standards should be used? What conditions should be associated with using the data? Below we describe some of the challenges that the broader community of biodiversity scientists could face in developing greater capacity, specifically to manage and meaningfully use biodiversity occurrence data.

The need for conceptual rigour in curating NSC or biodiversity data

Collectively, records of physical specimens and records of observations of organisms are termed 'occurrence records' – hence we speak about 'the occurrence of a species at a place and time'. This phrase encapsulates all the fundamental classes of knowledge (i.e. metadata) about most of the biodiversity data referred to above. Occurrence records are particularly important to anchor abstract knowledge of species in the observed world. For example, an ecologist may need to assess the (occurrence of) freshwater invertebrate indicator species in a particular

stream to evaluate its current state, or compare the arthropod community structure of a forest with that of a nearby crop to assess the availability of natural enemies.

All such biodiversity occurrence records need to be curated in a specifically designed biodiversity database, even if representative voucher specimens are not preserved and deposited in a natural science museum for future reference. Occurrences of certain species may not be found, and these absences can be meaningful, e.g. when plants or marine invertebrates are systematically sampled using quadrats or photo-quadrats. Records of systematically structured sampling events and transects are therefore important to know that any effort, or how much comparable effort, was made to find occurrences. Species' absences increase the rigour of analyses such as ecological niche modelling, in which the species distribution range is estimated.

To comprehensively characterise the context of data, sampling events and occurrences must be represented using a coherent conceptual model. At a higher level of conceptual abstraction in this model, physical specimens and human observations are represented by the same properties of metadata classes. After all, a bird (occurrence) may have been seen during a sampling event at a particular place, whether or not the bird was captured or preserved as a specimen (Figure 1).
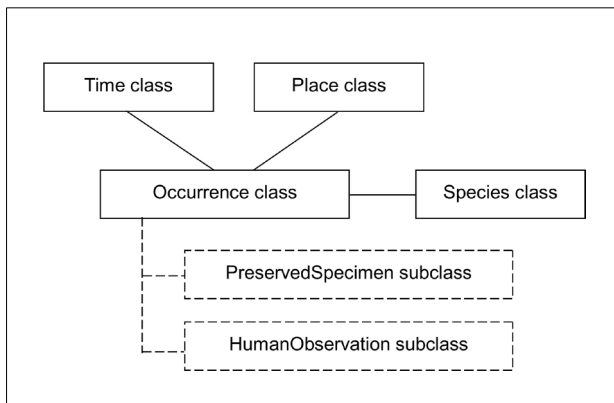
Figure 1: The properties of the occurrence class apply equally to the PreservedSpecimen subclass and the HumanObservation subclass, because the properties are related to the parent class: Occurrence.

There is thus a need to develop skills and capacity for generalised biodiversity data curation or stewardship, to integrate data records representing the full suite of concepts used by scientists, or to integrate typical NSC data with typical ecological data (i.e. to integrate specimen records with observations) for greater rigour or broader spatio-temporal coverage.

Below we elaborate on the basic idea, supported by the above reasoning, that a particular biodiversity database application (specifically its database schema), which is open-source software, is an ideal tool to use, both for physical specimens and biodiversity observations. In other words, it is an ideal database and application to manage biodiversity sampling event and occurrence records. Wider adoption of a common conceptual model, data management protocol, and approach will foster the development of a future class of biodiversity informatics technicians and analysts who will be able to efficiently manage and preserve our biodiversity research data.

Moving beyond traditional uses of NSC collection databases

The traditional specimen collection database is useful within the NSC museum, to document and manage a museum's specimen holdings by making inventories and keeping track of loans. More rigorous attention to the curation of biodiversity occurrence records will address other practical needs, e.g. the increasing requirements to include data

management plans in funding proposals and upload data sets to stable, online repositories.

Current biodiversity database applications include fields and functions which serve purposes other than collection management. For example, a globally unique identifier (GUID) ensures that a record can be uniquely identified, i.e. not confused with any other record published on the World Wide Web, which can be seen as the 'extended database' that is used to publish or share data. Such web technologies (of which GUIDs are just one example) are indicative of the changing culture of scientific data use typical of the Open Science Movement. These technologies imply that researchers ought to publish their biodiversity data in a way that makes the standardised (meta)data accessible to other researchers (i.e. researchers ought to use this extended database properly) (see Box 1).[9] The data steward therefore needs a thorough understanding of the conceptual model of the local database as well as that of the online repository or relevant data standard (see below).

The Open Science Movement[3,10] offers many diverse motivations to share scientific data, publications and knowledge, and mechanisms for conducting open scientific research. In South Africa, a new multi-institutional initiative, the Data Intensive Research Initiative of South Africa (DIRISA), is aligned with the principles of open science. The first National Research Data Workshop was held from 19 to 21 June 2018, and included presentations from astronomers, sustainable development researchers, bioinformaticians, biodiversity scientists and librarians, among others.[11] DIRISA is one of the three pillars of the National Integrated Cyber Infrastructure System (NICIS), an initiative of the DST that is implemented by the Council for Scientific and Industrial Research (CSIR). The other two pillars of the NICIS are the Centre for High Performance Computing (CHPC) and the South African National Research Network (SANReN).

Box 1: Data publication

Uploading occurrence data to the Global Biodiversity Information Facility (GBIF) Data Portal using the GBIF Integrated Publishing Toolkit is a common way to publish/share standardised biodiversity occurrence data, i.e. to make the data available for web integration with other data. Without access to more data of a high quality, we cannot expect to make progress in the more advanced uses of biodiversity data. Outdated opinions about data ownership, however, continue to cloud potentially progressive institutional data-sharing policies, and present a significant cultural barrier to the wider use of NSC data. The fact that scientists receive recognition (e.g. rating and publication subsidies) for publishing articles, but not for publishing data, has been addressed by the advent of the 'data paper', published by the *Biodiversity Data Journal*.[12] This journal accepts articles as long as they are accompanied by the underlying data in the form of links to data sets uploaded to online repositories. Such initiatives will allow authors' published data to be cited and recognised in the same way that research articles are recognised, and should therefore encourage researchers to publish biodiversity data. The general subject of scientific data archiving/sharing/publication has been addressed.[9,13-15]

The Global Biodiversity Information Facility (GBIF) currently publishes just over 1 billion standardised biodiversity occurrence records, from about 39 000 data sets contributed by about 1000 data providers around the world (Figure 2). Of these, 77% are records of human observations, and only 15% are records of physical specimens. Of the 19.2 million occurrences on the GBIF Data Portal originating from South Africa, only 1.8 million (9.4%) are preserved specimens. The GBIF data are freely available to be used in accordance with the terms of three Creative Commons (CC) licences. Many data providers will require attribution according to a supplied citation and will therefore publish their data under a CC-BY licence. Other data providers commit their data to the public domain and publish under a CC-Zero licence (not necessarily requiring acknowledgement or citation), or stipulate a CC-BY-NC licence, adding the requirement that use of the data will not be for commercial purposes. Occurrence records of southern African aquatic biodiversity are published on the GBIF Data Portal by the South African Institute for Aquatic Biodiversity (SAIAB) under a CC-BY licence.[16]
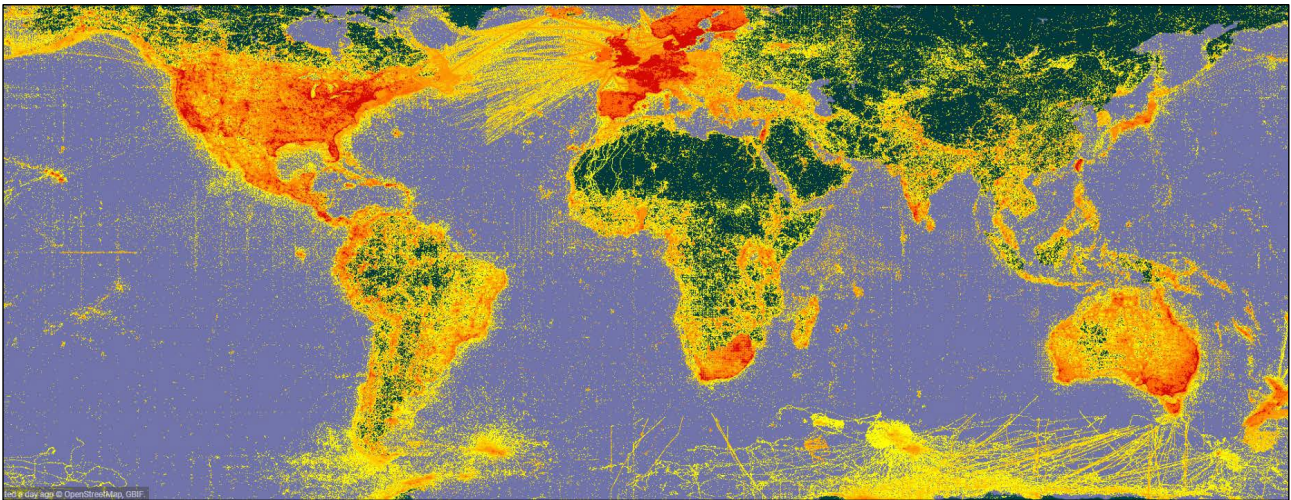
Source: GBIF (©OpenStreetMap contributors, ©OpenMapTiles)

Figure 2: A heat map published on the Global Biodiversity Information Facility (GBIF) website, showing the density of biodiversity occurrence records published by GBIF.

Improving biodiversity data curation in South African natural science collections

In 2012, the Museum Data Migration Project[18] was initiated by SAIAB to migrate the specimen records of selected museums to newly developed collection databases. Museum staff were then trained to use the databases to better manage specimen collections. Specify Software[19], which has been under development for about 30 years, was used to develop the databases. Specify Software is popular worldwide and is currently used by about 60 trained users working in 13 South African NSC museums (which house more than 50 specimen collections). From 2012 to 2015, another project, funded by the JRS Biodiversity Foundation, involved the cleaning and migration of significant data sets of arachnid and other data to new or existing collection databases, accompanied by further Specify Software training.

South African NSC have been periodically assessed since 1974.[20] Despite recognition of their importance, globally NSC have not fared well because of decreasing funding and the erosion of positions.[21] South Africa is a shining exception since the launch in October 2017 of the Natural Science Collections Facility (NSCF) – a much-anticipated response to concerns of neglected collections raised in recent years by the biodiversity research community. The NSCF is a virtual facility composed of a network of institutions that hold natural science collections which are accessible to external researchers. The overall aim of the NSCF is to ensure that natural science research collections and associated data are used for high-quality research and decision-making to address issues of socio-economic importance. The NSCF is funded as part of the DST's long-term funding programme, the Research Infrastructure Roadmap (SARIR), and administered by SANBI. A Coordinating Committee oversees operational management and is supported by several working groups made up of staff already employed by South African NSC museums. The Data Working Group includes representatives from various collection institutions who have experience in data management and strive to improve data curation and the use of appropriate data standards across institutions, to enable integration and publication of high-quality, standardised biodiversity data.

A new initiative, the Biodiversity Data Curation Platform (comparable to a cloud hosting service), initiated by SAIAB, will build on the Museum Data Migration Project by offering South African museums dedicated webservers and Specify 7 databases. Specify 7 is a web application that is the latest product released by the Specify Collections Consortium. It is hoped that the Biodiversity Data Curation Platform will ease museums' data management burden and contribute towards the objectives of the NSCF. Rather than requiring their own database server or systems administration expertise, staff of a participating NSC museum can gain access to a database on a virtual server, simply by loading a website using a standard web browser. Nothing else is required to make the museum's customised Specify 7 database and application available to perform routine collection management functions (e.g. catalogue specimens, query data, create loan records, and print loan invoices or specimen labels) or advanced informatics functions (e.g. export standardised data for publication on the GBIF Data Portal). In 2018, the vertebrate specimen records of four NSC museums, which have not previously used Specify Software for vertebrate specimens, were migrated to newly created databases hosted by the Biodiversity Data Curation Platform. Vertebrate specimens have been prioritised by the NSCF, both in terms of physical curation and data curation. These specimens and records are now in a better state to be examined by expert taxonomists, and brought to the requisite standard of preservation and information (e.g. specimens may need to be re-identified, and the taxonomy reflected in databases brought up-to-date). It is hoped that the Biodiversity Data Curation Platform will foster the development of biodiversity data curation expertise in South Africa's natural science museums.

Specify Software is not only for managing collections of physical specimens

An important SAIAB research project typifies the kind of biodiversity occurrence data in the South African community that need to be brought under formal data curation, namely the work on Baited Remote Underwater Video (or BRUV, another research platform offered by SAIAB)[22], and the closely related work on marine macrobenthos imagery[23]. The data will inform better decisions about the management of reef ecosystems and fish resources. In this underwater camera-trap and photo-quadrat sampling work, collection of physical specimens is not among the objectives. The BRUV videos and still images (Figure 3) of subtidal reef fish and
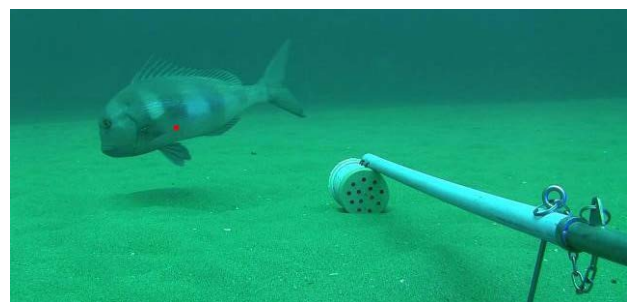
Figure 3: A still image from a video captured by a baited remote underwater video camera.

macrobenthos are associated with (meta)data that are used to assess fish assemblage structure, including species composition, abundance and size. The number of fish observed is recorded and the lengths of some of these are measured (if stereo cameras are used). Standard spatio-temporal metadata (place and time) as well as instrument settings are also recorded.

The data generated by this project are therefore typical human observations of biodiversity (marine fish; in the case of macrobenthos the data are typical photo-quadrat sampling events to estimate percentage cover, including species absences). The conceptual model and database schema underlying Specify Software was tested to evaluate whether any of the fields necessitated by the fish and macrobenthos data and metadata could be said to be excluded. It was found that all fields were easily accommodated by the database schema.

When Specify Software is re-used for biodiversity observation data, interaction with the data need not be limited to the use of the Specify Software interface, but can be achieved through a custom-developed user-form (Figure 4) specifically tailored to users' various requirements. In contrast, tailoring a database schema and input mechanism (by far the heavier infrastructure), or underlying conceptual model, to each biodiversity research project would be tantamount to re-inventing the wheel many times, and would complicate data integration.

We therefore argue that capacity development for the curation of biodiversity occurrence records, including many or most of the different biodiversity sampling protocols and objectives (i.e. not only traditional NSC objectives), can potentially be strengthened by the use of a common conceptual model (database schema) and related 'spoken language'. By re-using the Specify database schema we will be standardising the tools we use to carry out the same fundamental operations of information management across the community (e.g. data validation preceding batch data importation), which will make it easier for technicians to learn the techniques of biodiversity data curation.

The Biodiversity Data Curation Platform includes a tool to publish data, but the platform allows the NSC museum clients to execute information management functions independently and according to their own procedures and policies. The Integrated Publishing Toolkit (IPT) was developed by GBIF to simplify the process of publishing standardised biodiversity data on the GBIF Data Portal. At the national scale, fish and invertebrate sampling protocols may be differently designed and metadata classes differently defined from project to project, and this could complicate the storage, management, sharing, analysis and interpretation of data. Use of the Biodiversity Data Curation Platform could therefore be a first step to remedy this semantic heterogeneity, by allowing different users to manage their data independently but in a way that will allow the data to be vertically integrated (Figure 5) through the use of biodiversity information standards (specifically the Darwin Core metadata terms[24]).

When publishing biodiversity or occurrence data it is important to use ratified data standards. Compliance with these standards basically requires particular words to be used as field names (e.g. 'basisOfRecord') provided that a strict definition applies, as well as particular words to be used as the data values (e.g. 'HumanObservation') in these standardised fields. Biodiversity data standards are developed and published by the community of biodiversity informatics practitioners and researchers, through the organisation Biodiversity Information Standards (formerly the Taxonomic Databases Working Group).[25]
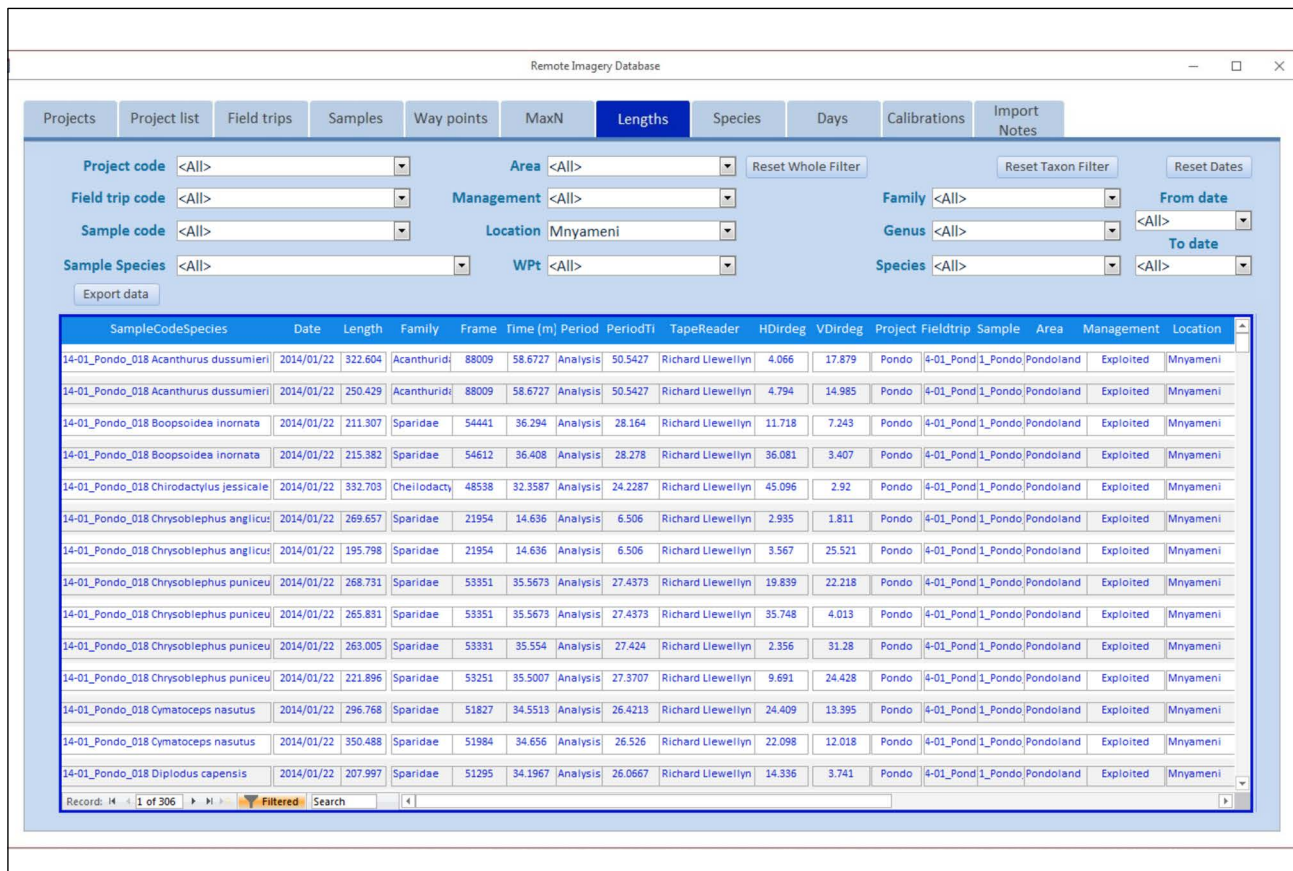
Figure 4: A simple user interface built in Microsoft Access, to filter and export the data from the back-end Specify (MySQL) database. This interface greatly enhances data accessibility within the South African Institute for Aquatic Biodiversity because it replaces hundreds of differently formatted spreadsheets.
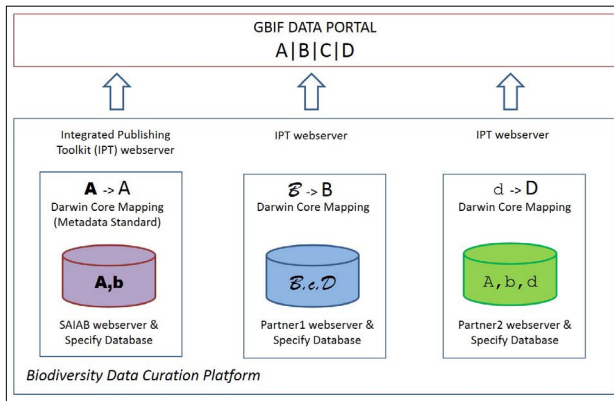
Figure 5: Whereas data from different sources are usually heterogeneous with respect to the meaning of terms/words, metadata standardisation can facilitate vertical integration of consistently defined terms.

Challenges and capacity development

We need to investigate ways to further develop technical skills to use Specify 7 technology effectively in NSC museums and biodiversity research institutes. Ensuring that the transition to Specify Software will be sustainable must be a high priority. It will be important to design a comprehensive training programme to improve data management, data curation and data publication skills in the NSC and biodiversity science community. Only then can we expect that the increasing use of information technology in NSC and associated institutes will become differentiated into new roles in these organisations. It is possibly this lack of differentiation that has held back the development of biodiversity informatics skills and professionals.

The process of 'cleaning' legacy NSC and biodiversity data, and migrating the data to a new, more rigorous database schema, is potentially a bottleneck to progress. Even when all the legacy data have been migrated to the new platform, a national-scale, sustained effort will be required to ensure that newly acquired data will continue to be imported into, and curated in, NSC databases consistently, timeously and accurately. The Specify suite of applications includes a 'Workbench' tool which can be used to map spreadsheet columns to database fields, for bulk data importation. The outcome of a given data importation routine will depend on the data steward's understanding of how the conceptual model represents knowledge concepts denoted by the (meta)data. This is therefore where the focus of capacity development for biodiversity data curation should be (i.e. rather than focusing simply on using the Specify application's interface correctly to catalogue individual records, which requires a much lower level of competence). The Specify Workbench is an important tool and level of technology development, around which concrete capacity development initiatives can be designed, to engage not only specialist data analysts but traditional NSC museum practitioners as well.

This new class of data stewards will be responsible for carefully channelling the flow of data into NSC or biodiversity databases and sharing the data for wider use. Enhancing data curation skills could contribute to the establishment of a new culture of data stewardship in NSC and biodiversity research institutes, in which South African biodiversity researchers and technicians can look forward to collaborating on exciting and creative projects to use new information technology and high-quality data in biodiversity science and ecological research.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3, Art. #160018, 9 pages. Available from: https://doi.org/10.1038/sdata.2016.18

2. Foundational Biodiversity Information Programme [homepage on the Internet]. No date [cited 2018 Jul 01]. Available from: http://fbip.co.za/

3. Schilthuizen M, Vairappan CS, Slade EM, Mann DJ, Miller JA. Specimens as primary data: Museums and "open science". Trends Ecol Evol. 2015;30:237–238. Available from: https://doi.org/10.1016/j.tree.2015.03.002

4. Coetzer W. Natural science collections [network on the Internet]. c2018 [cited 2018 Jul 01]. Available from: https://www.mendeley.com/community/natural-science-collections/

5. Ponder WF, Carter GA, Flemons P, Chapman RR. Evaluation of museum collection data for use in biodiversity assessment. Conserv Biol. 2001;15(3):648–657. Available from: http://doi.wiley.com/10.1046/j.1523-1739.2001.015003648.x

6. Drinkrow DR, Cherry MI, Siegfried WR. The role of natural history museums in preserving biodiversity in South Africa. S Afr J Sci. 1994;90(8–9):470–479.

7. Tweddle D, Bills R, Swartz E, Coetzer W, Da Costa L, Engelbrecht J, et al. The status and distribution of freshwater fishes. In: Darwall WRT, Smith KG, Tweddle D, Skelton P, editors. The status and distribution of freshwater biodiversity in southern Africa. Gland/Grahamstown: IUCN/South African Institute for Aquatic Biodiversity; 2009. p. 21–37.

8. Pauw A, Hawkins JA. Reconstruction of historical pollination rates reveals linked declines of pollinators and plants. Oikos. 2011;120(3):344–349. Available from: http://centaur.reading.ac.uk/20920/

9. Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE. Biodiversity data should be published, cited, and peer reviewed. Trends Ecol Evol. 2013;28(8):454–461. https://doi.org/10.1016/j.tree.2013.05.002

10. Lin T. Cracking open the scientific process. New York Times. 2012 January 16. Available from: https://www.nytimes.com/2012/01/17/science/open-science-challenges-journal-tradition-with-web-collaboration.html?_r=2

11. DIRISA. National Research Data Workshop [webpage on the Internet]. c2018 [cited 2018 Jul 07]. Available from: https://www.dirisa.ac.za/workshop/

12. Chavan V, Penev L. The data paper: A mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics. 2011;12(Suppl 15), S2, 12 pages. Available from: https://doi.org/10.1186/1471-2105-12-S15-S2

13. Koopman MM, De Jager K. Archiving South African digital research data: How ready are we? S Afr J Sci. 2016;112(7/8), Art. #2015–0316, 7 pages. https://doi.org/10.17159/sajs.2016/20150316

14. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: Practices and perceptions. PLoS ONE. 2011;6(6), e21101, 21 pages. https://doi.org/10.1371/journal.pone.0021101

15. Penev L, Mietchen D, Chavan V, Hagedorn G, Smith V, Shotton D, et al. Strategies and guidelines for scholarly publishing of biodiversity data. Res Ideas Outcomes. 2017;3, e12431, 12 pages. https://doi.org/10.3897/rio.3.e12431

16. Coetzer W. Occurrence records of southern African aquatic biodiversity [data set on the Internet]. c2010 [cited 2018 Jul 06]. Available from: https://www.gbif.org/dataset/1aaec653-c71c-4695-9b6e-0e26214dd817

17. Open Streetmap. Copyright and license [webpage on the Internet]. No date [cited 2019 Feb 13]. Available from: http://www.openstreetmap.org/copyright

18. Coetzer W, Gon O, Hamer M, Parker-Allie F. A new era for specimen databases and biodiversity information management in South Africa. Biodivers Informatics. 2012;8:1–11. https://doi.org/10.17161/bi.v8i1.4263%09%0A

19. Specify Software [software on the Internet]. No date [cited 2018 Feb 01]. Available from: http://www.specifysoftware.org

20. Coaton W. Status of the taxonomy of the Hexapoda of southern Africa. Entomol Mem Dep Agric Tech Serv Repub South Africa. 1974;38:1–124.

21. Kemp C. Museums: The endangered dead. Nature. 2015;518(7539):292–294. https://doi.org/10.1038/518292a

22. Bernard ATF, Götz A, Parker D, Heyns ER, Halse SJ, Riddin NA, et al. New possibilities for research on reef fish across the continental shelf of South Africa. S Afr J Sci. 2014;110(9/10), #a0079, 5 pages. http://dx.doi.org/10.1590/sajs.2014/a0079

23. Heyns ER, Bernard ATF, Richoux NB, Götz A. Depth-related distribution patterns of subtidal macrobenthos in a well-established marine protected area. Mar Biol. 2016;163(2):1–15.

24. Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE. 2012;7(1), e29715, 8 pages. Available from: https://doi.org/10.1371/journal.pone.0029715

25. Biodiversity Information Standards [homepage on the Internet]. No date [cited 2018 Feb 01]. Available from: http://www.tdwg.org/

AUTHORS:
Ian Glenn[1]
Sam M. Ferreira[2]
Danie Pienaar[2]

AFFILIATIONS:
[1]Research Fellow, Communication Sciences, University of the Free State, Bloemfontein, South Africa
[2]Scientific Services, SANParks, Skukuza, South Africa

CORRESPONDENCE TO:
Ian Glenn

EMAIL:
GlennIE@ufs.ac.za

Communication on rhino poaching: Precautionary lessons about backfires and boomerangs

Rhinoceros (rhino) poaching has become a dominant topic in wildlife management in southern Africa since the poacher onslaught started in 2008.[1] As social concerns about rhino poaching have grown, so have various forms of persuasive communication in attempts to curb, discourage, stigmatise and stop rhino poaching as well as discourage end users. Governmental and conservation agencies, non-governmental organisations (NGOs) dedicated to stopping rhino poaching, international organisations, documentary film-makers, television producers, and international figures like Prince William and celebrities like Jackie Chan, DJ Fresh and South African football star Maps Maponyane[2] have tried many different methods to persuade end users, local communities or intermediaries that killing rhinos for their horns is wrong.

How effective are such initiatives, particularly given that the demand for rhino horn shows little sign of abating and poached carcasses increased to over 1000 in South Africa by 2014? A reasonable response would be that such campaigns are usually ineffective particularly given that campaigns seldom consider outcomes such as changing or maintaining changed attitudes, or, even more difficult, changed behaviours.[3-6] The inducements driving rhino poaching[7] are so strong that it may be that no appeals are likely to deter offenders. Traditions of trade, inelasticity of demand and high profit potential may provide large enough incentives to override any persuasive appeals or threats[8], particularly when past social injustices may give an added sense of justification to poachers[7].

Here, however, we consider another possibility: that some communications about rhino poaching are counter-productive. To do this, we examine the literature on what has previously[9-12] been characterised as unintended effects or backfire or boomerang communications[13], in diverse areas such as health communications[9], political advertising[10], commercial advertising[11], or age warnings about unsuitable media content[12]. Here we look at how some of these lessons may apply to past communications that may help shape future communications about rhino poaching. These lessons may also apply to other areas of social concern where persuasive communication is crucial.

First, we give a brief summary of four theoretical frameworks of problematic or 'backfire' communications, considering the implications for communications around rhino conservation and conservation more generally. These four frameworks – reactance theory, descriptive norms and injunctive norms, mis-framing or cultural misreading, and 'loose lips' – may overlap, but they provide a useful checklist of communication traps and suggest better ideals for communication.

Reactance theory – wildlife officials and local communities

The theorist who has considered a directly negative reaction to persuasion most carefully in a series of articles and books is Jack Brehm.[14-17] Brehm's reactance theory argues that forceful arguments that limit our freedom to choose and behave as we like, or as we have in the past, may produce a counter-action, or resistance. This is particularly true if a powerful person is making the argument, in which case we are also likely to view the person negatively. Brehm argues that reactance may set in to threats by implication, so that what seems to be a simple statement in a familial or organisational setting may lead to reactions because of the larger power and relational dynamics.[15]

Brehm's work overlaps with what is known as 'forbidden fruit' dynamics.[18] The best-studied examples are those on the effects of age bans and censorship on computer games or films. Many studies have shown that attempted bans make films, television programmes, or computer games more appealing for many consumers.[12,18,19]

What does Brehm's work suggest for communication about rhino conservation, in particular for communications with communities around the Kruger National Park? First, threats of harsh consequences for poachers, particularly emanating from powerful outsiders, are likely to lead to resistance, particularly in communities in which many people may have strong historical reasons for feeling aggrieved by the Kruger National Park, or have memories of a racialised military struggle in the apartheid era.[20,21] If communications about rhino poaching are seen as restricting and limiting freedoms, they are likely to be resisted and be counter-productive.

There is another area where reactance theory may explain reactions. Furious moral denunciations of poachers or angry billboard slogans like 'We don't shoot rhinos, we shoot poachers', risk racialising the issue and alienating communities. When EFF leader Julius Malema accuses white people of caring more for rhinos than for black people, many black South Africans may share his view because of the vehemence of anti-poacher anger that may seem racially based.[22] Some of the protests at Kruger National Park entry gates and displays at funerals of poachers in villages abutting the Park also suggest that official attempts to demonise poachers may rebound and Hübschle points to ways in which poachers may enjoy a 'Robin Hood' status in their communities.[21]

Brehm's work suggests that communication about rhino poaching should come from non-threatening insiders, that it should stress positive values of opportunity and choice linked to the survival of the rhino and wildlife tourism. Such messages should suggest that, in opposition, the choice of rhino poaching is limiting and self-defeating.

Communications that downplay threats and powerful messages from authority figures or symbolic violence on social media are also likely to lead to less concern in scholarly analysis about what has been characterised as the 'rhinofication of security' or green violence.[23-25]

Descriptive norms and injunctive norms – media coverage of a crisis

One early concern for officials was that increased coverage of rhino poaching seemed perversely related to increased poaching – to the point that it almost felt as though the coverage was helping to drive poaching. Not surprisingly, senior staff at Kruger National Park in South Africa have come to see communication about rhino poaching as inherently problematic. How could this happen?

Studies of the health industry may help provide answers. In a major study of the effects of a billion dollar anti-drug campaign in the USA, researchers concluded with some surprise that the only significant statistical finding to emerge was that those exposed to the campaign were more likely to use drugs than those that had not been exposed.[26] Similar disconcerting results have been found in studies of campaigns to reduce cigarette smoking[27], drinking at colleges[28] and drug usage[26].

Why would this be? Cialdini[29], in looking at a famous advertisement decrying pollution in the USA, argues that the advertisement undermines itself by showing how commonplace littering and pollution are. The descriptive norm (many people pollute) overcomes the injunctive norm (you shouldn't pollute). When teenagers are told not to do drugs, or drink alcohol, the main effect may be to suggest, particularly to those who are behaving better than the norm described, that many people are doing drugs or drinking, and thus reduce the inhibition on such actions. Crucially, one study suggests that unless the descriptive norms are complemented by appropriate injunctive norms, boomerang effects are likely.[30]

In looking at news communications about rhino poaching, it is clear that many early news reports stressed the increasing numbers of rhino killed and the value of the horn. When it was also clear that relatively few poachers had been caught, the descriptive impression would have been to encourage rhino poaching, particularly given the publicity about the huge prices paid for rhino horn. Few of these news reports expressly condemned poaching or showed alternative positive role models. Concerns about the effects of ongoing coverage of losses of rhino and increasing poaching led to the South African Department of Environmental Affairs releasing figures on rhino poaching on only a quarterly basis.

Media reporting on rhino poaching may have been more responsibly undertaken and done more to discourage poachers had it pointed out the relatively low prices paid to local poachers, or shown an imprisoned poacher and the costs to a grieving family. So, while South African and international media may have assumed their readers would have reacted to reports of poaching negatively, it may be that inadvertent environmental media framing effects helped contribute to publicising and increasing poaching.[31]

When it comes to campaigns to discourage rhino poaching through persuasion, threats or warnings of dire consequences need to be carefully considered. Ideally, such campaigns should align descriptive and injunctive norms by, for example, showing positive role models in law enforcement as valued and heroic figures, or showing the benefits of not poaching. If there is to be a negative tactic, it should perhaps be linked to showing the consequences for poachers who have been caught, by letting them discuss their unhappy experience of imprisonment and the distance from their families.

While the exact number of deaths from poaching is unknown, analysing how devastating deaths, serious injuries, and imprisonment have been to particular communities could change descriptive norms. If communities see poachers and poaching leading to overall hardship for a community, it may be easier to discourage recruitment there. It may also be possible to contrast the hardship and paltry rewards of the 'foot soldiers' to the large, relatively risk-free rewards gained by others, thereby driving a wedge between the community and the recruits and the criminal networks that profit from them.[32]

Mis-readings and counter-readings – NGOs and messages about rhino horn

People of different genders, classes, ages, cultures and religious convictions may read and react to the same message very differently – a fact which impacts strongly on communication on rhino poaching. Many examples from health communication show how different or counter-readings can, for example, lead to young pregnant women taking up smoking because the threat that smoking would lead to them having smaller babies was perceived by them as a health benefit.[33]

In the case of rhino poaching, this is a problem particularly faced by NGOs in the way that they phrase their appeals for funds. Many appeals look at the dangers to the survival of rhinos as a species. A local NGO appeals for funds in the local mall with the message: 'Don't let the big 5 become the big 4!' A South African T-shirt has the image of a rhino merging with that of a dodo, warning of the possible fate of extinction. These warnings of extinction may seem to work to discourage poaching, but in many cases they have the opposite effect.

Shelley[32] points out that in Vietnamese gift culture – a major use of rhino horn – the very rarity of the rhino horn increases the value of the gift and makes it more desirable and valued. Empirical evidence on the value of animals put on the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) endangered lists reinforces the point: announcing that something is endangered or vulnerable to extinction increases its rarity value and thus makes it more likely that its value will increase and gives poachers an added inducement.[34] It may even be that some criminal syndicates see extinction as increasing the value of their stockpile.[35]

We thus have the tension between a message that may be effective in fundraising, yet has the unintended result of increasing the value of – and thus the risk to – the animal the fundraisers are trying to protect. What well-meaning environmentally conscious audiences perceive locally or internationally may be completely different from what end users perceive. In the rhino case, the message heard across the illegal supply chain may be significantly different from what communicators intended. It might be much better for rhino conservation, and much more discouraging for international speculators, to convey the message that there is a sustainable rhino population with little or no risk of extinction.

These kinds of moral panic attacks from NGOs can have unintended consequences more broadly in conservation areas. Media messages wrongly portraying widespread declines in Zimbabwean wildlife resulted in dis-investment in wildlife projects, risking the creation of a self-fulfilling prophecy.[31]

'Loose lips may sink ships' – official communication and links with the public

The old World War II phrase warning against unguarded talk is the obvious area for tactical focus in communications. Ill-timed information about the whereabouts of rhino, or new security measures, or even about success in capturing criminals, may give poachers crucial information that makes their task easier or allows them to anticipate and counter defences. During 2014, SANParks moved rhinos from poaching hotspots to areas of relative safety in the Kruger National Park. Local South African media attended the first captures and 126 media outputs resulted with the key message 'Kruger evacuating rhinos'. Poaching rates doubled in the following 4 weeks in those poaching hotspots. Foreign media attended the second captures and 80 media outputs resulted with the key message 'Integrated approaches to protect rhinos'. Poaching rates stayed the same in those areas and in the rest of the Park. Subsequent arrests confirmed that poachers in the first case decided to attack before all the rhinos were removed from the more vulnerable areas.

There are other examples in which technological hubris has contributed to authorities giving too much information to criminals, thus allowing them to counteract measures. One example from KwaZulu-Natal is where radio transmitters were inserted into rhino horns and local communities were shown the devices and warned that any poachers would be tracked

and punished. Soon, rhinos were shot, but tracking of the horns was not possible. When one poacher was caught with a roll of tin foil in his backpack, authorities realised that poachers had simply covered the poached horns in tin foil, thus disabling the tracking device.[36]

In other cases, the desire of fundraisers or local or international NGOs to have their efforts publicised or given media attention may make that very intervention less successful. For example, a group raising funding for, hypothetically, dogs to be dropped by helicopter or helium-powered drones, or rhinos wearing devices that record sound, will, in publicising the new intervention, give poachers time to consider how best to react.

In many cases, it will not be possible to stop fundraisers vaunting their particular contributions, but, in planning, care should be taken to ensure that any information released can be used strategically – perhaps by suggesting that the intervention may be more effective or more widespread than is the case, or by concealing complementary measures taken.

A similar problem arises when we look at successes in capturing rhino smuggling kingpins or in closing a particular smuggling route. The capture of a Mr Big may encourage a larger number of would-be Mr Bigs to try to take advantage. Celebrating the closing of one smuggling route may help smugglers find better ones. This is a common feature in crime networks – resilient crime networks become more robust once they recover from a disruption and then become significantly more difficult to destroy.[37]

This problem is exacerbated when we consider that in many cases tactical on the ground information or planning may be compromised by internal betrayal where officials collude with poachers or may themselves be poachers. The problem of internal collaboration with poachers raises complex questions on how to identify and counter possible internal help for poachers. This situation, however, may also provide opportunities for authorities to spread disruptive information.

In terms of dealing with internal communications, the problem of identifying and isolating criminal behaviour within relevant organisations raises complex issues of labour relations, ethics, morale building and team-building more generally.

Conclusion

Our attempt to provide a framework for an analysis of risks associated with communication about rhino poaching is largely retrospective and theoretical. It should, however, help identify problem areas for particular sectors and other areas of public concern. How can this framework help in formulating messages going forward for various role-players?

Authorities could do well to define priorities for their strategic communications and focus on communication with stakeholders that has a high impact on achieving objectives of rhino conservation. Authorities may even gain further by recognising the existing nature of their relationship with priority stakeholders and vary the tone of messages accordingly to increase awareness, change attitudes or change behaviour. This awareness could help shape press releases, engagements with non-government organisations and journalists, and any other kind of communication about rhino poaching.

News media should recognise that there is no such thing as simply reporting on rhino poaching and discouraging it by looking at the number of rhinos killed, the value of the horns, and the number of poachers captured or killed. The dangers of failure or the costs of success for communities were notably lacking in most early accounts. Authorities should do more to shape the larger narrative.

Many NGOs have helped raise awareness and concern about poaching, but they too need to consider the costs of raising moral panics and the dangers of how their messages will be understood outside of an original audience of potential funders. Nor should they imagine that only sympathetic figures are reading or hearing their messages and plans.

Finally, authorities have to consider the costs and benefits of transparency and the tactical use of when and how to deploy information about their intentions and successes.

While rhino poaching poses its own particular 'wicked' problem for effective communication, the relevance of these pitfalls is clear for communicators and journalists more widely. Effective persuasion demands expertise and sociological and anthropological sophistication.

References

1. Knight M. African rhino specialist group report. Pachyderm. 2013(53):7–24.

2. Celebs band together for rhinos. The Citizen. 2016 May 05 [cited 2017 Feb 20]. Available from: http://citizen.co.za/news/news-national/1101441/speak-up-to-stop-rhino-poaching/

3. Bandura A. Social foundations of thought and action. Englewood Cliffs, NJ: Prentice Hall; 1986.

4. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991;50(2):179–211. https://doi.org/10.1016/0749-5978(91)90020-T

5. Prochaska JO. Transtheoretical model of behavior change. In: Gellman M, Turner JR, editors. Encyclopedia of behavioral medicine: New York: Springer-Verlag; 2013. p. 1997–2000.

6. Zanna MP, Rempel JK. Attitudes: A new look at an old concept. In: Bar-Tal D, Kruglanski AW, editors. The social psychology of knowledge. Cambridge, UK: Cambridge University Press; 1988. p. 315–334.

7. Conrad K. Trade bans: A perfect storm for poaching. Tropical Conserv Sci. 2012;5(3):245–254. https://doi.org/10.1177/194008291200500302

8. Ferreira SM, Pfab M, Knight M. Management strategies to curb rhino poaching: Alternative options using a cost-benefit approach. S Afr J Sci. 2014;110(5–6), Art. #2012-0055, 8 pages. https://doi.org/10.1590/sajs.2014/20120055

9. Crano WD, Burgoon M. Mass media and drug prevention: Classic and contemporary theories and research. Mahwah, NJ: Lawrence Erlbaum Associates; 2002.

10. King JD, McConnell JB. The effect of negative campaign advertising on vote choice: The mediating influence of gender. Soc Sci Q. 2003;84(4):843–857. https://doi.org/10.1046/j.0038-4941.2003.08404012.x

11. Petrova PK, Cialdini RB. Fluency of consumption imagery and the backfire effects of imagery appeals. J Consumer Res. 2005;32(3):442–452. https://doi.org/10.1086/497556

12. Bushman BJ, Cantor J. Media ratings for violence and sex: Implications for policymakers and parents. Am Psychologist. 2003;58(2):130–141. https://doi.org/10.1037/0003-066X.58.2.130

13. Byrne S, Hart PS. The boomerang effect: A synthesis of findings and a preliminary theoretical framework. Ann Int Commun Assoc. 2009;33(1):3–37. https://doi.org/10.1080/23808985.2009.11679083

14. Brehm JW. A theory of psychological reactance. New York: Academic Press; 1966.

15. Brehm JW. Psychological reactance: Theory and applications. In: Srull TK, editor. NA – Advances in Consumer Research. Volume 16. Provo, UT: Association for Consumer Research; 1989. p. 72–75.

16. Brehm SS, Brehm JW. Psychological reactance: A theory of freedom and control. New York: Academic Press; 2013.

17. Miron AM, Brehm JW. Reactance theory – 40 years later. Zeitschrift für Sozialpsychologie. 2006;37(1):9–18. https://doi.org/10.1024/0044-3514.37.1.9

18. Bushman BJ, Stack AD. Forbidden fruit versus tainted fruit: Effects of warning labels on attraction to television violence. J Exp Psychol Appl. 1996;2(3):207–226. https://doi.org/10.1037/1076-898X.2.3.207

19. Bijvank MN, Konijn EA, Bushman BJ, Roelofsma PH. Age and violent-content labels make video games forbidden fruits for youth. Pediatrics. 2009;123(3):870–876. https://doi.org/10.1542/peds.2008-0601

20. Van Heerden P. Faces of Africa – Charlie 'Nyempere' Nkuna: Protector of Kruger National Park [video on the Internet]. 2018 March 13 [cited 2018 Sep 13]. Available from: https://www.youtube.com/watch?v=CkM_DRqoqfA

21. Hübschle AM. The social economy of rhino poaching: Of economic freedom fighters, professional hunters and marginalized local people. Curr Sociol. 2017;65(3):427–447. https://doi.org/10.1177/0011392116673210

22. Malema J. Why do white people despise blacks? Times Live. 2016 January 10. Available from: https://www.timeslive.co.za/sunday-times/opinion-and-analysis/2016-01-10-why-do-white-people-despise-blacks/

23. Lunstrum E. Green militarization: Anti-poaching efforts and the spatial contours of Kruger National Park. Ann Assoc Am Geogr. 2014;104(4):816–832. https://doi.org/10.1080/00045608.2014.912545

24. Massé F, Lunstrum E. Accumulation by securitization: Commercial poaching, neoliberal conservation, and the creation of new wildlife frontiers. Geoforum. 2016;69:227–237. https://doi.org/10.1016/j.geoforum.2015.03.005

25. Humphreys J, Smith M. The 'rhinofication' of South African security. Int Affairs. 2014;90(4):795–818. https://doi.org/10.1111/1468-2346.12141

26. Hornik R, Jacobsohn L, Orwin R, Piesse A, Kalton G. Effects of the national youth anti-drug media campaign on youths. Am J Public Health. 2008;98(12):2229–2236. https://doi.org/10.2105/AJPH.2007.125849

27. Wakefield M, Terry-McElrath Y, Emery S, Saffer H, Chaloupka FJ, Szczypka G, et al. Effect of televised, tobacco company-funded smoking prevention advertising on youth smoking-related beliefs, intentions, and behavior. Am J Public Health. 2006;96(12):2154–2160. https://doi.org/10.2105/AJPH.2005.083352

28. Wechsler H, Nelson TE, Lee JE, Seibring M, Lewis C, Keeling RP. Perception and reality: A national evaluation of social norms marketing interventions to reduce college students' heavy alcohol use. J Stud Alcohol. 2003;64(4):484–494. https://doi.org/10.15288/jsa.2003.64.484

29. Cialdini RB. Crafting normative messages to protect the environment. Curr Directions Psychol Sci. 2003;12(4):105–109. https://doi.org/10.1111/1467-8721.01242

30. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychol Sci. 2007;18(5):429–434. https://doi.org/10.1111/j.1467-9280.2007.01917.x

31. Gandiwa E, Sprangers S, Van Bommel S, Heitkönig IM, Leeuwis C, Prins HH. Spill-over effect in media framing: Representations of wildlife conservation in Zimbabwean and international media, 1989–2010. J Nature Conserv. 2014;22(5):413–423. https://doi.org/10.1016/j.jnc.2014.03.009

32. Shelley L. Convergence of the rhino horn trade with other criminal activity [document on the Internet]. c2016 [cited 2018 Sep 13]. Available from: http://defsec.csir.co.za/wp-content/uploads/2016/05/1.-Dr-Louise-Shelley-Convergance-of-Illicit-Actors-and-Flows.pdf

33. Dennis S. Golden chocolate olive tobacco packaging meets the smoker you thought you knew: The rational agent and new cigarette packaging legislation in Australia. Contemp Drug Problems. 2013;40:71–97. https://doi.org/10.1177/009145091304000105

34. Sand PH. Enforcing CITES: The rise and fall of trade sanctions. Rev Eur Comp Int Environ Law. 2013;22(3):251–263. https://doi.org/10.1111/reel.12037

35. Bulte EH, Mason CF, Horan RD. Betting on extinction: Endangered species and speculation. Land Econ. 2003;79(4):460–471. https://doi.org/10.2307/3147293

36. Lottering B. Interview with Ian Glenn 2015.

37. Ayling J. What sustains wildlife crime? Rhino horn trading and the resilience of criminal networks. J Int Wildlife Law Policy. 2013;16(1):57–80. https://doi.org/10.1080/13880292.2013.764776

AUTHOR:
Hlengisizwe Ncube[1] iD

AFFILIATION:
[1]Department of Forest Resources and Wildlife Management, National University of Science and Technology, Bulawayo, Zimbabwe

CORRESPONDENCE TO:
Hlengisizwe Ncube

EMAIL:
hlengisizwe.ncube@nust.ac.zw

A call to embrace adaptive management for effective elephant conservation in Zimbabwe

Wildlife conservation is at a critical juncture across Africa, hamstrung by bureaucratic incompetence and the erosion of ethical principles, while wildlife populations are predominantly threatened by habitat fragmentation and indiscriminate killings.[1,2] The Zimbabwe Parks and Wildlife Management Authority (ZimParks) was once quintessential in Africa, among pioneers of the wildlife management front, inter alia, with effective protected area management, even authorising wildlife conservancies outside protected areas.[1] ZimParks is expected to generate its own income from both non-consumptive and consumptive activities, such as ecotourism and sport hunting.[1] However, a perennially lean budget, use of obsolete equipment, low morale among the staff, and a high staff turnover because of a low skills retention capacity constrain the activities of the department. Although ZimParks patrol teams are augmented by state police in major protected areas, rangers are sometimes injured or killed during contact with armed poachers with sophisticated weapons. The situation is continually made more dire by fraudulent tendencies, where ZimParks officers, state police and politicians are incessantly implicated as accomplices in wildlife poaching syndicates. On the other hand, the techniques used by poachers are dynamic, with recent elephant *Loxodonta africana* poaching tactics involving lacing water sources and salt licks with cyanide poison, which also kills secondary targets such as scavenging vulture species.[3]

There is neither a 'silver bullet' nor a 'straight jacket' solution to indiscriminate wildlife killings. While consolidated initiatives are necessary, there seems to be a tragedy of policy inconsistency and duplication of efforts in wildlife conservation. First, there is a misguided tendency to draw contest between the old 'tried and tested' methods against the new innovative conservation initiatives. The old conservation paradigm is constructed on a fortress model, where a largely unfenced protected area is the epicentre of conservation activity.[1] The new conservation model is pragmatic, with emphasis on the management of protected areas based on the ecological requirements of wildlife and the socio-economic aspirations of surrounding communities.[1,4] Second, the emergence of conflicting spheres of influence is now characteristic of wildlife conservation, where non-governmental organisations (NGOs) safeguard their institutional niche by criticising ZimParks initiatives, fortified by unscrupulous media coverage. Such NGOs negate the ethos of their very existence, by duplicating state organs and assuming competitor roles, rather than complementing ZimParks. Such an approach ensures that NGOs hoodwink donors, while in reality goodwill funds are eroded by overhead expenses and sustaining lavish lifestyles for senior management, far away from conservation areas. Nonetheless, it is imperative to acknowledge that some NGOs promote wildlife conservation and capacity building among Africans, serving as vital conduits for skills and technology transfer between Africa and the rest of the world. The Kavango–Zambezi Transfrontier Conservation Area (KAZA TFCA) biosphere, encompassing IUCN conservation area Category II–VII, is therefore ideal for examining the human–elephant coexistence paradigm.

The conservation status of the African elephant, a flagship species

The African elephant has coexisted with humans, in fluxes, driven by both ecological and socio-economic cruxes.[5] Historically, elephants occupied niches most suitable for human agriculture, and now access approximately 5 million km^2, 10–25% of their historical range in southern Africa.[5,6] The KAZA TFCA, 440 000 km^2, comprising parts of Angola, Botswana, Namibia, Zambia and Zimbabwe, is a multinational initiative for promoting human–elephant coexistence.[6] Approximately 60% of the 230 000 elephants within KAZA TFCA share space with humans.[6,7] Large herbivore populations across Africa have experienced population declines in the past four decades, even in protected areas.[8] However, the elephant has been increasing, with fluctuations, at 1.5–5% of the annual population growth rates, within the KAZA TFCA biosphere, mainly in response to the suspension of culling in 1986 and the regulation of international trade.[6] In 1989, the African elephant was classified under the Convention on International Trade on Endangered Species of Fauna and Flora (CITES) Appendix I, and reclassified in 1997 to CITES Appendix II, only in Botswana, South Africa, Zambia and Zimbabwe, where international trade is regulated through a quota system, and a moratorium on ivory stock pile auctions.[6,9]

In protected areas, such as Hwange National Park, covering 14 651 km^2 and located between 18°30'–19°50'S and 25°45'–27°30'E, the elephant is a *de facto* keystone species, sometimes reaching densities of 4 elephants/km^2 and constituting over 90% of mammal biomass during the dry season.[10] In open ecological systems, where elephant population densities exceed 0.5 elephants/km^2, the 'elephant problem' has been significant, and is evidenced by stripped bark; broken, pushed over and uprooted trees; and the temporal niche shifts of other large herbivores.[10,11] Maintaining the carrying capacity, the dogma of wildlife management, is therefore reactionary and a symptomatic approach to wildlife conservation.[4] Surface water provision is a contentious issue, and is now considered outdated and at the root of the 'elephant problem'.[4,10] In the short to medium term, manipulating the distribution of key resources, such as dry season water provision, could ease pressure on vegetation, without causing mortality of other water-dependent wildlife species. Dispersal opportunities for elephants could also reduce the impact on vegetation around artificial watering points within protected areas.[4] Elephant culling for the protection of vestigial pockets of floristically impoverished reserves is not recommended, considering the relatively high rate of population declines elsewhere in Africa.[2,4,8] Culling is not only objectionable, it also stimulates an increase in the population growth rate, through the elimination of the effects of density-dependent factors.[12] The use of porcine zona pellucida (pZP) vaccines as contraceptives in elephants has been considered in South Africa.[13] However, the method is invasive and requires subsequent boosts, which is expensive and impractical for wild populations.[13] The growing elephant herd should therefore be recognised as an ecological entity – a resource to be conserved and harnessed, or 'something to cherish and to use'[14].

Outside protected areas, elephants compete for space with the human requirements for agricultural land, and there are no prospects for the restoration of prehistoric population levels.[4,15,16] An estimated 55 elephants are illegally killed every day across Africa to meet the demand for ivory, which is used to manufacture artifacts and ornaments.[15] Poaching pressure has been intense in East and Central Africa, and the poaching frontline has been slowly pushed southwards, such that Zambia and Zimbabwe began experiencing poaching pressure after the iconic Kenyan and Tanzanian elephant populations had been decimated.[15] The Sebungwe and Mid-Zambezi valley elephant population clusters are under siege from poachers, with Chizarira National Park experiencing a 75% elephant population decline during the 2001–2014 period.[3] Elephant populations are vulnerable, as is evidenced by the Kenya and Selous Game Reserve (Tanzania) herds that plummeted from 275 000 to 20 000 from 1973 to 1989, and from 109 000 to 13 000 from 1976 to 2013, respectively.[8] The CITES Monitoring Illegal Killings of Elephants (MIKE) and Elephant Trade Information System (ETIS) programmes remain informative, but not preventive.[9] Outside protected areas, sport hunting generates funds for ranchers and communal areas, through the Communal Areas Management Programme for Indigenous Resources (CAMPFIRE).[17] CAMPFIRE is a concept for devolution of wildlife conservation, where proceeds from wildlife products are used mainly for community infrastructure development.[17,18] However, illegal trade in wildlife products remains key in sustaining rural livelihoods among communities in close proximity of protected areas.[17,18]

What next? How to mitigate widespread wildlife population declines

Under the backdrop of such unprecedented rates of species depletion, the pertinent question remains poignant – what next? The magnitude of illegal elephant poaching calls for adaptive wildlife management, the re-thinking of wildlife conservation strategies, which responds to previous experiences.[4] Recently, ZimParks authorities have been granting concessions for protected area management to NGOs and eco-tourism establishments. The motive in granting such permits is mainly the lure of the windfall concession fees, misconstrued as 'free money'. Such a model of conservation has no guaranteed success, but presents secondary challenges, such as how to monitor concessionaire activities without seemingly interfering. The future of natural resources is therefore made dissolute, even looted and mortgaged to predominantly foreign influences, with unverified management concepts. There is an urgent need to optimise wildlife conservation, especially in human-dominated landscapes, through promoting human–wildlife coexistence.[1,19] It is irresponsible to entertain any aphorism that indigenous people are ignorant of wildlife conservation practices; they are an integral constituent of a composite human–wildlife biosphere. Active involvement of local communities in mainstream conservation could harness indigenous knowledge systems for effective wildlife conservation.[19] Local community involvement could be extended to quota setting of employment opportunities in mainstream conservation. Community scouts could be promoted, and groomed into professional rangers. A healthy relationship among the ecological and the socio-economic systems is the primary basis for adaptive wildlife management.[1,17] Legalised sport hunting, and its extreme form 'canned hunting', are increasingly becoming a contentious issue.[20] Arguments against the practice encourage non-consumptive use of wildlife resources, such as eco-tourism, as an alternative to cruel 'bloodsports'.[20] However, proponents of preserving sport hunting argue that funds realised are in turn used to reinforce wildlife conservation.[20]

At community level, poaching is poverty driven, especially where formal employment opportunities are limited. At a higher level, poaching is a high capital undertaking by international syndicates, and often finances political instability and terrorism, and thrives on corruption.[21] It is also a colossal task to attempt to change the socio-cultural consumption patterns for wildlife products at the main market, the Asian nations. Poaching is therefore more than a mere issue of enforcement, as is evidenced by sustained poaching rates, even when punitive fines and sentences are imposed.[21] The 'green military' approach, as practised in

Kenya, revokes a philosophical debate skirting on how much humanity is prepared to sacrifice to save wildlife. In Zimbabwe, public sentiment remains polarised on the endorsement of the shoot-to-kill policy with regard to wildlife poachers. Messer et al.[22] assert that the shoot-to-kill policy is the only viable solution in poor nations, where substantial economic gains are realised from wildlife poaching. However, there is a moral obligation for preserving basic human rights under International Law, where there are no provisions for extrajudiciary killings, except in self-defence.[22,23] It can also be argued that, at its best, the shoot-to-kill policy risks obstruction of justice, by killing suspects, while at its worst, it increases the wager for poaching, as it speculates a risk.[23] In most conservation areas, wildlife benefits accrue at community level, while human–wildlife conflicts are experienced at household level, seldom with direct compensation.[1,17] The shoot-to-kill policy therefore stands to further alienate protected area management from local communities.

It is envisaged that a successful re-alignment of the human–wildlife coexistence conservation paradigm can be attained through the establishment of Community Share Trusts, funded through levies on local ecotourism, sport hunting and the extractive industries, such as mining. Such devolution could finance compensation schemes and capital development projects such as value addition to non-timber forest products, livestock and crop insurance, and mitigatory initiatives such as the 'fencing people in, fencing elephants out' concept.[1,4,17] Zonation of land could delineate conservation networks, catering for conservation even where humans dominate, without detracting from sustainable livelihoods.[24]

Acknowledgements

References

1. Child G, Child B. The conservation movement in Zimbabwe: An early experiment in devolved community based regulation. Afr J Wildl Res 2015;45:1–16. https://doi.org/10.3957/056.045.0103

2. Chase MJ, Schlossberg S, Griffi CR, Bouche PJC, Djene SW, Elkan PW, et al. Continent-wide survey reveals massive decline in African savanna elephants. Peer J. 2016;4, e2354, 24 pages. https://doi.org/10.7717/peerj.2354

3. Muboko N, Muposhi V, Tarakini T, Gandiwa E, Vengesayi S, Makuwe E. Cyanide poisoning and African elephant mortality in Hwange National Park, Zimbabwe: A preliminary assessment. Pachyderm. 2014;55:92–94.

4. Van Aarde RJ. Elephants: A way forward. Pretoria: University of Pretoria; 2013.

5. Van Aarde RJ, Jackson TP. Megaparks for metapopulations: Addressing the causes of locally high elephant numbers in southern Africa. Biol Conserv. 2007;134:289–297. https://doi.org/10.1016/j.biocon.2006.08.027

6. Blanc JJ, Barnes RFW, Craig GC, Dublin HT, Thouless CR, Douglas-Hamilton I, et al. African elephant status report 2007: An update from the African elephant database. Occasional paper series of the IUCN species survival commission, No. 33. IUCN/SSC African elephant specialist group. Gland: IUCN; 2007.

7. Wittemyer G, Daballen D, Douglas-Hamilton I. Comparative demography of an at risk African elephant population. PLoS ONE. 2013;8, e53726, 10 pages. https://doi.org/10.1371/journal.pone.0053726

8. Craigie ID, Baillie JEM, Balmford A, Carbone C, Collen B, Green RE, et al. Large mammal population declines in Africa's protected areas. Biol Conserv. 2010;143:2221–2228. https://doi.org/10.1016/j.biocon.2010.06.007

9. Convention on International Trade on Endangered Species of Fauna and Flora [webpage on the Internet]. No date [cited 2018 Jul 16]. Available from: https://www.cites.org

10. Valeix M, Fritz H, Dubois D, Kanengoni K, Alleaume S, Said S. Vegetation structure and ungulate abundance over a period of increasing elephant abundance in Hwange National Park, Zimbabwe. J Tropic Ecol. 2007;23:87–93. https://doi.org/10.1017/S0266467406003609

11. Cumming DHM, Feneton MB, Rautenbach IL, Taylor RD, Cumming GS, Cumming MS, et al. Elephants, woodlands and biodiversity in miombo woodland in southern Africa. S Afr J Sci. 1997;93:231–236.

12. Cumming DHM. The decision-making framework with regard to the culling of large mammals in Zimbabwe. In: Owen-Smith NR, editor. Management of large mammals in African conservation areas. Pretoria: Haum Educational Publishers; 1983.

13. Asa CS, Parton IJ. Wildlife contraceptives: Issues, methods and application. Baltimore, MD: The Johns Hopkins University Press; 2005.

14. Mundy PJ. The African elephant – something to cherish and to use. Int J Environ Stud. 2006;63:587–597. https://doi.org/10.1080/00207230600963171

15. Shaw M, Rademeyer J. A flawed war: Rethinking 'green militarisation' in the Kruger National Park. Politikon. 2016;43(2):173–192. https://doi.org/10.1080/02589346.2016.1201379

16. Douglas-Hamilton I, Krink T, Vollrath F. Movements and corridors of African elephants in relation to protected areas. Naturwissenschaften. 2005;92:158–163. https://doi.org/10.1007/s00114-004-0606-9

17. Muphree MW. Congruent objectives, competing interests, and strategic compromise concepts and process in the evolution of Zimbabwe CAMPFIRE programme, 1984-1996. In: Brosius JP, Tsing LA, Zerner C, editors. Communities and conservation histories and policies of community based natural resources management. Walnut Creek, CA: AltraMira Press; 2005.

18. Frost P, Bond I. The campfire programme in Zimbabwe: Payment for wildlife services. Ecol Econ. 2008;64:776–787. https://doi.org/10.1016/j.ecolecon.2007.09.018

19. Pandey DN. Cultural resources for conservation science. Conserv Biol. 2003;17:633–635. https://doi.org/10.1046/j.1523-1739.2003.01578.x

20. Lindsey PA, Frank LG, Alexander R, Mathieson A, Romanach S. Trophy hunting and conservation in Africa: Problems and one potential solution. Conserv Biol. 2007;21:880–883. https://doi.org/10.1111/j.1523-1739.2006.00594.x

21. Challender DWS, MacMillan DC. Poaching is more than an enforcement problem. Conserv Lett. 2014;7:484–494. https://doi.org/10.1111/conl.12082

22. Messer KD. Protecting endangered species: When are shoot-on-sight policies the only viable option to stop poaching? Ecol Econ. 2010;69(12):2334–2340. https://doi.org/10.1016/j.ecolecon.2010.06.017

23. Mogomotsi GEJ, Madigele PK. Live by the gun, die by the gun: An analysis of Botswana's 'shoot-to-kill' policy as an anti-poaching strategy. S Afr Crime Q. 2017;60(June):51–59. https://doi.org/10.17159/2413-3108/2017/v0n60a1787

24. Roever CL, Van Aarde RJ, Leggett K. Functional connectivity within conservation networks: Delineating corridors for African elephants. Biol Conserv. 2013;157:128–135. https://doi.org/10.1016/j.biocon.2012.06.025

AUTHOR:
David A. Coldwell[1]

AFFILIATION:
[1]School of Economic and Business Sciences, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
David Coldwell

EMAIL:
david.coldwell@wits.ac.za

A fair individualised university researcher rating system? A rejoinder to the current NRF debate

Given the impact of individualised research rating classifications on academic careers, further effort needs to be dedicated to finding systems and methods aimed at ensuring that rating evaluations are as fair as possible. However, the idea that this aim is possible through the application of a *more scientific approach* in the classification of individual academic researchers' 'pecking orders'[1] is questionable.

Boshoff's[2] 'rebuttal' points to some of the inaccuracies of Callaghan's[1] description of South Africa's National Research Foundation's (NRF) researcher evaluation process, but, in my view, he places too much faith in the *fairness* of the NRF's current researcher evaluation system and at times seems to rather uncritically ingratiate himself to the existing system despite its obvious faults. An issue that Boshoff[2] conflates, which is central to his argument against Callaghan's[1], is that of discrimination between different levels of academic performance. Boshoff[2] finds it unacceptable that Callaghan[1] regards discrimination between academics as elitist and unfair. However, the real issue is surely not whether it is unfair to discriminate between academic performance as such, but whether the NRF's discrimination in evaluating the excellence of researchers' outputs and the grade allotted to individual researchers is fair *sui generis*. I argue that it is not fair because it does not take account of the life circumstances of specific researchers which undermine their equality of opportunity to scarce research resources controlled by the NRF in its capacity as the sole official social institution in South Africa for this purpose.

Callaghan[1] suggests that the NRF rating system is not sufficiently scientific and that it makes, at least occasional, errors in rating categories allotted to individual scientists. And, in some cases, it is suggested these errors are gross and could be avoided by being more scientific. I argue that the realistic achievement of a formal scientific rating system is unattainable. However, with greater attention to qualitative aspects of the evaluative process, in particular, rather than greater objectivity, many of these gross errors of individual researcher prowess classification could be avoided. Of course in this approach, the assumption is that an individual form of researcher evaluation such as that used by the NRF in South Africa is the best form of assessment and the one most conducive to promoting more quality research across all disciplines.

From this point of view, South Africa currently operates an individualised form of research evaluation that is shunned by highly research productive countries in Europe. For example, in the UK, the Netherlands and Germany, a group-centred, departmental view to the quality and quantity of research output is adopted.[3] This fact and the reasons for it, cast suspicion on the likely efficacy and durability of the South African system managed by the NRF whose viability has further recently become threatened through a drastic reduction in research funding for individual rated scientists (falling from a research subsidy of ZAR200 000 over 5 years to a once-off payment of ZAR30 000 over the same period for C-rated researchers). However, unlike Callaghan's[1] recent argument that the NRF is not scientific enough in its individualised rating assessments and should aim to be more so, I suggest that a truly objective and fair individualised rating system is not achievable and that rather than attempting to be more scientific in its evaluations, the NRF should rather be fairer in its assessments of a researcher by evaluating *more data* relating to individually specific, qualitative circumstances. Individual circumstances concerning ethnicity, gender and disability need to take central stage in individualised researcher grading if the grading is to be considered fair in a country such as South Africa with a long history of exclusion. The fact that it is not possible to consider in a scientifically rigorous way, the effects of ethnicity, gender and disability on research performance does not mean that consideration of such factors should be discounted altogether from NRF evaluations. In fact, and speaking generally, while race and gender have rightly attained central focus, and prescriptive remedial steps have been strongly implemented to try to address past discrimination, disability has become the 'Cinderella' of this thrust for redress and often is left out of equity deliberations and policies altogether. Although the proportion of disabled people in the South African general population (and therefore those who have been excluded by discriminatory practices) is comparatively smaller than those affected by past policies of race and gender exclusion, it does not mean that they should not also be dealt with fairly. In other words, disabled South African researchers must be given a fair opportunity to take up their rightful place by fairly assessing their contribution in relation to their particular circumstances and thus allowing a fair recognition of their talents to the overall research programme.

Of course, it should be stated at the outset that in any form of human assessment or judgement, errors are inevitable. It might appear to follow from this assertion then that the best way to avoid human subjective error is to use hard objective data, i.e. to become more objective and scientific as Callaghan[1] suggests, and thus eliminate as far as possible errors that arise from human subjective judgement. However, I argue that a truly scientific rating system worthy of that name is beyond reach and that objective assessments using hard, objectively verifiable data alone, would tend to *lessen* the fairness of such assessments, particularly in the context of South Africa where past injustices have left a legacy of inequality and injustice. It is my view that it is not that the NRF system is not scientific enough in its evaluations, as Callaghan[1] suggests, but that a more scientific approach to such evaluations is beyond reach, and that the adoption of such an approach, even if practically possible, would tend to *diminish* the validity of such evaluations (although it might increase their reliability), because crucial *qualitative* criteria for making fair assessments would summarily be ruled out of order.

Science and non-science demarcation and the NRF rating system

Callaghan[1] maintains that the current NRF system is not scientific. Perhaps the most generally accepted view of the demarcation between science and non-science is presented by the seminal work of Karl Popper in the philosophy of science. Popper's[4] notion of demarcation between science and non-science is methodological, i.e. it is not

the subject matter itself that makes being scientific possible, but the methodological approach used. Thus, for Popper, sociology can be as scientific as physics. Being scientific is adopting a method which is, for Popper[4], 'falsificationism'; that is, scientific hypotheses must be both testable and open to falsification. Clearly, in this strict sense, evaluations in which expert reviews are used as a core resource for evaluation, such as those by the NRF, cannot be formally tested nor open to falsification. To that extent, NRF individualised researcher evaluations are not truly scientific. Such evaluations may be made more objective through the manner Callaghan[1] suggests, but by doing so, will lose a crucial criterion of fairness.

The review process

It is often said by NRF administrators that the process of grading depends very largely on the detail and content of the peer-review reports. Boshoff[2] gives a good outline of the NRF review process which he regards as both elaborate and sophisticated in its apparent thoroughness. However, I do not take this sophistication to mean, as Boshoff[2] apparently does, that the system is not prone to errors of evaluation because of the number of checks and balances in the evaluation process. Nor does the possibility of appeal against the outcome of the evaluation process seem to be one that bears much fruit for those who have taken this course of action to redress what they consider to be substantive errors in the outcome of such evaluations. For example, at the University of the Witwatersrand, very few appeals for reconsideration of grading were successful in 2017. Such an apparent mismatch between the number of appeals and successful outcomes thereof, casts doubt on the efficacy of the appeals process. The number of unsuccessful outcomes is surprising given that this University has expert knowledge and considerable experience of what is required for specific researcher grading categories, and that appeals are vetted for their reasonableness and cogency by the University before they are lodged with the NRF for consideration. On the face of it, this situation suggests that either the University is operating with a 'pie in the sky' notion of the NRF classification system, and one which seeks to maximise the chances of obtaining the highest number of research grades in the highest possible categories (which seems highly unlikely), or there are fundamental misunderstandings between the two bodies. Whatever the reason, there can be little doubt that the general tendency for appeals to be rejected evokes negative reactions in the researchers concerned and is generally highly counterproductive to the fundamental research enterprise.

As Callaghan[1] points out, the current review process is not blind. Not only do evaluation candidates indicate their reviewer choices, they are able also to request exclusion of reviewers who they feel might be most harshly opposed to them. This makes the evaluation process open to subjective assessment at best, and vulnerable to unbridled bias at worst.

I do not subscribe to the view that large sophisticated institutions simply because of their size and sophistication, or indeed the apparent 'good' for which they aim, cannot be wholly misguided.[2] History is littered with the failure of institutions precisely of that type, which may have been initiated with the intent to do good, yet later were found to have created at least as many problems as they solved. (The British National Health system might, arguably, serve as but one modern example.)

However, Rawl's[5] theory of justice which regards justice as *fairness* in social institutions, presents two basic priority rules which are worth considering in this context. The first rule states that principles of freedom – such as freedom of speech, association and religion – can only be restricted when doing so results in other freedoms. The second rule, which has much greater relevance to the argument here, is that justice or fairness is always more important than the efficacy or utility of outcomes, and that, in particular, equal opportunity is more important to justice and fairness than utility. As a social institution providing benefits and responsibilities of scarce resources, which in this case is research funding money to academic researchers, the NRF must ensure that fair opportunities exist for all researchers to fully enjoy these benefits through a fairly managed system of access to extant opportunities.

Lack of specificity and definition of graded categories

I have operated several times as a reviewer for the NRF and I have been amazed by the lack of specific guidelines on how to proceed with my review. For example, the grading system is aimed at categorising a particular individual into one of several possible grades:

- A – Leading international researchers
- B – Internationally acclaimed researchers
- C – Established researchers
- P – Prestigious awards
- Y – Promising young researchers

However, even as an established researcher with some knowledge of the NRF grading system, it becomes difficult to make an assessment of individual candidates' applications in terms of in which grades they might reasonably fit, because there are no clear guidelines in extant evaluator documents indicating how to do so. The reviewer is left to describe and assess an applicant's research without having any formal parameters to make specific categorisations. In effect, the reviewer is tasked with the job of 'measuring' an applicant's research prowess without having a common yardstick with which to perform this difficult task.

Weasel words in NRF grading

Merriam-Webster Dictionary[6] defines a weasel word as 'a word used in order to evade or retreat from a direct or forthright statement or position'. The American Heritage Dictionary[7] defines it as 'an equivocal word used to deprive a statement of its force or to evade a direct commitment'. As far as I am aware, the use of equivocal words in the NRF grading procedure is not a deliberate policy to make their validity in comparisons of individual rating outcomes difficult to judge, but it is, nevertheless, a direct consequence. In this regard, for example, the meaning and interpretation of the word 'acclaim' is crucial in that it affords the springboard to promotion from the NRF's lowest grade grouping.

The classification into separate categories of research excellence by the NRF is relatively clear for Grades A and C. An A-rated researcher is defined as 'a leading international researcher', which is fairly clear in the sense of what it purports to cover, although *leading* in international research may refer to different spheres of influence and therefore impact. For example, a leading international scholar of South African English literature will have a smaller critical audience and rivalry for the accolade of 'leading', and significantly less international impact and recognition for their research, than a research scientist working on a cure for HIV or malaria. This would suggest that a leading researcher in the field of medical science will be required to reach a quite different level of research excellence and to find a space as a leader in a much larger competing group of potentially leading international researchers than their colleagues in the humanities. In this regard, for example, Fedderke[8] found that biological C-rated researchers had on average the same h-index as A-rated researchers in the social sciences.

The NRF C-rating for established researchers is, similarly, a reasonably clearly defined and easily operationalised category, although again the requirements to become an 'established researcher' are quite different in stringency demands between disciplines in terms of what is regarded as minimum acceptable performance.

However, the concept of 'international acclaim' is largely a subjective one and one which cannot be effectively tested in the Popperian[4] sense, because the concept itself cannot be effectively operationally defined, and the meaning and objective measurement of, for example, the 'relationship' between what is meant by 'acclaim' and what may be considered internationally recognised scientific contributions to knowledge may not be one and the same thing. The Cambridge Dictionary[9] defines acclaim as 'public approval and praise', but in the context of scientific research can one reasonably expect all valuable and internationally recognised research to be 'acclaimed'. Lakatos'[10] idea of a scientific research programme is that it is not scientific for all time

but slips over time from being 'progressive' to becoming degenerative[11]. Lakatos[10] talks of a scientific 'hard core', of theory and methodology which is treated as irrefutable by its promoters with a protective belt of auxiliary hypotheses that have still to be fully tested. The protective belt acts as a defensive mechanism, thereby allowing a period of 'normal science' before a paradigm shift occurs through a revolutionary change in scientific perspective.[12] These ideas are particularly problematic in this regard because research that is publicly approved and praised today, may not be so tomorrow. But more importantly for the work of a researcher who is undergoing evaluation, there are, in the social sciences, multiple methodologies and theories co-existing and while one group of the research audience may regard a researcher's work as attaining the 'holy grail' of theoretical and methodological perfection, another group in the same audience may regard it in much the same way as a natural scientist today would regard astrology or alchemy. It is true to say, however, that the NRF has recently tried to put the 'concept' of 'international acclaim' on a more intelligible footing, by offering further criteria for its identification and including an addendum to the existing definition that states that despite this change, it may 'not be considered exhaustive'. This change is tantamount to an admission that the word 'acclaim' is next to impossible to define clearly in any scientifically operational way that would allow objective measurement. In short, therefore, any attempt at making an evaluation more scientific in a rigorously testable way is likely to fail; and to generate as many Type 1 as Type 2 errors.

Of course, this is not to say that objective criteria of 'international acclaim', such as the number of invited plenary presentations or the number of international citations in quality journals, should not be considered – clearly they should, but holistic qualitative data regarding a particular scientist's specific circumstances also need to be carefully considered. For example, is it fair that a scientist who suffers from a sensory disability (e.g. blindness or deafness) should be judged in the same manner as other scientists in relation to key criteria for a NRF rating? Criteria such as the size of their scholarly network and reputation (obtained significantly through interaction and communication with colleagues at local and overseas conferences) and the number of invited plenary sessions hold the same weight for applicants whether or not they are disabled.

Conclusion

While agreeing with Callaghan[1] that there are flaws in the NRF's current grading of individual researchers, some of which may be endemic to a system that tries to distinguish between the excellence of research output between individual researchers, I disagree that the solution is for the institution to become more objective (or 'scientific') in its evaluations. My concern is that if it were to do so, it would not apportion fairness in its assessments and ignore the particular circumstances of, for example, researchers with disabilities, thus lessening their equality of opportunity and access to scarce research funding resources.

It needs to be said that the current debate has been ongoing in the literature for some time. Campbell[13] for example refers to it as the binary option of university research evaluation; the binary aspects being:

- *peer review* consisting of evaluations based on subjective expert opinion which broadly corresponds to the current NRF system, and

- *indicators* consisting of judgements of research or researcher excellence based on objective quantitative data.

It is noted that while peer-review evaluations allow a higher degree of complexity in assessments, they have a strong dependence on the composition of the panels which can create personal, methodological and theoretical bias. Indicators can be more objective but tend to be superficial. Whether a fair balance can be found in NRF-type individual researcher evaluations, to fairly assess equality of opportunities of particular researchers presented with specific life realities, remains doubtful and is a significant factor motivating the use of the more generally applied *group-centred* researcher evaluation systems in most other countries today.

References

1. Callaghan C. A review of South Africa's National Research Foundation's rating methodology from a social science perspective. S Afr J Sci. 2018;114(3/4), Art. #2017-0344, 7 pages. https://doi.org/10.17159/sajs.2018/20170344

2. Boshoff C. Chris Callaghan's criticism of the National Research Foundation's rating methodology: A rebuttal. S Afr J Sci. 2018;114(7/8), Art. #a0278, 4 pages. https://doi.org/10.17159/sajs.2018/a0278

3. Barker K. The UK Research Assessment Exercise: The evolution of a national research evaluation system. Res Eval. 2007;16(1):3–12. https://doi.org/10.3152/095820207X190674

4. Popper KR. Science: Conjectures and refutations: The growth of scientific knowledge. London: Routledge; 1963.

5. Rawls J. A theory of justice. Cambridge, MA: Harvard University Press; 1971.

6. Merriam-Webster English Dictionary [online]. Weasel [cited 2018 Jul 27]. Available from: https://www.merriam-webster.com/dictionary/weasel%20word

7. American Heritage Dictionary. 5th ed. Boston, MA: Houghton Mifflin Harcourt Publishing Company; 2016.

8. Fedderke J. The objectivity of National Research Foundation peer review based ratings in South Africa. ERSA Working Paper 300 [document on the Internet]. c2012 [cited 2017 May 15]. Available from: https://econrsa.org/system/files/publications/working_papers/wp300.pdf

9. The Cambridge English Dictionary [online]. Acclaim [cited 2018 Jul 27]. Available from: https://dictionary.cambridge.org/dictionary/english/acclaim

10. Lakatos I. The methodology of scientific research programmes: Philosophical papers volume 1. Worral J, Currie G, editors. Cambridge: Cambridge University Press; 1978.

11. Blaug M. The methodology of economics. Cambridge: Cambridge University Press; 1980.

12. Kuhn TS. The structure of scientific revolutions. 2nd ed. Chicago, IL: University of Chicago Press; 1970.

13. Campbell DFJ. Conceptual framework for the evaluation of university research in Europe: Working paper. Washington DC: Center for International Science and Technology Policy (CISTP), George Washington University; 2002.

AUTHOR:
Jennifer M. Fitchett[1] iD

AFFILIATION:
[1]School of Geography, Archaeology and Environmental Studies, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Jennifer Fitchett

EMAIL:
Jennifer.Fitchett@wits.ac.za

On defining droughts: Response to 'The ecology of drought – a workshop report'

Situated between the subtropics and the mid-latitudes, South African climate is dominated by persistent anticyclonic conditions year round.[1,2] As a consequence, the region is relatively water scarce by global standards.[2,3] During years of below-average rainfall, significant biometeorological impacts are felt by natural systems and human populations, yet often these are under-researched and under-reported, particularly in developing regions.[4] Efforts to bring together interdisciplinary teams to collaborate in quantifying the impacts of adverse climatic conditions on ecosystem function are thus tremendously valuable. Swemmer et al.[5] report on one such workshop, presenting important insights into the detriments, and in some cases benefits, of the below-average rainfall experienced in 2015–2016.

The findings regarding the impacts on herbivore mortality, migration and the utilisation of artificial waterholes, and the spatial heterogeneity in vegetation changes including grass and forb growth patterns and tree deaths[5], are valuable in tracking the impacts of regional-scale water scarcity during events of below-average rainfall. These findings are not disputed. Rather, this Commentary reflects on issues surrounding the definition of a drought event, and in particular the use of a drought classification scale that discriminates exclusively between 'wet' and 'drought' conditions. It must be stated outright that the definition of drought does not affect the veracity of the workshop report by Swemmer et al.[5] pertaining to the impacts of a rainfall reduction on herbivores and vegetation, nor do I contest that below-average rainfall occurred during this period.

Swemmer et al.[5] present a map demonstrating 24-month standardised precipitation index (SPI) scores for an end run-date of May 2016. On the basis of this map, the authors argue that a severe to extreme drought 'appears' to have been experienced in northeastern Mpumalanga, the eastern half of the Free State and northern KwaZulu-Natal.[5] It is curious that the authors use such tentative language, when the occurrence of a drought formed the primary basis of their workshop. It is also curious that they did not perform independent climatological analysis, nor present the methodology for SPI calculation conducted by the ARC or justification for the use of a 24-month SPI plot (as opposed to 6- or 12-month plots which are also produced by ARC in their UMLINDI newsletters).

The SPI is one of the most widely used measures of meteorological drought conditions globally[6], and, following the Lincoln Declaration[7], it has been formally adopted by the World Meteorological Organization (WMO) as the formal metric on which drought conditions should be classified and quantified, with a *WMO SPI User Guide* published to standardise calculation and interpretation of SPI score[8]. It is therefore highly appropriate to use the SPI in a workshop on drought and any publications that follow, but such use of the SPI should follow WMO conventions to facilitate comparison of results across WMO member countries. The *WMO SPI User Guide* groups SPI scores into seven discrete classifications ranging from extremely dry, with scores <-2, to extremely wet, with scores >+2, and with scores ranging from +0.99 to -0.99 classified as near normal (Table 1[8]). A drought event is thereafter declared on the basis of SPI scores of <-1.5, i.e. severely dry conditions.[9] The map presented by Swemmer et al.[5] is not consistent with the *WMO SPI User Guide* classifications. The plot discriminates eight rather than seven classifications which span 'extremely wet' to 'extreme drought'.[5] There are no SPI values assigned to any of these classifications, so it is not possible to determine whether the upper boundary of 'mild drought' is associated with an SPI score of -1.5, or whether it refers to all SPI scores of less than zero. The plot also contains no classification for normal conditions.

Table 1: *WMO SPI User Guide*[8(p.4)] standardised precipitation index (SPI) value classifications

SPI value	Classification
2.0+	Extremely wet
1.5 to 1.99	Very wet
1.0 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1.0 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2.0 and less	Extremely dry

This classification system is reminiscent of the original presentation of the SPI by McKee et al.[10] in a 1993 conference proceedings paper (Table 2), which provides four classifications of drought conditions with SPI scores spanning 0 to <-2. However, this first publication predates the Lincoln Declaration, and the initial SPI had been revised in the inter-leading period in terms of both the method of calculation and the classifiers. The re-categorisation in particular followed critiques by leading academics in the domain of drought climatology, such as Prof. Agnew of University College London who wrote 'In McKee's classification…all negative indexes (SPI) are taken to indicate the occurrence of drought; this means for 50% of the time, drought is occurring. This is clearly

nonsense!'[11]. This sentiment was supported by Hayes[12], Guttman[13] and the US National Drought Mitigation Centre[14] among others, which lead to the re-categorisation in the *WMO SPI User Guide*[8]. It is curious that this plot uses a classification that has been rejected in published literature and revised in a formal WMO user guide. Unfortunately, as the map uses an outdated classification system with no quantifiable metrics on which a reclassification could be inferred, the definition by Swemmer et al.[5] of 2014–2016 as a period of drought cannot be supported by the data presented.

Table 2:   Initial standardised precipitation index (SPI) value classifications by McKee et al.[10(p.2)]

SPI value	Drought category
0 to -0.99	Mild drought
-1.00 to -1.49	Moderate drought
[-*]1.5 to -1.99	Severe drought
≤-2.00	Extreme drought

In the McKee et al. paper this is presented as 1.5 to -1.99, but is no doubt a typo.

The findings regarding the ecological impacts of rainfall reduction would be more strongly presented if they were coupled by independent climatological analyses, with clearly outlined methodology that is consistent with WMO user guides. The authors themselves express certain reservations with their approach to the measurement of drought in their discussion section:

> It was clear that annual rainfall is not a sufficient metric to properly understand, and predict, the ecological impacts of drought. Other metrics such as the length of intervals between rainfall events or the incidence of rain in the dry season, also need to be considered, as well as the effect of higher temperatures […] spatial variability in rainfall also needs to be measured adequately to properly understand the severity of drought.[5]

Indeed, an analysis of a wider range of climatological variables, the exploration of a wider range of drought definitions including hydrological drought (where dam or river levels are reduced) and agricultural drought (where soil moisture levels decline, often as a function of coupled temperature-precipitation changes) would strengthen their analysis of any drought-related ecological changes.[15] A key threat posed by climate change is an increased frequency, intensity and severity of extreme climatic events – it is therefore imperative that we accurately monitor, record and reflect on such hazards.

References

1. Tyson PD, Preston-Whyte RA. The weather and climate of southern Africa. Oxford: Oxford University Press; 2005.

2. Roffe SJ, Fitchett JM, Curtis C. Classifying and mapping rainfall seasonality in South Africa: A review. S Afr Geog J. In press 2019. https://doi.org/10.1080/03736245.2019.1573151

3. King J, Mitchell S, Pienaar H. Water supply and demand. In: King J, Pienaar H, editors. Sustainable use of South Africa's inland waters. Pretoria: Water Research Commission; 2011. p. 1–16.

4. Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. Climate extremes: Observations, modelling and impacts. Science. 2000;289(5487):2068–2074. https://doi.org/10.1126/science.289.5487.2068

5. Swemmer AM, Bond WJ, Donaldson J, Hempson GP, Malherbe J, Smit IPJ. The ecology of drought – a workshop report. S Afr J Sci. 2018;114(9/10), Art. #5098, 3 pages. https://doi.org/10.17159/sajs.2018/5098

6. Amin Asadi Zarch M, Sivakumar B, Sharma A. Droughts in a warming climate: A global assessment of standardized precipitation index (SPI) and reconnaissance drought index (RDI). J Hydrol. 2015;526:183–195. https://doi.org/10.1016/j.jhydrol.2014.09.071

7. World Meteorological Organization. Lincoln Declaration on Drought Indices [document on the Internet]. c2009 [cited 2018 Dec 21]. Available from: http://www.wmo.int/pages/prog/wcp/agm/meetings/wies09/documents/Lincoln_Declaration_Drought_Indices.pdf

8. World Meteorological Organization. Standardised precipitation index user guide. WMO-no. 1090. Geneva: World Meteorological Organization; 2012.

9. Roault M, Richard Y. Intensity and spatial extension of drought in South Africa at different time scales. Water SA. 2003;29(4):489–500.

10. McKee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration to time scales. Paper presented at: Eighth Conference on Applied Climatology; 1993 January 17–22; Anaheim, CA, USA.

11. Agnew CT. Using SPI to identify drought. Drought Network News. 2000;1:6–12.

12. Hayes MJ. Revisiting the SPI: Clarifying the process. Drought Network News. 2000;18:13–14.

13. Guttman NB. Accepting the standardized precipitation index: A calculation algorithm. J Am Water Resour As. 1999;35(2):311–322. https://doi.org/10.1111/j.1752-1688.1999.tb03592.x

14. NDMC. Drought monitoring: SPI program [software on the Internet]. No date [cited 2018 Dec 20]. Available from: https://drought.unl.edu/droughtmonitoring/SPI/SPIProgram.aspx

15. Duan K, Mei Y. Comparison of meteorological, hydrological and agricultural drought responses to climate change and uncertainty assessment. Water Res Man. 2014;28(14):5039–5054.

AUTHORS:
Juanette John[1,2] (iD)
Sonali Das[1,2,3,4] (iD)

AFFILIATIONS:
[1]Council for Scientific and Industrial Research, Pretoria, South Africa
[2]DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa
[3]Department of Statistics, Nelson Mandela University, Port Elizabeth, South Africa
[4]The School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa

CORRESPONDENCE TO:
Juanette John

EMAIL:
jjohn@csir.co.za

State of African research: Snapshot of abstracts for the 2018 Women in Science Without Borders indaba

According to the African Union's *Agenda 2063*, the development of Africa depends critically on science, technology and innovation.[1] However, Africa contributes to only about 1.3% of the global investment in research and development (R&D). In response, the African Union and the New Partnership for Africa's Development (NEPAD) initiated a series of consultations in 2003 which resulted in Africa's Science and Technology Consolidated Plan of Action (2005–2014).[1] In South Africa, the Draft White Paper on Science, Technology and Innovation was released for public comment in September 2018.[2] This document acknowledges that there has been slow progress on the gender agenda in many countries and includes gender transformation in the pillars required for a South African responsible research and innovation approach. Measures to increase cooperation between relevant stakeholders include the development of gender-sensitive monitoring and evaluation mechanisms. There are, however, concerns that this document does not clearly articulate the required strategy that will enable science and technology to contribute to the solutions for some of the country's challenges.[3] A number of independent views have suggested that a variety of methods and processes is required to find solutions to these challenges.[3,4]

In 2018, the annual Women in Science Without Borders (WiSWB) Indaba (conference) was held in Johannesburg (South Africa) to highlight and emphasise particularly the contribution of women researchers within the science, technology, engineering, mathematics, medicine and innovation (STEMMI) ecosystem to effect positive outcomes in a world in which, in fact, half of the consumers of STEMMI offerings are women. The 2018 WiSWB conference was thus an opportunity to focus on, inter alia, transdisciplinary issues that particularly have an Africa-wide urgency, such as institutional capacity development, inclusive sustainable growth and regional integration, through a gender lens with the overarching theme of the conference in 2018 being 'Resilience in Diversity'. During the preparation leading up to the 2018 WiSWB conference it was also realised that there is a dearth of centralised online information related to transdisciplinary research for both South Africa and Africa in general, that has captured both technical expertise along with demographic details of researchers operating within Africa. In order to achieve the targets of *Agenda 2063*, it is essential that information about transdisciplinary research, including information about the researchers, be continuously captured and updated. So over and above the core purpose of the conference which was to provide a platform to present transdisciplinary research, an additional objective of the 2018 WiSWB was to investigate how this process of collecting metadata on research might work by piloting a template which captures both conference-submission details as well as details about the researchers themselves, including, but not limited to, information on demographics and research interests at the time of submission. A particular feature of interest was to investigate the mobility of researchers, that is, from where they were born to where they resided during submission for the conference. We would like to remind the reader that the conference was not discipline focused but rather problem focused. Hence to encourage interdisciplinarity, submissions from prospective participants (gender-inclusive) had to fall within one of a number of themes, which also cut across a number of South Africa's national priorities, namely, Clean Energy, Climate Change, Digital Revolution, Disaster Management, Education and Outreach, Food Security, Gender Studies, Health, Industrialisation, Science Diplomacy, Smart Cities and Water. The intent was to also facilitate the identification of potential national as well as international collaborators working within the same themes with the purpose of strengthening further research impact.

Data

Data used in the analysis were acquired from the submission portal that was used by the conference organisers to enable the authors to submit their abstracts for peer review for the conference. The drop-down options for various items in the submission form was designed to capture various aspects of the research as well as of the researchers. The analysis presented below is a subset from the submissions database restricted to the African-Arabian region submissions only. No information presented here is meant to identify any individual submission; rather the intent is only to use this information to present a snapshot of the science landscape during the period leading up to the 2018 WiSWB conference. This snapshot is also not meant to be generalised.

Summary statistics

Descriptive statistics of all African and African-Arabian submissions

Preliminary results reveal that the country with the most submissions was Nigeria (55%); more than 60% of submitters were in the 30–45 age group and almost 50% of the main authors had a doctoral qualification. Most of the first authors who submitted abstracts were female (Figure 1a) and in the 30–45 age bracket (Figure 1b), and had a postgraduate or doctoral qualification (Figure 1c). This was true for both the South African and the African-Arabian submissions. Most submissions had one or two authors (Figure 1d). Most affiliations were research and/or educational institutions (Figure 2a and 2b).

Descriptive statistics by theme

The largest number of submissions overall was under the Health theme (26%), followed by Food Security (17%) (Figure 3a) – which is indicative of these being the two most urgent transdisciplinary problems in the African-Arabian region in this period. The trend in South Africa was similar to the overall trend, although Clean Energy was somewhat more prevalent and Food Security somewhat less (Figure 3b).

The proportion of countries that submitted in each theme (Figure 4) also shows that the Health and Food Security themes were addressed by almost half of the 27 countries represented (14 and 13, respectively). Nigeria and South Africa generally had the most or second-most submissions for every theme.
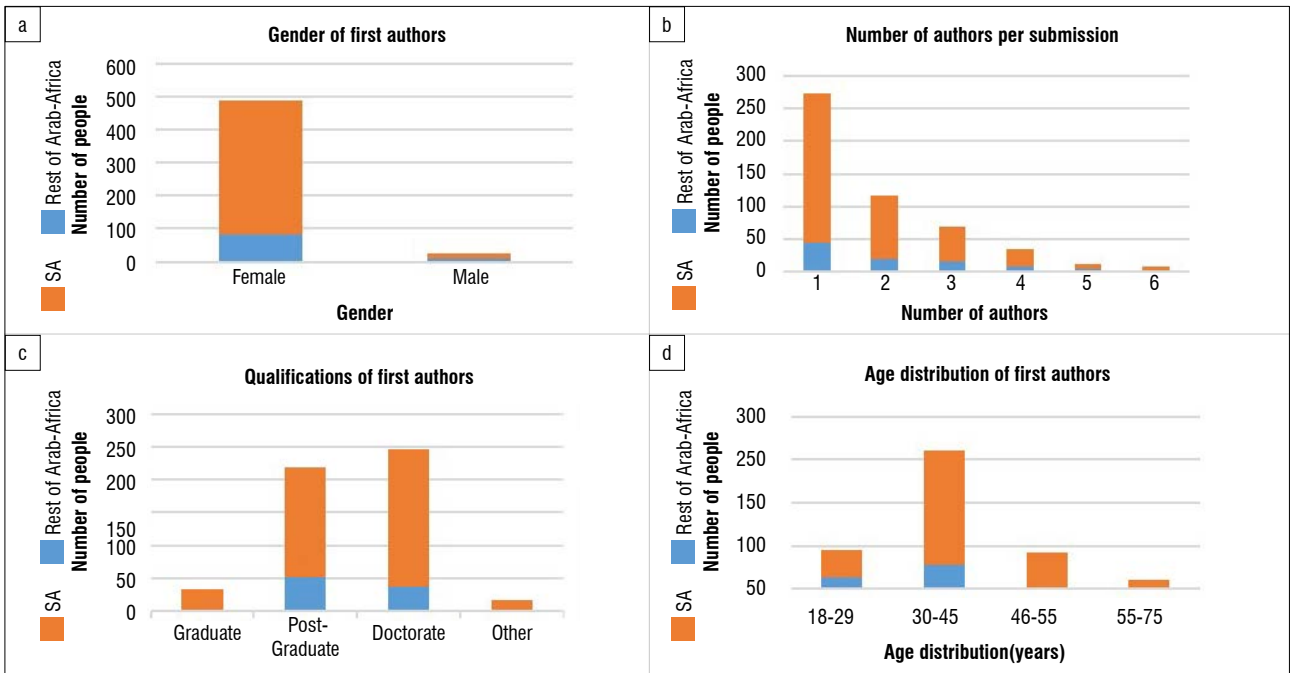
Figure 1: Descriptive statistics of all African-Arabian submissions including South Africa (SA): (a) gender, (b) number of authors, (c) qualification and (d) age.
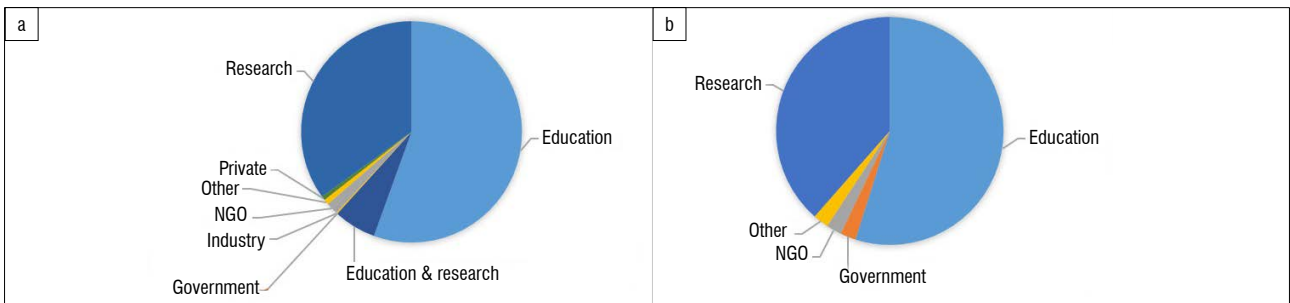
Figure 2: Affiliations represented by (a) African-Arabian and (b) South African submissions.
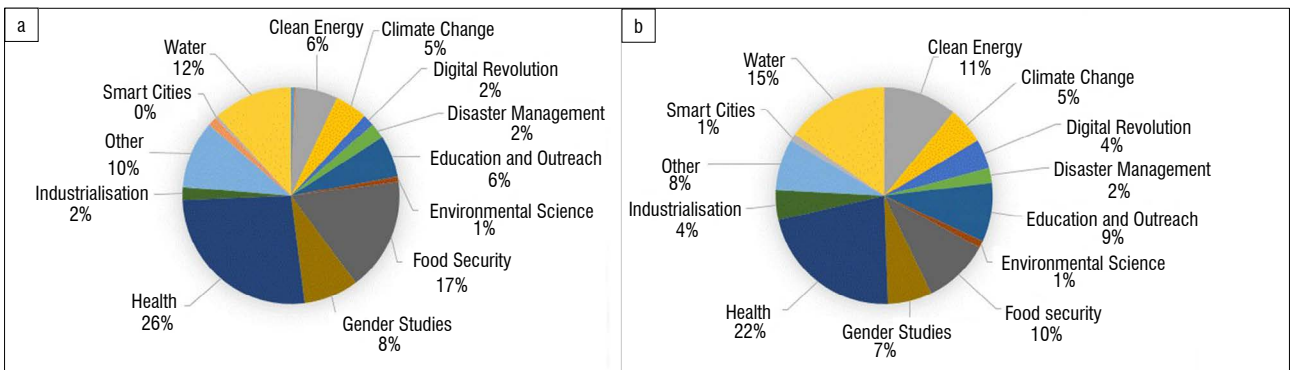
Figure 3: Themes represented by (a) African-Arabian and (b) South African submissions.
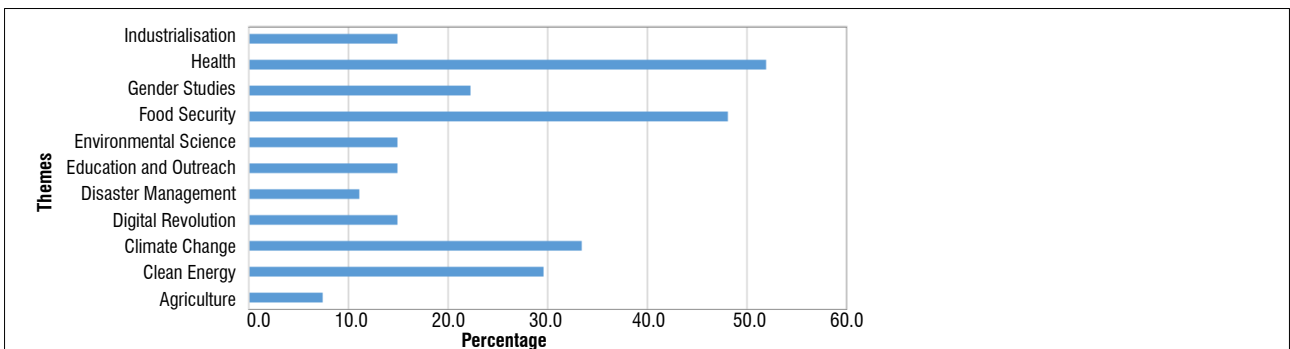
Figure 4: Proportion (%) of African-Arabian countries from which authors submitted in each theme.

Table 1 shows an example of the themes that were selected by the most submitters of abstracts in a particular country. This table shows that the Health theme was the most-selected theme in 12 countries. Although many of these countries had very few submissions and it is thus not generalisable to the whole country, this table provides an indication of the type of information that could be useful for a variety of purposes should the sample be increased. The table shows, for example, that 56% of abstract submissions from Tunisia were classified under the Health theme, 50% of all submissions from Ghana were in Gender Studies and 23% of all submissions from Ethiopia were Food Security related.

Kumu analysis of submission data for the African-Arabian region

An envisaged outcome of the conference was to establish a database of metadata from the submissions to identify common goals in the region using Kumu, an open-source mapping tool. Kumu (meaning teacher or source of wisdom in Hawaiian), is a web-based visualisation platform that can be used to map systems and improve the understanding of relationships (https://docs.kumu.io/about/what-is-kumu.html). Kumu allows shared access, which enables collaboration between individuals and teams. Kumu provided an ability to explore and visualise inherent complex relationships within the submission metadata, by location, age distribution, submission themes and research interests. A spatial component was applied to show different features by location. All features were displayed by age category.

Figure 5a represents the spatial distribution of all African-Arabian submissions (any type), colour coded by the age range of the main author. The locations of South African submissions are shown in Figure 5b. In the 2018 WiSWB, Nigeria was the leading contributor with the majority of them in the 30–45-year age group. Very few submissions came from north and west Africa. Overall, there were very few submissions from authors older than 46 years of age.

Submissions by theme and age category

The themes most selected by submitters, by age category, are mapped in Figure 6a–f.

South African themes

South African submissions were mostly from metropolitan areas, which is a consequence of the location of the leading tertiary institutions. More data visualisations can be obtained from the link: https://embed.kumu.io/ce17de8afaf6e5a6577f5f4ffe79e83a.

Areas of interest

In addition to the theme, submitters were requested to indicate area of interest. These areas may give a broader perspective on potential areas of collaboration.

Further analysis for combination of submission features

An example of further ways to interrogate the collected data is the location of different age groups by their qualifications (Figure 7a–c). As expected, most of the graduates were in the lowest age group (18–29 years).

Table 1: Theme with the highest number of submissions for each country

Most-selected theme by each country	Country	% Total submissions by country (total number by country)
Clean Energy	Burkina Faso	100% (2)
	Uganda	100% (1)
	Algeria	50% (2)
Health	Jordan	100% (2)
	Malawi	100% (1)
	Oman	100% (1)
	Mauritius	75% (4)
	Benin	67% (9)
	Yemen	67% (3))
	Kenya	50% (8)
	Cameroon	40% (5)
	Egypt	30% (20)
	Nigeria	28% (286)
	Sudan	24% (21)
	South Africa	22% (92)
Water	Mozambique	100% (1)
	Tunisia	56% (9)
Food Security	Rwanda	100% (1)
	Botswana	67% (3)
	Ethiopia	23% (13)
	Congo	50% (2)
	Tanzania	29% (7)
Climate Change	Zimbabwe	100% (1)
	Madagascar	50% (2)
	Tanzania	29% (7)
Gender Studies	Congo	22% (2)
	Ghana	50% (18)
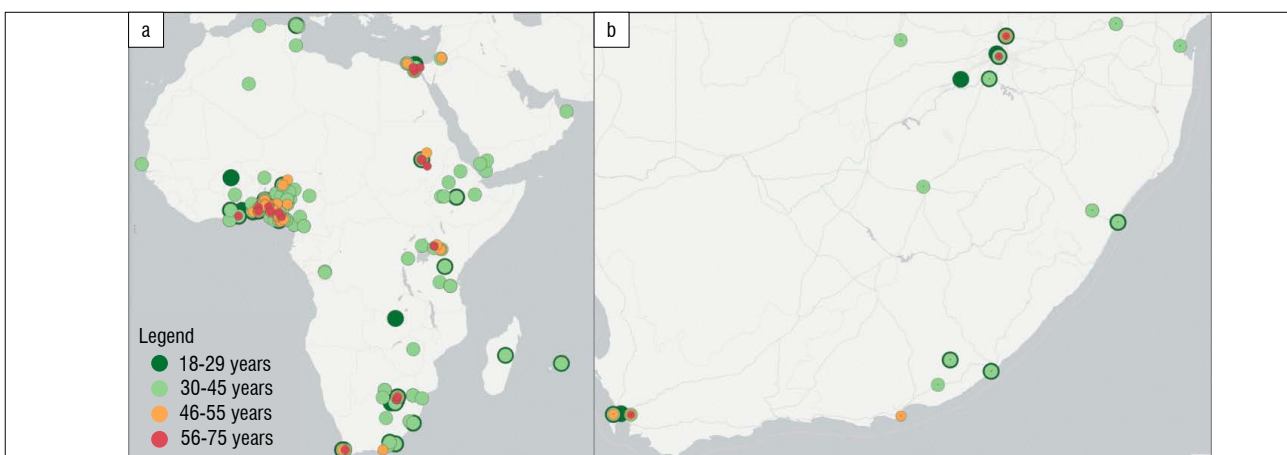
Figure 5: Cities represented by all (a) African-Arabian and (b) South African submissions, by age category.
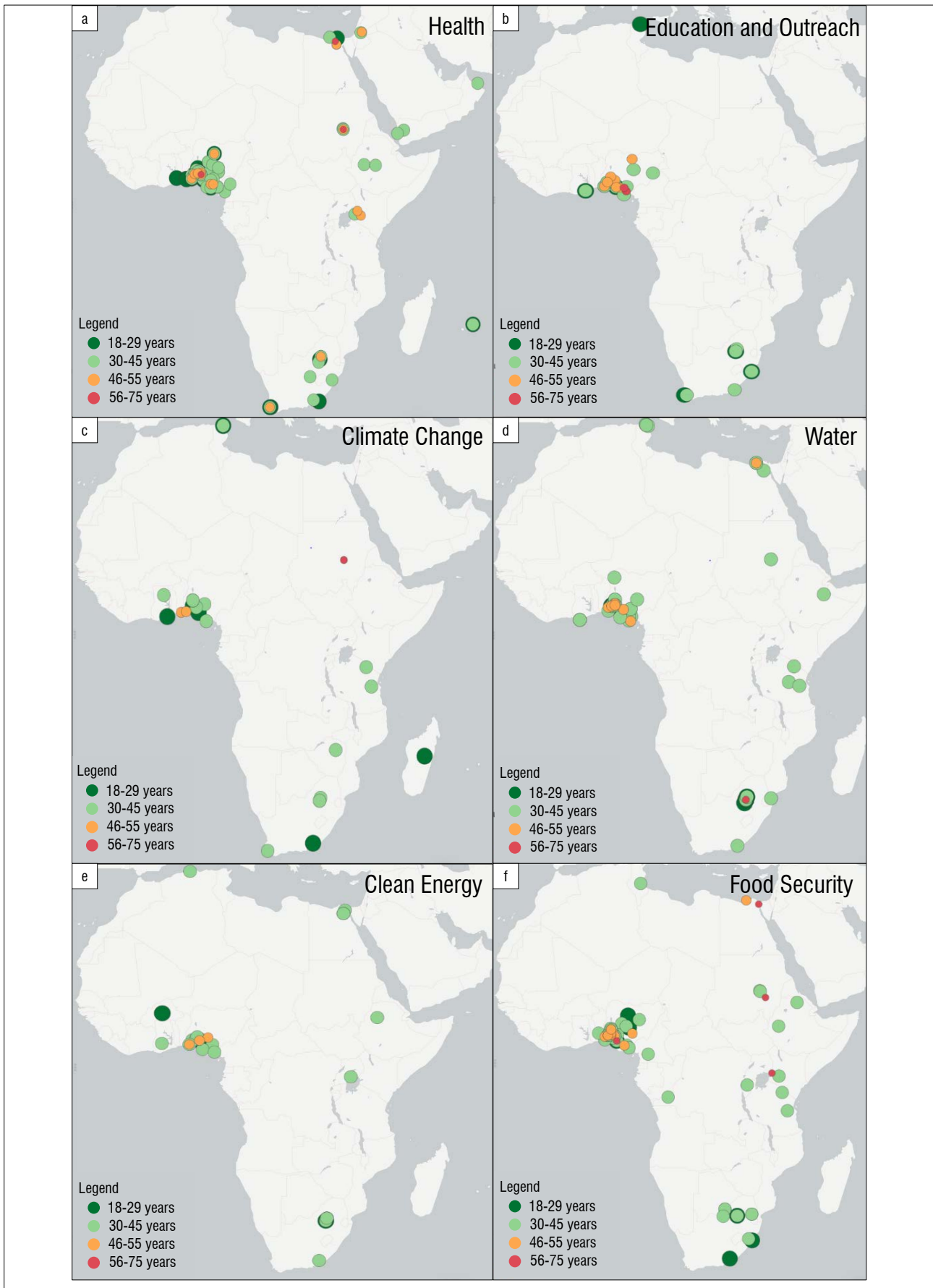
Figure 6: The most common themes selected by submitters, by age category.
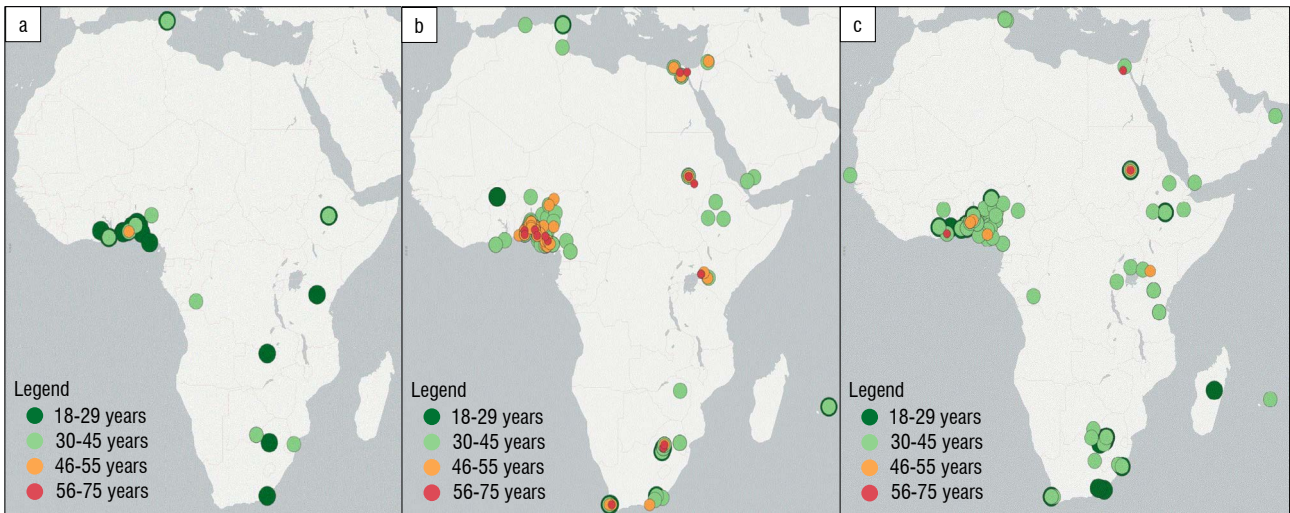
Figure 7: Age distribution by (a) graduate, (b) postgraduate and (c) doctoral qualifications.

It is interesting to note that most of the 56+ age group featured in the postgraduate category, while the 30−45-year group had a prominent presence in the doctorate category.

The birth countries of first authors who indicated that their country of birth was different from the country from which they were submitting, are indicated in Figure 8. This figure provides an indication of the degree of migration by researchers. It is encouraging to note that researchers from a variety of countries, even from outside of Africa, are working within the African continent to pursue higher education and research.
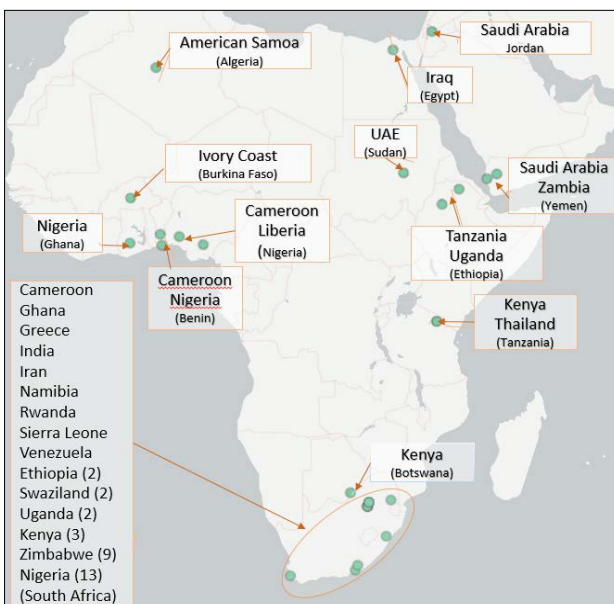
Figure 8: Birth country of author where it differed to country of residence (country of submission in brackets).

Concluding remarks and recommendations for the way forward

We have presented some findings from the abstract peer-review process of the 2018 WiSWB. A specific focus has been on the technical details as well as details related to the broader demographic and interdisciplinary nature of the submissions.

To effect any of the objectives of Africa's Science and Technology Consolidated Plan of Action (2005–2014), it is vital to measure what

needs to be managed. During the preparations leading up to the conference, a rapid desktop search on R&D expenditure in Africa found only a few online references that provide government expenditure in STEMMI R&D, and then for only some countries.[5-7] Even for South Africa, there is no single portal that provides details of headcount and full-time equivalents by gender, age, qualification or specialisation, let alone over time. No central location was found where similar data are archived on a regular basis for reference purposes. This gap in readily available R&D metadata motivated the collection of submission information to allow a snapshot view of the interplay among gender, age, qualification, location and research interests at the time of submission. The intent was thus to not only focus on traditional disciplines but particularly on how it can be applied across different themes to solve real-world problems. This information enabled us to visualise the state of various related factors across the continent.

Although the 2018 WiSWB conference was gender-inclusive, almost all submissions overall, as well as from the African-Arabian region, were by female authors. Hence, it provided an adequate data set to assess the landscape of research opportunities for female STEMMI researchers, by age, country and qualification. Also, the themes provided a potential proxy for the pressing problems in their respective countries. Given that the submission portal was piloted, it is acknowledged that it can, and should, continuously evolve to be more representative and relevant. It goes without saying that analysing obsolete data for impactful planning in R&D is a futile exercise. African countries should also adopt the practice of maintaining and sharing records on R&D expenditure on an annual basis regarding all aspects of data using the format of the US-NSF,[8] and make the data accessible for monitoring and evaluation purposes.

In conclusion, it is recommended that the African Union champions the setting up of a taskforce to coordinate the collection of detailed research profiles of researchers and research outputs. This initiative will assist in meaningful consolidation of R&D data to understand the status quo, and to identify locations for interventions as appropriate, to provide tangible inputs to help make the African Renaissance a reality.

References

1. Marsh K. How Africa can close its continent-wide science funding gap. The Conversation. 2016 April 12; Science and Technology [cited 2018 Dec 9]. Available from: https://theconversation.com/how-africa-can-close-its-continent-wide-science-funding-gap-55957

2. South African Department of Science and Technology (DST). Draft white paper on science, technology and innovation. Pretoria: DST; 2018 [cited 2018 Sep 26]. Available from: https://www.dst.gov.za/images/2018/Draft-White-paper--on-STI-7_09.pdf

3. Walwyn D, Cloete L. Draft White Paper on Science, Technology and Innovation neglects to prioritise issues of performance and human capability. S Afr J Sci. 2018;114(11/12), Art. #5679, 6 pages. https://doi.org/10.17159/sajs.2018/5679

4. Carvalho A. Wishful thinking about R&D policy targets: What governments promise and what they actually deliver. Sci Public Policy. 2017;45(3):373–391. https://doi.org/10.1093/scipol/scx069

5. Index mundi. Research and development expenditure (% of GDP) – Africa [webpage on the Internet]. c2017 [cited 2018 Dec 09]. Available from: https://www.indexmundi.com/facts/indicators/GB.XPD.RSDV.GD.ZS/map/africa

6. Index mundi. Researchers in R&D (per million people) – Africa [webpage on the Internet]. c2017 [cited 2018 Dec 09]. Available from: https://www.indexmundi.com/facts/indicators/SP.POP.SCIE.RD.P6/map/africa

7. Unesco Institute for Statistics (UIS). Science, technology, and innovation [webpage on the Internet]. c2018 [cited 2018 Dec 09]. Available from: http://data.uis.unesco.org/#bThemes

8. National Science Foundation (NSF) International [homepage on the Internet]. No date [cited 2018 Dec 09]. Available from: www.nsf.org

'Gender shouldn't matter because we are all scientists here': A narration of the panel discussion at the 2nd International Women in Science Without Borders conference

AUTHORS:
Antoinique van Staden[1,9]
Nova Ahmed[2]
Yoseph Getachew[3,9]
Irvy M.A. Gledhill[4,9]
Maria Kanjere[5,9]
Sibusisiwe Khuluse-Makhanya[6]
Sonali Das[6,7,8,9]

AFFILIATIONS:
[1]Department of English, University of Pretoria, Pretoria, South Africa
[2]Department of Electrical Engineering and Computer Science, North South University, Dhaka, Bangladesh
[3]Department of Economics, University of Pretoria, Pretoria, South Africa
[4]School of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand, Johannesburg, South Africa
[5]Turfloop Graduate School of Leadership, Turfloop Campus, University of Limpopo, Polokwane, South Africa
[6]Council for Scientific and Industrial Research, Pretoria, South Africa
[7]Department of Statistics, Nelson Mandela University, Port Elizabeth, South Africa
[8]The School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa
[9]DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa

CORRESPONDENCE TO:
Antoinique van Staden

EMAIL:
ant_vst@yahoo.com

The obligation for working mothers is a very precise one: the feeling that one ought to work as if one did not have children, while raising one's children as if one did not have a job.[1]

The disparities that exist between men and women, and more so working mothers, in the participation of science, technology, engineering, mathematics, medicine and innovation (STEMMI) are a global concern.[2] In the past, efforts to address gender inequalities in STEMMI progressed at a slower pace because of the fragmented nature of gender equality advocacy efforts. In explicitly defining gender equality as a goal for sustainable development (SDG 5), there is renewed vigour in the pursuit of solutions to address discrimination on the basis of gender.[3] Challenges that constrain women's full participation in political, economic and public life are being identified, and as a result policies and strategies, including those concerning women in STEMMI in developing countries like South Africa, are being reviewed to ensure that the gender gap is reduced.[4,5]

It is against this backdrop that a discourse on gender biases in science was deemed necessary for the 2nd International Women in Science Without Borders (WiSWB) conference that was held in Johannesburg (South Africa) on 21–23 March 2018. Here we reflect the thoughts and discussions around the softer issues faced by women in the sciences which emerged from a panel discussion at the conference. The conference consisted of technical sessions which attracted peer-reviewed papers from various scientific disciplines, as well as dedicated sessions and a keynote address by Minister Naledi Pandor. Although the conference had as its primary focus the showcasing of technical research, there was a need identified in the organising phase to also discuss some of the challenges facing female scientists within the ecosystem of scientific research, as well as to obtain insights from senior male and female scientists and leaders on strategies for overcoming some of those challenges. This article reflects primarily the opinions and statements made by the various panellists. Formally, a panel discussion was facilitated in which three questions were presented to the panellists who were encouraged to share their experiences, opinions and advice with the delegates. Although there are many more challenges faced by women in the sciences, for the panel discussion the focus was on three specific issues that were deemed internationally applicable to all fields and scientists alike: (1) gender wage gap; (2) cultural perceptions and encouraging young girls to become scientists; and (3) the need for women in sciences. While specific questions were aimed at specific panellists, it was expected, and encouraged, that other panellists add to the discussion. There was also a request to the audience to defer questions until after the panel discussion was concluded because, as is almost always the case, time was limited. However, the audience was encouraged to continue the conversations amongst themselves, and with the panel members if necessary, and to use the rest of the conference as an extension of the platform aimed at providing support for female scientists. It was also clarified that 'science' is used as a blanket term to include all forms of sciences and there is no distinction between what is defined as hard and soft science. We summarise the discussion related to each of the three questions and follow by some concluding remarks.

Gender wage gap

In all fields of science, it was evident that it is a generally accepted fact that women are paid less than their male counterparts for the same work or task and comparable qualification. Theories were advanced in the discussion as to why gender wage gap persists. One theory is that women are expected to take more leave, and therefore work less than men, attributed to familial responsibilities. This theory results in the notion that women are less productive in the workplace, and hence creates the impression that paying women less is justified. Statistical discrimination theory was mentioned and translates to the interplay between cultural stereotypes and gendered preferences or outcomes such as the willingness of employers, especially in STEMMI occupations, to pay a premium for men who are viewed favourably in terms of agency, intelligence and analytical competence compared to women.[6]

Economists confirm that the gender wage gap is a universal problem that still exists in both the developing and developed world. In mainstream economics, the International Monetary Fund studies, conducted in the labour market, have shown that there is no difference in the productivity between male and female workers. Despite some economists arguing that there is no gender discrimination in the market (as discrimination could be inefficient, it may not be tolerated in a perfectly competitive market), practical evidence shows that it exists. According to Getachew et al.[7], there exists an inequality of opportunities (parental gender bias) at home, and sharing parental responsibility creates 'inequalities' within households. Getachew et al.[7] argue that such parental gender bias is a result of non-pecuniary cost associated with parental investment in children. In this research, gender bias is treated as the difference in the parents' psychic cost. This is a reflection on their optimism or pessimism towards their investment in their children's education, which leads to different human capital accumulation, and therefore to differing social mobility thresholds for daughters and sons.

In a study of a cohort of mathematically gifted individuals as adults (being in the top 1% of mathematical reasoning ability at age 13 years), Lubinski et al.[8] found that the incomes of the men were significantly higher than those of the women even though the differences in university educational attainment were not significant. Further, being married was more of an advantage for men than for women, as married men had higher incomes than married women

and unmarried men. While both genders had achieved exceptionally well in their chosen fields, their values were different, with men, as a group, valuing full-time high impact work as compared to women, as a group, valuing flexibility at work and balancing work with other aspects of life.

A similar notion regarding gender wage gap is evident from the Uber algorithm (referred to as the Gig economy[9]) which assigns customers to everybody in the Uber driver database equally without gender preference. However, there is evidence that there is a difference between the incomes of the female drivers and their male counterparts. It has since been discovered that the algorithm has identified that female drivers are more reluctant to take risks, such as taking late-night passengers to a dangerous destination, and, as a result of this choice, they were earning a lower income. In this case, equal opportunities exist, but not everybody is able, or willing, to seize opportunities for various reasons. In scientific research careers, a similar phenomenon occurs in which the career growth of women is constrained by decisions that limit mobility, such as the choice to have a family, which often coincides with the period after graduate studies when both men and women are expected to establish themselves by travelling to conferences and taking up postdoctoral positions.[10]

A further explanation for the wage gap was offered in the discussion: female scientists do not have competitive negotiation skills. There exists a perception, described in this panel session, that women are not as adept at asking for pay increases because they believe that their managers will recognise and reward their achievements if they 'just work hard enough'. A general observation was that their male counterparts tend to be more assertive. There is a perception that it is easier for men to vocalise the value they add to a company, and hence generally find it easier to talk about their own accomplishments and behave as better self-marketers. It was agreed that female scientists need to develop the skill and confidence to be able to negotiate for pay increases. Learning from other female scientists and mentors can help with understanding experiences.

An additional experience many female scientists have encountered is that of the 'imposter syndrome'. Imposter syndrome is a psychological pattern in which individuals doubt their accomplishments and have a persistent, often internalised, fear of being exposed as a 'fraud'.[11] Imposter syndrome occurs more frequently among women than men. Despite evidence to the contrary, a woman may develop the belief that she is not competent, or not considered competent. It was suggested that this may have become linked to performance evaluation in recent years and be related to the assertion that women are consistently under-evaluated by both women and men.[12] One of the best ways to combat this phenomenon is to know that other women experience it. Making sure that young female scientists have objective, validated confidence in their achievements is important, and alerting them to the existence of imposter syndrome, and other obstacles, can be very useful. Attributing success to one's own intelligence can be helped by consciously building one's own expertise and knowledge. One way of mitigating this is through mentorship programmes. Mentors can assist in reiterating the intrinsic value of the research being done and the real contribution it is adding to the field. The support that women need can often be found by working with other women in the same field. It was also noted that it is common for most women to put undue pressure on themselves to 'perform' in a perceived way.

Traditionally, men have engaged with the sciences across the entire socio-economic landscape. It must be recognised that in many countries, cultures and institutions, men are active advocates for the advancement of women. However, in many cases men still tend to act as gatekeepers, and, in some instances, are reluctant to accept contributions from female researchers. There is an urgent need for this situation to change. The balance in the gatekeeper positions will change when women are better represented in the senior levels of the sciences, and are seen and known as experts in their fields.

Two additional factors were discussed. In some cultures, it is possible that men are intimidated by women who are high achievers.[12] The second factor is the known perception that when a field becomes dominated by female researchers or scientists, the field becomes less prestigious for men.

According to Thébaud and Charles[6], deep-rooted gender stereotypes have interactional effects at both the individual and broader societal levels. At the societal and cultural levels this takes the form of overt and subtle biases, held by both men and women, that men and women should fit neatly into the popular notion that 'men are from Mars and women are from Venus' in terms of behaviour and career choices.[13] At the individual level, gender stereotyping shapes how women perceive themselves in terms of aptitude in studying mathematics and science at an early age and later on in their confidence to pursue, persist and thrive in STEMMI careers. Female researchers and scientists are encouraged to persevere, to stand out, and to make progress. At the same time, strategies and policy actions are required to ensure that girls and boys are exposed to the same STEMMI opportunities at school and that gender biases are eradicated. At institutional and organisational levels, policies and practices need to be reassessed from a gender perspective.[14] Increased financial investment in female participation in STEMMI is a step in the right direction, but concrete steps to create inclusive organisational and societal environments are needed to maximise returns on investments.[4,5]

Cultural perceptions and encouraging young girls to become scientists

The idea that young girls are discouraged from pursuing careers in the hard sciences, as well as the cultural perceptions about why it is better for women to marry and have children, still exist, and although times are changing, the process is slow. Families do invest in a daughter's education, but there is still a perception, often by extended family, that daughters' achievements are less worthy than sons' achievements. Young girls still grow up with an expectation that they should get married and have a family.

An example was described from traditional Zulu culture. Male children continue to be more highly valued than female children. Sons are expected to become providers in the homestead whereas daughters are viewed less favourably by their families because they are expected to marry into another family, where they will assume the role of caretaker of their future homes and in-laws. Therefore, characteristics such as the courage to pursue a high-impact career, independence and leadership are often not expected from daughters.[15] Family members, especially elder men, can sometimes be heard saying to female relatives who display such characteristics that they 'have manly courage' and it would have benefitted the family if they had been born male. Cultural gender stereotypes have been used to deny female children an education in many parts of the world. Ironically, regions which are strongholds of cultural gender stereotyping outperform egalitarian countries in terms of the proportion of women obtaining higher degrees in STEMMI fields and persevering in STEMMI careers.[15]

Another member of the panel shared her personal experience on the topic of culture. As a daughter in a single-parent household, her experience was that her mother was more than willing to invest in her education, but also had the strong expectation that she would also marry and be responsible for a household. Her experience is in line with the view expressed by Getachew et al.[7] that parents should reconsider their perceptions of psychic cost, in this case, their expectation that daughters should marry. This is a critical point of the human capital development of children in terms of parental gender bias. Although gender equality has come far, from these discussions it was clear that there is still a long way to go in terms of societal expectations.

The interactional processes that serve to stymie the full participation of women in STEMMI are often compounded for African women. Often, the twin subtle biases of race and gender are not openly discussed. However, it is a professional experience of African women that anything less than excellence is not enough to be placed on a par with one's peers as 'you are assumed to be incompetent until you can prove otherwise'.[6,10]

The need for women in the sciences

Women want the choice to work in the sciences, or, as it was expressed by a panel member, 'People want choices!'. In a recent study of the demand and supply of skills in South Africa[4], the authors could not elaborate on

trends of women's participation in the different industrial sectors because of the lack of historical gender disaggregated labour data. However, they noted that while the rate of female student enrolments at universities is surpassing that of male students, the proportion of male students graduating with STEMMI degrees was higher. Interestingly, at the time of writing this piece, a question was asked by the then CEO of the South African Institution of Civil Engineering: 'Given that money, time and resources are constrained, and evidence pointing to women being predisposed to caring and people careers, should we be investing so heavily in attracting women into STEM careers, specifically engineering, or should we invest in creating more gender-equal societies?'[16]. The answer to his question is an emphatic 'yes'. We should continue investing because people, irrespective of gender, want to have the freedom to choose. Men and women should have equal opportunities to pursue careers of their choice in order to empower them to contribute to their fullest capabilities. Women comprise half the population and continued financial investment and skills policy reforms are recommended to remove the societal – and institutionalised – discrimination that limits women from attaining positive higher educational outcomes in STEMMI, and subsequently in their participation and growth in the workforce.[13] The marginalisation of women amounts to the exclusion of half of the population's needs and forfeiture of the value thereof that could have been added.

The discussion reached the consensus that a woman who chooses to focus on her career should not be considered 'selfish'. Similarly, parental and household labour should be shared responsibilities and acknowledged as productive time, rather than assuming that women do not have the 'appetite for workload and extreme performance requirements'[12,16]. Female scientists and researchers need more female role models. These role models need to encourage their female counterparts to not feel guilty about making career choices; but the reverse also needs to be addressed. If a woman chooses household responsibilities and decides that she prefers to focus on her family, there should be no stigma attached to that choice. It is indeed about choices and the power to make the choice that fits an individual the best.

Marginalising the needs and expectations of women when constructing research strategy excludes half of the population's requirements and aspirations. The presence of women in the sciences, and in decision-making positions, can result in valuable input, and improve aggregate efficiency. As more women are marginalised, more productive investment opportunities will be forgone which can negatively impact aggregate income and welfare.

There is a need to assess the challenges that limit the number of women climbing the career ladder to positions of leadership, including leadership in STEMMI organisations.[9,11] An increase in female leadership would mean an increased availability of mentors, as well as advocacy against practices that may be hindering female participation within organisations.

Audience responses

A working mother related that she is often blamed and pressured into focusing on her family, and identified as guilty if mishaps occur within the family. On the other hand, if a man helps out in the same family, he is applauded for doing the unexpected. This discrepancy between expectations and rewards held by societies of the different roles and responsibilities between the genders within the household is frequently problematic for women. Many women may observe the need to be very persistent about their goals and their career expectations. Many young women take on the role of instigators of change, and it is partially up to this generation of young scientists to address these feelings of guilt. The remark was made that when a woman cries or feels guilty, then she is 'emotional', but when a man cries, he is perceived as 'caring'.

An audience member posed the question: 'If maternity is criticised, where will the next generation come from?'. The response shifted towards a question in behavioural economics: 'What is in it for men to share their power with women?'. It was stated that there are men who do not want their wives to have power within the household, but they do want their daughters to have more rights and power. This may create an incentive for men to relinquish their power.[17] In terms of the bargaining power

of women in the household, a new balance is emerging. The session ended with a call for more female scientists to be visible, and to continue fighting stereotypes and being persistent in their quest for equality.

As a closing thought, Figure 1 depicts wall art in a boardroom at a leading tertiary institution. What is of particular interest is that this boardroom is in the Faculty of Science and this photo was taken in 2018. When the figures are scrutinised, it is obvious that the images used to represent the sciences appear to be predominantly male. At the bottom of the wall, only one small figure can be identified as female, and even then she is depicted as a cheerleader. It is imperative that institutional leadership, as well as all members of society, become acutely aware of the importance of highlighting women in sciences with utmost urgency, because – *gender should never matter, we are all scientists here*.

Figure 1: Wall art in a boardroom of the Faculty of Science at a leading tertiary institution. The image was captured in 2018.

Acknowledgement

References

1. Parentwood [homepage on the Internet]. No date [cited 2018 Nov 22]. Available from: www.parentwood.co.za

2. Elsevier. Gender in the global research landscape: Research report. Amsterdam: Elsevier; 2017. Available from: http://elsevier.com/research-intelligence/resource-library/gender-report

3. United Nations Sustainable Development Goals [homepage on the Internet]. No date [cited 2018 Sep 13]. Available from: https://sustainabledevelopment.un.org/sdgs

4. Department of Science and Technology (DST). Human capital development strategy for research, innovation and scholarship. Pretoria: DST; 2016.

5. Department of Science and Technology (DST). Draft white paper on science, technology and innovation. Pretoria: DST; 2018.

6. Thébaud S, Charles M. Segregation, stereotypes and STEM. Soc Sci. 2016;7(7), Art. #111, 18 pages. https://doi.org/10.3390/socsci7070111

7. Getachew Y, Azomahou T, Yitbarek E. Share the love: Parental gender bias, women empowerment and intergenerational mobility [document on the Internet]. c2017 [cited 2018 Dec 10]. Available from: https://pdfs.semanticscholar.org/0f77/89b281695b28486289fd9578157356cb6618.pdf?_ga=2.3036661.459824647.1551349127-512262569.1550151602

8. Lubinski D, Benbow CP, Kell HJ. Life paths and accomplishments of mathematically precocious males and females four decades later. Psychol Sci. 2014;25(12):2217–2232. https://doi.org/10.1177/0956797614551371

9. Cook C, Diamond R, Hall J, List JA, Oyer P. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers [document on the Internet]. c2018 [cited 2018 Dec 10]. Available from: https://web.stanford.edu/~diamondr/UberPayGap.pdf

10. Bondarescu R, Balakrishna J, Moran CC, DeSilva A. Women in science: Surpassing subtle and overt biases through intervention programs [preprint]. arXiv 2018; #1808.05549. Available from: https://arxiv.org/abs/1808.05549

11. Clance PR, Imes SA. The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. Psychother Theor Res Pract. 1978;15(3):241–247. https://doi.org/10.1037/h0086006

12. Valian V. Why so slow? The advancement of women. Cambridge, MA: MIT Press; 1997.

13. Gray J. Men are from Mars, women are from Venus. New York: HarperCollins; 1992.

14. Dlamini NJ. The impact of the intersection of race, gender and class on women CEOs' lived experiences and career progression: Strategies for gender transformation at leadership level in corporate South Africa [doctoral thesis]. Pretoria: University of South Africa; 2013. http://uir.unisa.ac.za/handle/10500/13828

15. Stoet G, Geary DC. The gender-equality paradox in science, technology, engineering, and mathematics education. Psychol Sci. 2018;29(4):581–593. https://doi.org/10.1177/0956797617741719

16. Pillay M. Out on a rib. Civil Engineering Magazine. 2018 July;26(6):1.

17. Doepke M, Tertilt M. Women's liberation: What's in it for men? Quart J Econ. 2009;124(4):1541–1591. https://doi.org/10.1162/qjec.2009.124.4.1541

AUTHORS:
Irvy M.A. Gledhill[1,2] (iD)
Marie-Françoise Roy[3]
Mei-Hung Chiu[4]
Rachel Ivie[5]
Silvina Ponce-Dawson[6]
Helena Mihaljević[7]

AFFILIATIONS:
[1]School of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand, Johannesburg, South Africa
[2]DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa
[3]Rennes Mathematical Research Institute, University of Rennes 1, Rennes, France
[4]Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan
[5]Statistical Research Center, American Institute of Physics, College Park, Maryland, USA
[6]Department of Physics and IFIBA (CONICET), University of Buenos Aires, Buenos Aires, Argentina
[7]College of Engineering and Economics, University of Applied Sciences, Berlin, Germany

CORRESPONDENCE TO:
Irvy Gledhill

EMAIL:
Igle.Gledhill@wits.ac.za

A global approach to the gender gap in mathematical, computing and natural sciences: How to measure it, how to reduce it?

The participation of women in many fields of science remains a subject of concern worldwide. In this Commentary, we describe one of three collaborative projects – the Gender Gap project – funded by the International Science Council (ISC) and the 11 partners of the project. This project is the only one of the three that addresses an issue of such relevance for society. The objectives of the project are to provide evidence on which interventions can be based, and to make available material on best practice that has been proven by test. The project includes a joint global survey and a bibliometric analysis, both of which have an emphasis on comparing and contrasting results from less developed countries and from more developed countries.

Background

Mathematical and natural sciences have long traditions of women who have made significant contributions. However, participation by women remains small across the globe. While there are excellent studies, our data remain largely local, and much of the existing data are now out of date in the light of recent socio-political changes.

Two major surveys provide the foundation for this work. In physics, the International Union of Pure and Applied Physics (IUPAP) conducted a Global Survey of Physicists in 2009/2010.[1,2] The survey was conducted in eight languages, with 14 932 respondents in 130 countries. The results were first announced at the 4th IUPAP International Conference of Women in Physics[3], and have proved useful in guiding the choice of interventions appropriate in physics. In mathematics, a web-based bibliometric survey of publication patterns among men and women was carried out, using comprehensive metadata sources.[4] Among the findings was that over the years 1970 to 2010, the fraction of publications by women in top journals was of the order of 10% of total authorship.

The Gender Gap project

Since these surveys, much has changed. In geopolitical terms, the Arab Spring has affected academic environments, and anecdotal evidence indicates that the effects on women have been significant. The number of women in science is no longer rising in the European Union; and the number of bachelor's degrees in physics earned by women in the USA has fallen significantly since 2015.[5] Science funding has fallen in a significant number of countries, and science itself is subject to 'post-expert' public opinion. The education of girls is under serious threat in certain countries. However, there are many successful new initiatives in place, and many projects are addressing better science education and gender in science.

Therefore, a proposal was made to ISC to engage in a joint project with the following objectives. Firstly, as up-to-date evidence is needed, a global survey and a study of publication patterns are currently being conducted. Secondly, this joint project provides an excellent opportunity for collaborating with social scientists working in the fields of gender and science, noting that ISC was formed by the merger of the International Council of Science (ICSU) and the International Social Sciences Council (ISSC) in 2018. Thirdly, easy access to materials that have proved successful in encouraging women and girls in science will be provided, and fourthly, based on the evidence, successful practical policies and actions for reducing the gender gap will be recommended.

The project[6] is led by the International Mathematical Union, with the International Union of Pure and Applied Chemistry and IUPAP as executive partners. The other project partners are: IAU (Astronomy), ICIAM (Industrial and Applied Mathematics), IUBS (Biological Sciences), UNESCO (United Nations Educational, Scientific and Cultural Organization), IUHPST (History and Philosophy of Science and Technology), ACM (Computer Science), GenderInSite (Gender in Science, Innovation, Technology and Engineering) and OWSD (Organisation of Women for Science for the Developing World).

Global Survey of Mathematical, Computing and Natural Scientists

The survey has been developed and implemented through the American Institute of Physics Statistical Research Center as part of the collaboration of social scientists and natural scientists. It is probable that the results of the global survey will provide comparisons between disciplines, answering some persistent questions about why one discipline is generally more attractive than another to women. Contrasts and similarities will be sought between regions; countries that are very highly developed and less developed; employment in academia, industry and schools; and younger and older respondents. The survey addresses a series of experiences through sections that cover the development of interest in science, education and career experiences, family support for career choices, access to resources needed to conduct science, and opportunities to contribute to the scientific enterprise, for both women and men. It is designed to provide longitudinal data through compatibility with the Global Survey of Physicists of 2011.

This survey is based on a snowball sample, and is not designed to provide data on numbers or proportions of women among scientists. In this regard, it is complementary to the SAGA surveys[7] carried out through UNESCO, and use of both sets of results, where available, promises insightful conclusions.

The survey was translated, with the aid of scientists working with professional translators, into seven languages: English, French, Chinese, Japanese, Russian, Spanish and Arabic. The choice of languages was based both on those in use at the United Nations and on those used for the Global Survey of Physicists.
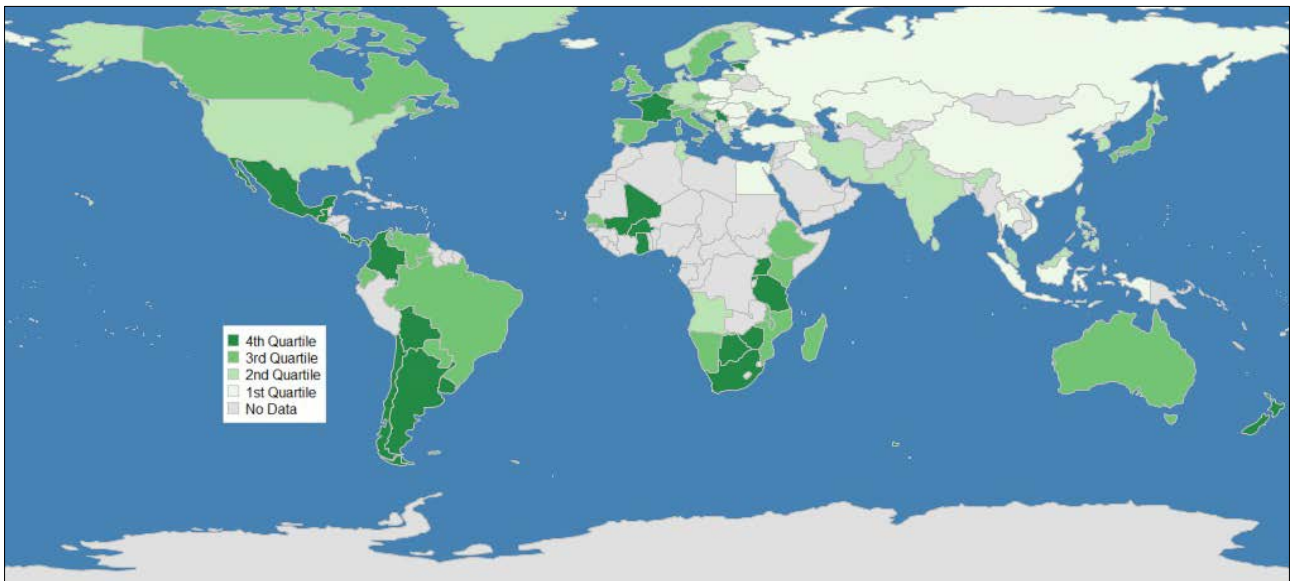
Figure 1: Proportion of responding scientists by country (1 November 2018).

Because multiple cultures are surveyed, three regional workshops were held: in Taiwan (National Taiwan Normal University, 7–8 November 2017), Colombia (Universidad de los Andes, 22–24 November 2017) and South Africa (African Institute of Mathematical Sciences, South Africa, 1–2 December 2017). The workshops' objectives were to inform diverse regional science communities about the project, present its three tasks, make contacts with people who are instrumental for the success of the project, and get input from Asia, Latin America and Africa to ensure the project reflects local realities. Additionally, these workshops were aimed at organising the dissemination of information about the project in Asia, Africa and Latin-America and to encourage the active participation of individuals and organisations of these regions. These surveys used special input from the three major regions of the ISC Regional Offices – Africa; Asia and the Pacific; and Latin America and the Caribbean – and were generously welcomed through these organisations. In each workshop, 30–40 participants could be invited. A principle of this project is that it includes men, both in the surveys and in the organisation and implementation of the project. The countries represented were chosen for the optimal representation of regions and cultures. In South Africa these were: Algeria, Burkina Faso, Botswana, Cameroon, Ethiopia, France, Kenya, Lesotho, Morocco, Madagascar, Malawi, Nigeria, South Africa, Swaziland, Uganda, USA and Zimbabwe; in Taiwan: Australia, China, France, India, Israel, Japan, Korea, Nepal, Malaysia, Taiwan, Thailand and USA; and in Colombia: Argentina, Brazil, Chile, Colombia, Costa Rica, Cuba, El Salvador, Mexico, Perú and USA.

Through consensus, debate and written responses, participants at each workshop worked hard to provide critical feedback for the survey creation team. Once completed, results from each of the workshops were organised and compared to determine the final survey instrument. In all cases, contributions from all attendees and insights from different disciplines and countries were shared, further helping refine the global survey to be culturally appropriate both in terms of language as well as in substance.

The Latin American workshop took place immediately after a workshop on professional skills for young people in science and engineering, with an exchange session to learn about the special needs of young people.[8] During the Asian workshop, participants emphasised the need to ensure that the questionnaire was applicable to individual respondents in industry. As a result, questions and response items were added that include patents and other measures of success in industry. Consequently, the language of the survey is relevant to industry-based participants, as well as academics and other professionals. During the African workshop, participants worked to ensure that the questions on career disruptions included the realities that scientists throughout Africa face. As a result, responses were expanded to include health, conflict,

natural disasters, and other continent-specific answers, and Arabic was added as a language.

The survey was released in May 2018 and can be accessed at https://gender-gap-in-science.org. The number of responses relative to the estimated number of scientists in each country is illustrated in Figure 1.

Analysis will take place at the American Institute of Physics Statistical Research Center in 2019. Preservation of personal confidentiality is a fundamental tenet, and the guiding principle is that analysis cannot be carried out down to samples with fewer than 30 respondents, or with fewer than 5 respondents in a cell.

Joint data-backed study on publishing patterns

The second task is a bibliometric study extended across disciplines, with an ongoing sustainable methodology designed to allow longitudinal studies and updates into the future.

A solid publication record is a key factor in a successful academic career. In mathematics, a recent study by data scientists Mihaljević-Brandt et al.[4] on publication patterns based on comprehensive metadata sources showed a systemic gender imbalance in the publication distribution of mathematicians. Using four decades of data, it was shown that the number of women mathematicians who are publishing has tripled since 1970, but that women publish less than men at the beginning of their careers, and leave academia at a higher rate. High-ranked journals publish fewer articles by women. Some show less than 5% authorships by women, with no change over time. Women published fewer single-authored papers, although their co-author networks are similar in size to those of men.

Ideally, for the purposes of this study, a volunteered identification of gender would be used. However, this is rarely available in the scientific literature. The methodology in automated studies relies on gender inference methods. The data science challenge is to estimate gender (in the present case confined to female, male or unknown) based only on an author's name string, which, in the case of academic papers, is one of the few gender indicators in the public domain. Gender inference application programming interfaces (APIs) are widely used by web services and rely on a collection of records from, for example, baby name lists, censuses and social media profiles. The tool returns the estimated gender with a weighted certainty of assignment. In the study quoted above, gender assignments were made with acceptable confidence for 61% of the authors identified. The scope of this research includes providing sufficient evaluation of the method used. In the first year of the project, the authors benchmarked and compared the gender

inference services available to the project, making this one of the few well-validated studies. In order to define benchmarks, error metrics and constraints have had to be devised.

This survey covers multiple languages, alphabets and cultures. An estimate of the origin of the name can be made, and confidence limits given, for African, Asian and European names, and names derived from these, using additional APIs, which are also undergoing evaluation for the present purpose.

The study of publishing practices in mathematics is now undergoing extension to astronomy, and will then be applied to physics, chemistry, biology and computer science. It is intended that the resulting engines could be automated and used in longitudinal studies in the future.

Database of good practices

The third task is to use the accumulated evidence to advise on initiatives which have a successful track record, taking regional and cultural differences into account. There are many initiatives across the world, but relatively few have undergone evaluation.

The task of finding published evaluated interventions aimed at increasing the participation of girls and women in science will lead to a database, which will be hosted by the International Mathematical Union and made available through the partners. It is a goal to add searches that make the material accessible. Databases have a limited life, although much effort is invested in them; therefore, guidance for the identification and evaluation of initiatives, and additions to the database, must be included.

This work is aimed both at young women, and at those who provide guidance to girls and young women, including parents and teachers – an important point in developing countries. Experience indicates that good information is needed, particularly by the family, about the stability of careers in science and the availability of jobs. The challenge of reaching parents is a significant one, and for that reason is rarely undertaken.

Conclusions

Knowledge of science and technology is universal, but it is shaped by local culture.[9] The Gender Gap project aims to provide a global set of data on experiences of both women and men in science, with comparisons across regions and for less developed and very highly developed countries.

The Gender Gap project was initiated in 2016 and will conclude in 2019. It is intended that the database of good practices and results of the survey and the publication study will be carried forward by all partners and ISC

in formulating coherent, sound, evidence-based initiatives, in which good intentions are guided by solid data, and in which regional and cultural relevance can be found.

Acknowledgements

References

1. Ivie R, Tesafaye C. Women in physics: A tale of limits. Physics Today. 2012;65(2):47–50. https://doi.org/10.1063/PT.3.1439

2. Ivie R, Tesfaye CL, Czujko R, Chu R. The global survey of physicists: A collaborative effort illuminates the situation of women in physics. In: AIP Conference Proceedings. 2013;1517:53–61. https://doi.org/10.1063/1.4794221

3. Proceedings of the 4th IUPAP International Conference of Women in Physics; 2011 April 5–8; Stellenbosch, South Africa. College Park, MD: American Institute of Physics; 2013.

4. Mihaljević-Brandt H, Santamaria L, Tullney M. The effect of gender in the publication patterns in mathematics. PLoS ONE. 2016;11, e0165367, 23 pages. https://doi.org/10.1371/journal.pone.0165367

5. American Physical Society [homepage on the Internet]. No date [cited 2018 Dec 10]. Available from: https://www.aps.org/

6. A global approach to the gender gap in mathematical, computing, and natural sciences [webpage on the Internet]. No date [cited 2018 Dec 10]. Available from: https://gender-gap-in-science.org/

7. Measuring gender equality in science and engineering. Working papers. Paris: UNESCO; 2016.

8. Meza Montes L, Ponce Dawson S. The Gender gap in mathematics, computing and natural sciences: An approach from Latin America. Mexico City: Mexican Physical Society; 2018.

9. Lin MT. Impact of science policy on the professional development for women in physics. Paper presented at: 6th IUPAP International Conference of Women in Physics; 2017 July 16–20; Birmingham, UK.

AUTHORS:
Heidi Prozesky[1,2] iD
Johann Mouton[1,2] iD

AFFILIATIONS:
[1]Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa
[2]DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Heidi Prozesky

EMAIL:
hep@sun.ac.za

A gender perspective on career challenges experienced by African scientists

Empirical knowledge of the career challenges that confront African scientists, and women scientists in particular, holds an important key to achieving future success in the science systems of the continent. In this article, we address a lack of evidence generally on the careers of scientists in Africa, by providing the first continent-wide description of the challenges they face, and how these challenges differ between women and men. Our analysis of questionnaire-survey data on approximately 5000 African scientists from 30 countries shows that women are not more challenged than men by a variety of career-related issues, with the exception of balancing work and family, which the majority of women, regardless of age and region, experience. Contrary to expectations, women are not only less likely than men to report a lack of funding as having impacted negatively on their careers, but have been more successful at raising research funding in the health sciences, social sciences and humanities. These results, as well as those from a comparison of women according to age and region, are linked to existing scholarship, which leads us to recommend priorities for future interventions aimed at effectively ensuring the equal and productive participation of women in the science systems of Africa. These priorities are addressing women's work–family role conflict; job security among younger women scientists; and women in North African and Western African countries.

Significance:
- This study is the first to describe, on a multinational scale, the career challenges that confront African scientists, and women scientists in particular.

- Contrary to expectations, we found that African women scientists do not report experiencing career challenges to a larger extent than men do, and have been more successful at raising research funding in three of the six major scientific fields.

- However, the findings highlight the significance of the challenge that balancing work and family poses to the majority of African women scientists.

Introduction

Our focus on gender in science in general and on publication output in particular is motivated by a general lack of empirical knowledge about the careers of scientists in Africa.[1] Very few studies have surveyed scientists across different African countries to gain insight into their career-related perceptions, and, to our knowledge, no study has attempted to understand the challenges that confront specifically African women scientists, such as the resources at their disposal (including non-material ones, such as mentorship and support), work–family balance, and mobility. Indications[2-8] are that patriarchy still pervades the majority of African societies, with its resulting gender based divisions of labour in both the family and scientific institutions, disparities between men's and women's access to power and resources, gender biases in rights and entitlements, etc.

These challenges and others have the potential to negatively affect women scientists' research performance and their retention, especially as emerging researchers, to the detriment of the research system that trained them. The contribution of this paper is to address this gap in the literature, and thereby to better understand what might be done to nurture the full potential of women scientists from 30 African countries. In addition to gender, our analysis also takes into account three other variables. Firstly, our focus on chronological age is informed by the continent wide challenge of retaining especially emerging researchers in science careers. Secondly, taking into account nationality recognises socio-cultural differences among countries that impact on the role and status of women. Some countries, such as Uganda, are considered more 'gender progressive'[8] than others, such as Tunisia[9], and these differences seem to be more significant than differences between institutional settings within any given country.[10] Thirdly, scientific fields differ in terms of the extent to which they are characterised by various forms of gender inequality[11], as well as many other aspects relevant to career challenges in general.

Methods

Data collection

A web-based survey was conducted between 2016 and 2017. The survey generated data from slightly more than 5000 scientists born and currently working in an African country. For the purpose of this survey, scientists are defined as individuals who dedicate at least a portion of their professional activity to research. As members of a scientific community, they communicate – primarily through peer-reviewed journal publications – their results and findings to their peers. Thus, to identify and contact African scientists, we extracted corresponding authors' emails from the Web of Science and Scopus databases for each article published from 2005 to 2015 with an institutional address in Africa. For Zambia, we also used articles in journals not indexed in the Web of Science and Scopus databases. Other sources of email addresses were the South African Knowledgebase database, the Internet, as well as snowball sampling.

Data were collected via a self-administered, structured questionnaire, which was piloted in Zambia, and translated into French for respondents in French-speaking countries. It comprised sections on educational background; employment; working conditions; research output; funding; career challenges; international mobility; collaboration; mentoring; and demographic background.

Prior to commencement of data collection, the study was approved by the Research Ethics Committee of Polytechnique Montréal (N/Réf: Dossier CÉR-1516-43) and by the Research Ethics Committee: Human Research (Humanities) of Stellenbosch University (Proposal #: SU HSD-002130). Informed consent was obtained from all respondents, whose participation in this study was voluntary. Respondents could decline to answer any question and could withdraw from the study at any time without negative consequences. All data collected were treated as confidential and the respondents' anonymity is protected in this publication.

Data processing and analysis

The focus of this study was on career challenges, i.e. the extent to which 10 identified factors may impact negatively on the scientists' careers. These factors range from personal (e.g. balancing work and family demands) to contextual (e.g. political instability or war) in nature. Originally measured with three response options ('Not at all'; 'To some extent'; and 'To a large extent') the variables were recoded into binary variables ('No' and 'Yes', with the latter including at least to some extent) for ease of comparison. In addition, a score of the original responses across the 10 challenges was calculated as a composite measure of the extent to which respondents experience challenges in their careers in general. This involved assigning numerical values to each response option ('0' to 'not at all', 1 to 'to some extent' and 2 to 'to a large extent') and adding these for the 10 challenges for each respondent. Scores therefore ranged from 0 to 20.

Where relevant, results on perceived challenges are accompanied by an analysis of related variables. These variables include reported average amount of research funding (in USD) received in the 3 years preceding the survey; average and maximum number of children and dependants; percentage of care-work and general housework undertaken; having studied or worked abroad during the 3 years preceding the survey; and holding a contract-based position.

Nationalities were recoded into four regions: Southern Africa (Botswana, Lesotho and South Africa); Western Africa (Benin, Burkina Faso, Cote d'Ivoire, Ghana, Mali, Nigeria, Senegal and Togo); North Africa (Algeria, Egypt, Morocco and Tunisia) and Eastern and Central Africa combined, because of the small numbers in these two (Cameroon, Central African Republic, DRC, Gabon, Ethiopia, Kenya, Madagascar, Malawi, Mauritius, Mozambique, Seychelles, Tanzania, Uganda, Zambia and Zimbabwe). Age was recoded into a binary variable to distinguish between young scientists (defined as 45 years or younger) and scientists older than 45. Scientific field was recoded into five broad categories: natural and agricultural sciences; engineering and applied technologies; health sciences; humanities; and social sciences.

Data were analysed with IBM SPSS Statistics 24. Subgroup comparisons were drawn between women and men in terms of challenges experienced, and their negative impact on career development, as well in terms of a number of other, related, variables, where relevant. The focus then shifted to an analysis of only women scientists, to determine whether the challenges they had experienced varied according to chronological age and nationality. Again, where relevant, other comparisons between these various categories of women were also drawn in terms of related variables. In the case of categorical variables, cross-tabulations were generated, and in the case of continuous variables, means (or medians when standard deviations were high) were compared.

Results

A comparison of women and men scientists

The greatest differences between female and male scientists were found with regard to the impact that they perceived a lack of funding for research equipment had had on their careers, with a much larger percentage of male respondents reporting a negative effect of this challenge (Table 1). This finding may be explained by the dominance of men in equipment-intensive scientific fields such as the natural sciences and engineering. Controlling for scientific field supports this explanation to some extent, as it reduces by half the difference between men and women in those scientific fields, but increases it quite substantially (from 14% to 25%) in the female-dominated social sciences which are less equipment intensive. Also, the second- and fourth-largest gender differences are observed for a lack of access to a library and/or information sources and a lack of research funding in general; and, in both cases, again a greater percentage of men reported the negative impact of these resource deficiencies.

Table 1: Percentage of men and women respondents reporting the negative effect, to some or a large extent, of career challenges

Challenge	Women	Men
Balancing work and family demands	80%	71%
Lack of research funding	80%	87%
Lack of mentoring and support	72%	71%
Lack of funding for research equipment	69%	83%
Lack of training opportunities	64%	70%
Lack of mobility opportunities	64%	69%
Lack of access to library and/or information sources	40%	55%
Limitation of academic freedom	40%	43%
Job insecurity	38%	37%
Political instability or war	29%	31%

It should be noted that men reported having received on average slightly less research funding (USD72 304) than women (USD75 691) did in the 3 years preceding the survey. However, when we controlled for field, a more differentiated picture emerged. In only three scientific fields – the humanities, social sciences and health sciences – women reported having received more funding than their male counterparts in those fields. In the other three scientific fields – engineering and applied technologies, the natural sciences and the agricultural sciences – the opposite applies (Figure 1).
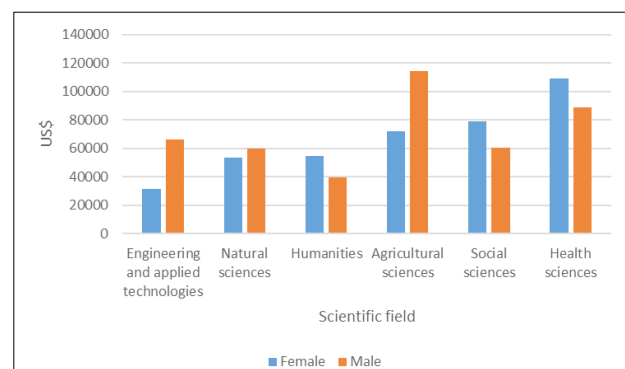
Figure 1: Reported amount of funding received, by gender and field.

The only challenge which women were more likely than men to have experienced concerns balancing work and family demands, for which the third-greatest gender difference is observed. Interestingly, men reported a larger average and maximum number of children and dependants than the women, but their care (and general housework) was not the men's main responsibility. On average, male scientists undertake a much lower percentage (37%) of such work themselves than is the case among women

(58%), while their partners contribute a much higher percentage (47%) than women scientists' partners do (23%) to care-work and general housework.

Among those for whom a lack of mobility and training opportunities have had a negative impact on their careers, we found a slightly greater percentage of men (although they are actually 10% more likely than women to have studied or worked abroad during the 3 years preceding the survey). Very small gender differences (3% or less) were found with regard to the remaining four challenges, i.e. a lack of mentoring and support, job insecurity, political instability or war, and a limitation in terms of academic freedom. In general, however, an analysis of a composite score of the responses across the 10 challenges shows that men reported having experienced, on average, challenges to a larger extent (mean = 8.56) than women did (mean = 7.71).

A comparison of women scientists in terms of age and nationality

When only women are analysed, balancing of work and family demands emerges as the challenge that has negatively affected the careers of the largest percentage (80%). A very similar majority (close to 80%) have experienced a lack of funding as a career challenge, and more than two thirds (69%) have experienced a lack of funding for research equipment specifically. However, this challenge does not seem to be a gender-related concern, as women are less likely than men to have experienced a lack of funding as a career challenge, and an exploratory survey[12] found this to be the most common challenge noted by 57 faculty members from 10 African countries (Botswana, Ethiopia, Ghana, Kenya, Nigeria, Rwanda, Senegal, South Africa, Tanzania and Uganda).

A lack of mentoring (72%) and a lack of mobility and training opportunities (64% in each case) had also been experienced by the majority of women. Poor mentoring and inadequate exposure to training have also been highlighted by qualitative research[4,6,13] among African women scientists. On the other hand, only approximately 40% of women indicated that a lack of access to library and/or information sources, a limitation in terms of academic freedom and job insecurity were major challenges to their careers. Only 29% experienced political instability or war as such.

In general, the composite score shows that women older than 45 seem to have experienced challenges to a lesser extent (mean = 7.06) than their younger counterparts (mean = 8.20). The age difference is most notable with regard to job insecurity: those who have experienced it as a career challenge are, on average, 6 years younger than those who have not. (Because of a high standard deviation in the case of age, the median rather than the mean is used as the measure of central tendency). This may be explained by the younger scientists being in the early stages of their careers. However, changes in the environment external to African research institutions – for instance volatile or uncertain economic conditions, or changes in government policy – may also lead institutions to cut costs by increasing their use of contract positions.[14] Without longitudinal data, a definitive answer in this regard is not possible. However, it is noteworthy that women 45 or younger are indeed more likely to hold contract-based positions (16%) than their older counterparts (10%).

It therefore also makes sense that younger women scientists are slightly more likely than their older counterparts to have experienced a lack of training and mobility opportunities as career challenges. At the same time, younger women have actually been more mobile: 26% have worked or studied abroad in the 3 years preceding the survey, compared to only 18% of those older than 45. No age difference was found with regard to having experienced a lack of funding in general as a challenge. However, this is not reflected in the actual amount of funding received in the 3 years preceding the survey, which was much higher for women older than 45 (USD94 443) than for those 45 or younger (USD60 551). When we controlled for scientific field, we found that this applies to all fields except the social sciences (Figure 2).
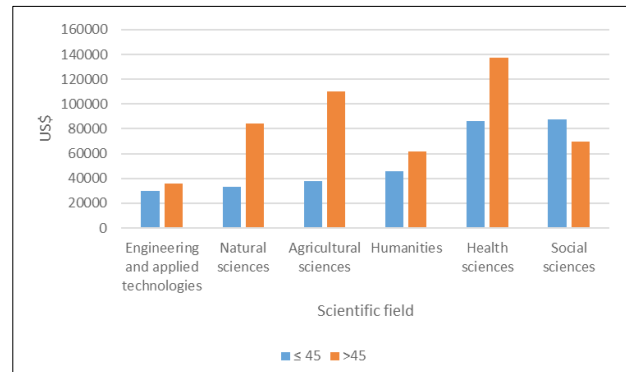

Figure 2: Reported amount of funding received by women older than 45 years and those 45 years or younger, by field.

On average, older women (but not by much) reported finding balancing work and family demands as significant career challenges. This was despite the fact that these older women had, on average, fewer dependants younger than 18 and were very similar to younger women in terms of the distribution of care-work and general housework between themselves and their partner. However, older women did have, on average, more elderly dependants in their care than their younger counterparts. It also needs to be taken into account that respondents were asked to reflect on challenges they had experienced in the past. Thus, although older women were past their intensive child-bearing and child-rearing years, certain opportunities (such as scholarships for PhD training) that they had to forego during that period have age limits, and they were no longer legible.[6]

When region of nationality is considered, the most salient pattern is that North African women scientists are more likely to have reported the negative effects of 6 of the 10 challenges on their careers: a lack of (1) mentoring and support, (2) mobility opportunities, (3) training opportunities, (4) research funding and (5) funding for research equipment; and (6) a limitation in terms of academic freedom. The composite score supports the observation that North African women scientists have experienced more and/or greater challenges than women in the other regions, and especially when compared to those in Southern Africa (Figure 3).

With regard to funding in particular, the data on the amounts received in the 3 years preceding the survey match the experience of North African women: they received by far the lowest amount at USD13 331, compared to USD57 613 for Western Africa, USD99 900 for Southern Africa, and USD203 093 (15 times more) for Eastern and Central Africa. With regard to mobility, Western African women were relatively unlikely (18%) to have worked or studied abroad recently (only Southern African women were found to be less mobile, at 16%).
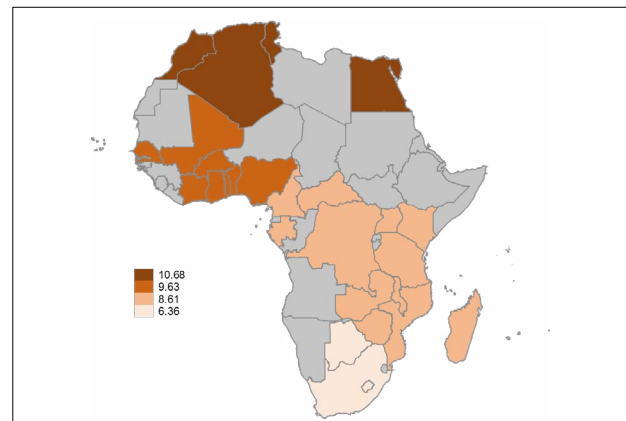

Figure 3: Women respondents' average career-challenge score, by region, from lowest to highest: Southern Africa, Eastern and Central Africa, Western Africa, and North Africa.

Furthermore, Western African women were most likely to highlight the balancing work and family demands as a career challenge. We also found them to have the highest average number of dependants older than 5, and the second-highest average number of dependants younger than 6. This finding resonates with Tsikata's[6] observation that, '[i]n keeping with the pro-natalist national cultures of Ghana, women academics are expected to marry and have children'. Interestingly, however, when compared with women from the other regions, Western African women did the smallest percentage of housework themselves. As Tsikata[6] notes with regard to Ghana, women scientists:

> differed in how they experienced the impacts of marriage and family on their work. The ages of their children, their marital status and who they were married to, were all important factors.

The women scientists most burdened by housework and caregiving were found in North Africa, which is explained, at least with reference to Tunisia, by Hassine's[9] observations that:

> academic women, with few exceptions, have internalised the traditional values that define gender roles and continue to prevail in Tunisian society and are even on the increase with the rise of Islamism.

Women from Western Africa were most likely to have experienced a lack of access to library and/or information sources, and political instability or war. The remaining challenge – job security – was most challenging for the careers of Central and Eastern African nationals.

The largest regional difference was observed with regard to a lack of access to library and/or information sources, which was a career challenge for only less than 20% Southern African women respondents compared to three-quarters of Western African women respondents. A lack of funding for research equipment and of training opportunities to develop professional skills also differed quite substantially between the regions, with Southern African women again less likely than their counterparts from other regions, especially Western Africa, to have experienced these challenges. In fact, the only challenges Southern African women were more likely to have experienced were those relating to job security and balancing work and family demands. Women across all the regions, however, were by far the most similar in terms of their experience of balancing work and family demands, with very high percentages (76–85%) citing this as a career challenge.

Conclusions

The large body of literature on women scientists leads one to expect that these women experience challenges to a larger extent than their male counterparts do. Our survey of African scientists does not, however, support this general hypothesis. The notable exception is the challenge of balancing work and family demands, which is the only challenge women are more likely than men to have experienced. It is also the challenge most frequently experienced by women, particularly those women who are older than 45 (who have a greater number of elderly dependants), and those from Western African countries.

A number of scholars[4,6,8,9,15-20] have highlighted the potential negative effects of particularly African women scientists' reproductive responsibilities and a traditional gendered division of labour within households. It has often been noted that many women scientists globally are limited in their geographic mobility by family demands and the occupational contingencies of their husbands (for a review see Prozesky[21]). Female scientists in Africa are reported to experience difficulties travelling to conferences, for example, because of the assumption that they are the primary domestic caregiver at home.[4,22] In addition, in some countries (e.g. Sudan) restrictions on women's interactions with non-familial men and norms inhibit their movement outside of the local area.[8,23]

These challenges may very well explain why the women we surveyed had been less mobile recently than the men, and that almost two-thirds of the women had experienced a lack of mobility as a career-related challenge. Campion and Shrum[22] refer to this scenario as 'educational

and research localism', which leads to restricted professional networks. What is surprising, however, is that women were less likely than men to report a lack of mobility as a challenge they had experienced. A possible explanation that should be considered for further research may be found in Stouffer et al.'s[24] concept of relative deprivation: because of the normative restrictions of women's geographic mobility noted above, their expectations of mobility are arguably lower than men's. Differences between women and men in terms of their expectations have been found to apply to other work-related aspects, such as the paradox that women report higher levels of job satisfaction than do men, although, by most objective standards, women's jobs are worse than men's.[25]

Our survey shows that women's greater likelihood to experience work–family role conflict is not a function of the number of dependants they have (which is greater for men), but it corresponds well with the findings of previous research outside of Africa (and reviewed elsewhere – see Prozesky[21]) that women scientists reported having fewer children than their male counterparts did. In Africa, women who do enter scientific careers are more likely to postpone or avoid family responsibilities than men[22], but when they do have children, they take the main responsibility for 'traditional obligations' of childcare and housework, while men tend to delegate these responsibilities to their (female) partners[8,26].

However, contrary to the literature – in particular the 'deficit model' which hypothesises that in patriarchal societies men and women do not share equally in the means of scientific production[27] – we find that African women scientists are not only much less likely than men to have experienced financial resources as a career challenge, they have also received more research funding than men in the humanities, social sciences and health sciences.

However, funding is still highlighted as a career challenge by the second-largest majority of women scientists. The need for funding for research equipment is most salient among North African women (particularly when compared to Southern African women) and among those in the natural and agricultural sciences (particularly when compared to their counterparts in the less resource-intensive humanities and social sciences).

Younger women are not more likely than their older counterparts to have experienced such financial-resource deficits as a career challenge, but they do, on average, obtain less funding than their older peers, except in the social sciences. They are also much more likely to have experienced job insecurity as a career challenge and in general have experienced challenges to a larger extent than their older counterparts.

In terms of region, the 'most career-challenged' women are found amongst North African nationals. A more detailed analysis shows that Western African women are particularly likely to report a lack of material and other resources (library and/or information sources, funding for research equipment and training opportunities to develop professional skills), especially when compared to their Southern African counterparts.

To policymakers and granting councils who are tasked with designing research-support programmes to optimise the performance of women scientists, our results are in agreement with Campion and Shrum's[22] argument that gender differences in research systems 'are rooted in systemic inequities in social rather than material resources'. This seems to suggest that increasing funding for women scientists (the proverbial 'throwing money at the problem') should be less of a priority than work–family role conflict. The latter is a much more persistent challenge that is specific to women scientists in societies characterised by traditional gendered division of labour within households. It should be recognised, as Tamale and Oloka-Onyango[8] observe, that 'the roots of patriarchal oppression lie in the smallest unit of societal organisation which is the family', and therefore 'the root causes of inequities' within science are 'based on the underlying gendered division of labour'.

In addition to the need for a fundamental, although complex, change in gender relations which would 'allow men to share in both the pleasures and the burdens of time-consuming domestic and parenting responsibilities'[20], research institutions should also provide allowances in their policies or employment contracts for women in the role of caregiver (e.g. maternity, child-care and domestic-support provisions, as well as family-

responsibility leave). In short, institutions cannot remain ostensibly gender neutral while critical differences exist between men and women in African societies, especially those that have experienced decreased access to social services.[8]

Job security among younger women scientists is also highlighted as an area of concern. Finally, from a regional perspective, efforts to address women scientists' career-related challenges should be directed first and foremost towards North African and Western African countries.

Acknowledgements

Authors' contributions

The study which produced the data analysed for this article was primarily conceptualised and methodologically designed by J.M., with H.P. providing some input during proposal writing. H.P. designed the data collection instrument, while J.M. managed the online questionnaire administration. Data analysis and interpretation was conducted by H.P., as was preparation and creation of the article, specifically writing the initial draft. J.M. provided critical review, commentary and revision of writing of that draft, which required additional analysis. Project leadership and management, as well as funding acquisition, were the responsibilities of J.M.

References

1. New Partnership for Africa's Development Planning and Coordinating Agency. African Innovation Outlook 2014 [document on the Internet]. c2014 [cited 2018 Aug 01]. Available from: https://www.nepad.org/publication/african-innovation-outlook-ii

2. African Development Bank. African gender equality index 2015 – empowering African women: An agenda for action [document on the Internet]. c2015 [cited 2018 Aug 01]. Available from: https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/African_Gender_Equality_Index_2015-EN.pdf

3. Olaogun JA, Adebayo AA, Oluyemo CA. Gender imbalance in the academia in Nigeria. Eur Sci J. 2015;Nov:294–306.

4. Akinsanya OO. The role of women in academics: Issues, challenges and prospects. J Res National Dev. 2012;10(2):136–141.

5. Guramatunhu-Mudiwa P. Addressing the issue of gender equity in the presidency of the university system in the Southern African Development Community (SADC) region. Forum on Public Policy. 2010;2010(2):1–20.

6. Tsikata D. Gender, institutional cultures and the career trajectories of faculty of the University of Ghana. Feminist Africa: Rethinking Universities. 2007;1(8):26–41.

7. Zewotir T, Maqutu D. Gender differences in attributions of success by academics in Africa: A study of the National University of Lesotho. Afr Insight. 2006;36(1):35–40.

8. Tamale S, Oloka-Onyango J. Bitches at the academy: Gender and academic freedom at the African university. Afr Develop. 1997;22(1):13–37.

9. Hassine OKB. Personal expansion versus traditional gender stereotypes: Tunisian university women. In: Buskens I, Webb A, editors. Women and ICT in Africa and the Middle East: Changing selves, changing societies. London: Zed Books; 2014. p.81–95.

10. Chakravarthy R, Chawla A, Metha G. Women scientists at work: An international comparative study of six countries. Scientometrics. 1988;14(1–2):43–74. https://doi.org/10.1007/BF02020242

11. Stack S. Gender and scholarly productivity: The case of criminal justice. J Crim Just. 2002;30(3):175–182. https://doi.org/10.1016/S0047-2352(01)00134-9

12. Darley WK, Luethge DJ. The role of faculty research in the development of a management research and knowledge culture in African educational institutions. Acad Manag Learn Educ. 2016;15(2):325–344. https://doi.org/10.5465/amle.2013.0337

13. Udegbe IB. Preparedness to teach: Experiences of the University of Ibadan early career academics. Stud High Educ. 2016;41(10):1786–1802. https://doi.org/10.1080/03075079.2016.1221656

14. Subotzky G. Addressing equity and excellence in relation to employment in higher education. EduSource Data News. 2003;38(January):1–13.

15. Arthur P, Arthur E. Tertiary institutions and capacity building in Ghana: Challenges and the way forward. Commonw Comp Polit. 2016;54(3):387–408. https://doi.org/10.1080/14662043.2016.1175690

16. Raburu P. Motivation of women academics and balancing family & career. J Educ Soc Res. 2015;5(1):359–370. https://doi.org/10.5901/jesr.2015.v5n1p359

17. Tettey W. Challenges of developing and retaining the next generation of academics: Deficits in academic staff capacity at African universities [document on the Internet]. c2010 [cited 2018 Aug 01]. Available from: http://www.foundation-partnership.org/pubs/pdf/tettey_deficits.pdf

18. Anagbogu MA, Ezeliora B. Prospects and problems of Nigerian women in science and technology for national development. Ghana J Dev Stud. 2008;5(2):17–26. https://doi.org/10.4314/gjds.v5i2.35072

19. Muula AS. Status of scholarly productivity among nursing academics in Malawi. Croat Med J. 2007;48(4):568–573.

20. Mama A. Restore, reform but do not transform: The gender politics of higher education in Africa. J High Educ Afr. 2003;1(1):101–125.

21. Prozesky HE. Gender differences in the publication productivity of South African scientists [doctoral dissertation]. Stellenbosch: Stellenbosch University; 2007.

22. Campion P, Shrum W. Gender and science in development: Women scientists in Ghana, Kenya, and India. Sci Technol Hum Val. 2004;29(4):459–485. https://doi.org/10.1177/0162243904265895

23. Miller BP, Shrum W. Isolated in a technologically connected world: Changes in the core professional ties of female researchers in Ghana, Kenya, and Kerala, India. Sociol Quart. 2012;53(2):143–165. https://doi.org/10.1111/j.1533-8525.2012.01229.x

24. Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM Jr. The American soldier. Princeton: Princeton University Press; 1949.

25. Clark AE. Job satisfaction and gender: Why are women so happy at work? Labour Econ. 1997;4(4):341–372. http://dx.doi.org/10.1016/S0927-5371(97)00010-9

26. Tamale S. Taking the beast by its horns: Formal resistance to women's oppression in Africa. Afr Develop. 1996;21(4):5–21.

27. Sonnert G. Women in science and engineering: Advances, challenges and solutions. In: Selby CC, editor. Women in science and engineering: Choices for success. New York: The New York Academy of Sciences; 1999. p. 34–57. https://doi.org/10.1111/j.1749-6632.1999.tb08353.x

AUTHORS:
U. Marie Clémentine Nibamureke[1]
Irene E.J. Barnhoorn[2]
Gesina M. Wagenaar[1]

AFFILIATIONS:
[1]Department of Zoology, University of Johannesburg, Johannesburg, South Africa
[2]Department of Zoology, University of Venda, Thohoyandou, South Africa

CORRESPONDENCE TO:
Ina Wagenaar

EMAIL:
inaw@uj.ac.za

Assessing the potential effects of nevirapine in South African surface water on fish growth: A chronic exposure of *Oreochromis mossambicus*

Aquatic environments around the world have become mixtures of different types of pollutants, including pharmaceuticals. The presence of pharmaceuticals in aquatic environments has raised concerns regarding the possibility of unintended effects on aquatic animals. South Africa is currently the largest consumer of HIV antiretroviral drugs (ARVs) worldwide. Nevirapine (NVP), a first-line ARV, has been associated with serious liver toxicity in humans and has been repeatedly detected in South African surface water. We investigated the potential effect of NVP on the growth of larvae and juveniles of the Mozambique tilapia (*Oreochromis mossambicus*) through a chronic laboratory exposure. Larval and early juvenile stages were exposed to the highest reported environmental relevant concentration of NVP in South African surface water (1.48 μg/L) for 60 days in a static renewal system. Body mass and total length measurements were recorded and analysed for individuals aged 1, 5, 30 and 60 days. In total, 455 fish were assessed. The growth parameters of larvae exposed to NVP were not statistically significantly different ($p > 0.05$) from those of control larvae. However, the juveniles exposed to NVP showed a slightly lower mean growth rate between the 30th and 60th day compared with the control fish. These results suggest that the concentration of NVP in South African surface water has no significant detrimental effects on fish growth during the first 2 months of their life. Further studies to investigate the effects on all life stages of fish are needed as it is evident that the growth rate of exposed fish could be affected after this stage.

Significance:
- This study was the first to investigate the effect of an antiretroviral drug in surface water on fish growth.
- Chronic exposure to the highest environmentally relevant concentration of nevirapine in South African waters did not affect the growth of early life stages of Mozambique tilapia.
- The levels of antiretrovirals in aquatic systems should be monitored closely as their consumption is likely to increase in the future.

Introduction

The presence of human pharmaceuticals in aquatic environments around the world is considered to be one of the pollution problems of emerging concern as the effect of these pharmaceuticals on aquatic animals is still largely unknown.[1,2] There are growing concerns that pharmaceuticals in water may have unintended effects on fish.[3,4] Case studies include the collapse of a fish population exposed to a synthetic oestrogen in Canada[5]; reduced hatching success of zebrafish (*Danio rerio*) embryos exposed to sulfonamides[6]; histological changes in the organs of rainbow trout (*Oncorhynchus mykiss*) and common carp (*Cyprinus caprio*) exposed to diclofenac[7]; and liver damage and altered health in Mozambique tilapia (*Oreochromis mossambicus*) exposed to efavirenz, an antiretroviral (ARV)[8]. It is evident that pharmaceutical pollution in aquatic environments needs to be investigated thoroughly as pharmaceuticals are important in our everyday life to prevent and treat diseases in humans, animals and crops, leading to increased economic income of households and countries.[1]

Consumption of pharmaceuticals around the world is diversified, and, therefore, the types and concentrations of pharmaceuticals in surface water differ from one region to another.[9] South Africa is amongst the countries with the highest HIV prevalence in the world, with approximately 7.9 million people infected.[10] More than 60% (4.4 million) of people living with HIV are receiving ARVs – making South Africa the biggest ARV treatment programme in the world.[10,11] Studies in different countries have showed that the current technologies used to treat waste water in wastewater treatment plants do not remove, or only partially remove, some of the pharmaceuticals, including ARVs.[12-14] In addition, during the disinfection processes, some pharmaceuticals form compounds which can be more active than their parent compound.[15] It is therefore not a surprise that South African surface water contains all types of ARVs used in the country, and that these ARVs have also been detected in tissues of fish.[16,17] Currently, little information is available on the effects ARVs may have on fish health.

Nevirapine (NVP), a first-line ARV prescribed in South Africa, has been repeatedly detected in surface water around the country[16], because it is resistant to biodegradation[18] and because the chlorination process used in wastewater treatment plants in the country results in the formation of various NVP degradation products, some of which have similar antiviral activity to the parent compound[15,19]. NVP is known to cause serious undesirable side effects in humans including liver toxicity and skin rash.[20] Being continuously added to surface water as a result of daily consumption, NVP as well as all other ARVs and pharmaceuticals in general, are constantly present in surface water and available to fish living in the water.[21] This situation raises concerns about the possible effects ARVs and NVP in particular may have on fish health.

Human pharmaceuticals are highly active chemical compounds designed to have specific physiological effects on target tissues.[1] Therefore, serious concerns arise when non-target organisms such as fish come into contact with such highly active compounds which were not intended for them. Nobody can predict with certainty what effects these compounds may have on fish. Recent laboratory exposure studies have shown that some of those

pharmaceuticals can disturb fish physiology and metabolism.[22-24] Some studies have also emphasised that early life stages of fish may be more vulnerable to pollutants in water which can impede their growth and development and these pollutants could, therefore, negatively affect fish communities and populations.[25]

Fish growth is determined amongst others by the type and quantity of food, the physicochemical water quality and the competition among individuals.[26] When conditions become unfavourable, for example, in the case of pollution or food scarcity, fish will show impaired growth.[27] In fisheries, mass and length measurements are important growth parameters which are recorded frequently to calculate mass–length relationships to estimate the well-being and age of the fish.[28,29] The mass–length relationship of a healthy fish should show a strong positive correlation as the fish is growing.[29]

One of the most used mass–length relationships in fisheries is the condition factor (CF) which is expressed as the ratio between fish mass and length.[28-30] Researchers in fisheries using different fish species have shown that the CF of a healthy adult fish is close to 1.[28-30] Pharmaceuticals which may affect fish growth by disturbing metabolism, may lead to reduced growth which may be noticed through changes in growth parameters.[30] Potentially detrimental effects of pharmaceuticals of concern on aquatic organisms should be determined, and effects of individual pharmaceuticals should be assessed on individual species to determine specific effects.[31]

We report on the potential effects of NVP on the growth of larvae and juveniles of the southern African indigenous freshwater fish species *Oreochromis mossambicus* through a laboratory chronic exposure.

Methods

Ethical clearance

This study was approved by the University of Johannesburg's Ethical Committee on 20 November 2015 (ref no. 201242617).

Breeding of fish

Sexually mature *O. mossambicus* were chosen from the University of Johannesburg's aquarium stock and organised into two breeding sets (comprising one male and two female individuals) in two 700-L glass tanks. The experiment was conducted in an environment-controlled room: the temperature of the water was kept at 27 $\pm$ 1 °C, a photoperiod cycle of 14 h light:10 h dark was maintained, and an oxygen line was connected to each tank (providing $\geq$ 80% dissolved oxygen). The fish were inspected every morning for possible eggs, and once eggs were available, they were collected from the mouth of the female fish approximately 24 h after fertilisation and placed in 1-L glass bottles fitted with small hatching jars.[32]

Exposure conditions and procedures

The experiment was conducted four times and included three groups: control group with dilution water only; the solvent control group (<0.01% v/v dimethylsulfoxide or DMSO in water) and the NVP group (1.48 μg/L[16]). The exposure started with eggs in 1-L glass jars containing the exposure media. The start of the larval stage was recorded as the first day after hatching.[33] On the fifth day after hatching, the larvae were transferred to 20-L glass tanks where the exposure continued until they were 30 days old. Then they were transferred to 90-L tanks and the exposure continued until they were 60 days old. The transfer to bigger tanks was done in order to give the fish more space for adequate growth and development.[33] As the exposure lasted 60 days, exogenic feeding started when the yolk sac was almost depleted, which was approximately 12 days post-fertilisation. The start of exogenic feeding also marked the end of the larval stage and the start of the juvenile life stage.[33,34] The fish were fed three times per day with complete tilapia fry crumble #1 (500–750 μm) (Tilapia Fry Crumble, AVI Products (Pty) LTD 2001/015923/07, Durban, South Africa).

The experiment was conducted in the same climate-controlled room as the breeding to avoid exposure of the embryos to different temperatures; the temperature in the different exposure containers was kept at 27 $\pm$ 1°C and a 14:10 h light:dark photoperiod was maintained. Oxygen was supplied and kept at $\geq$ 80%. Each exposure container was checked twice

daily to remove organic waste including dead embryos/larvae if any, food remains and excretion. The pH in each container was monitored daily and maintained between 7.6 and 8.1.

At least 10 individuals aged 1, 5, 30 and 60 days old were sampled per group by placing them on ice for 3–5 min, which served as an anaesthesia. Then the total length and wet mass of each individual were measured using a small graduated board (in mm) under a dissecting microscope and an electronic scale accurate to 0.001 g. The sampling of specimens was done randomly using a plastic pipette with the tip cut to make a $\pm$3-mm opening for 1- and 5-day-old larvae, and a small net for 30- and 60-day-old juveniles. After taking the measurements, the fish were killed and fixed in 10% neutral buffered formalin for further analyses. The CF of each fish was calculated from the recorded mass and length measurements following Carlander[35]. The age of the fish was recorded and expressed in days post-hatching, with Day 1 being the first day after hatching.

$$CF = \frac{\text{Total body mass (g)} \times 10^5}{\text{Total body length (mm)}^3}$$

The mean specific growth rate (SGR) which expresses the mean daily mass gain was also calculated in % for each group for 30- and 60-day-old juveniles using the following formula[36]:

$$SGR = \frac{\text{LnFinal mass - LnInitial mass}}{T_2 - T_1} \times 100,$$

where Ln represents the natural logarithm while T_1 and T_2 represent the days of initial and final mass recordings, respectively.

Water physical and chemical parameters

The physical parameters of the water were monitored daily. Every 96 h the exposure medium was renewed (to a half) to maintain NVP concentration and water quality as per recommended guidelines.[37]

The exposure media were sampled for instrumental analysis to determine the concentration of NVP. The analyses were done by an ISO 17025 accredited laboratory using ultra-high-pressure liquid chromatography (UPLC) coupled to quadrupole time-of-flight mass spectrometry (QTOF/MS).[38]

Statistical analysis

The IBM SPSS Statistics software (version 24) was used for statistical data analysis with a significance level of $p < 0.05$. Data were checked for normality of distribution using the Shapiro–Wilk test; the Spearman correlation coefficient moment (rho) was used to test the strength and direction of association between length and mass measurements; and non-parametric Kruskal–Wallis and Mann–Whitney U tests were used to test the significance of differences between the NVP-exposed fish and control fish.

Results

The analysis of test media samples showed that NVP concentrations never fell below 80% of the nominal concentration. NVP was below the detection limit in both the control and solvent control media.

At the end of the exposure, the mean survival rate of 60-day-old fish was above 78% in all the groups (Figure 1). No statistically significant difference in survival rate was found between the NVP-exposed fish and the control fish ($p = 0.683$). The length and mass measurements of larvae in all exposure containers and groups were very similar for 1- and 5-day-old fish, but 30- and 60-day-old fish showed variations in growth rate in the same container and in all groups. The mean length and mass, respectively, of 60-day-old juveniles were 26.5 $\pm$ 8 mm and 0.410 $\pm$ 0.30 g for the control group; 26 $\pm$ 6.1 mm and 0.336 $\pm$ 0.24 g for the solvent control fish; and 25.1 $\pm$ 7 mm and 0.319 $\pm$ 0.33 g for the NVP-exposed fish (Figures 2 and 3). The mean length and mass from Day 1 to Day 60 are presented in Table 1. The largest and heaviest 60-day-old juvenile fish was from the NVP-exposed fish with a total length of 49 mm and mass of 1.976 g; and the smallest fish was from the control group with a total length of 13 mm and mass of 0.032 g. Statistical comparison of length and mass measurements of larvae and juveniles in the NVP-exposed group with those from the control and solvent groups showed no

statistically significant difference (p=0.995 and p=0.808, respectively). As the fish aged, they were also growing in length and gaining mass. To assess the strength of the relationship between length and mass measurements in all three groups, the Spearman correlation coefficient rho was used. A strong positive and significant (p=0.0001) correlation was found between length and mass measurements from Day 1 to Day 60 post-hatching in all the groups: control group rho=0.958, solvent control group rho=0.955 and nevirapine group rho=0.949 (Figure 4). The relationships length–age and mass–age were also strong, positive

and significant (p=0.0001) in all groups, respectively: 0.883 and 0.691 for the control fish; 0.917 and 0.696 for the solvent control fish; and 0.890 and 0.574 for the NVP-exposed fish. The length–age relationship was stronger than the mass–age relationship in all groups. A standard multiple regression showed that in all groups, although both length and mass could be used to predict the age of the fish, length was a better predictor than mass, with adjusted R^2=0.862 for the control group (beta=1.54, p=0.0001), R^2=0.918 for the solvent control group (beta=1.447, p=0.0001) and R^2=0.872 for the NVP group (beta=1.303, p=0.0001). The calculated mean CF for the 60-day-old fish was 1.8±0.6, 1.7±0.1 and 1.7±0.8 for the control, the solvent control and NVP-exposed fish, respectively (Figure 5).

Statistical analysis showed no significant differences between the NVP-exposed fish and both control groups (p=0.427). The mean CF values for the fish in each group are presented in Table 1.The mean daily specific growth rate (SGR) of the fish was calculated as a percentage using mass measurements from Day 1 to Day 30 and from Day 30 to Day 60 post-hatching. SGR shows the mean percentage increase in body mass per day. The SGRs for the first month were 8.4±1.6%, 8.4±1.5% and 8.9±2.3% for the control, solvent and NVP-exposed fish, respectively, while for the second month they were 7.5±0.7%, 7.2±0.8% and 6.4±0.7% (Table 2).

The SGR for NVP-exposed fish between the 30th and 60th day was noticeably lower than that of the control fish, but statistical analyses showed no significant differences in SGR between the three groups (p=0.138).

Figure 1: Survival rate of juvenile fish from the three exposure groups on Day 60 post-hatching.
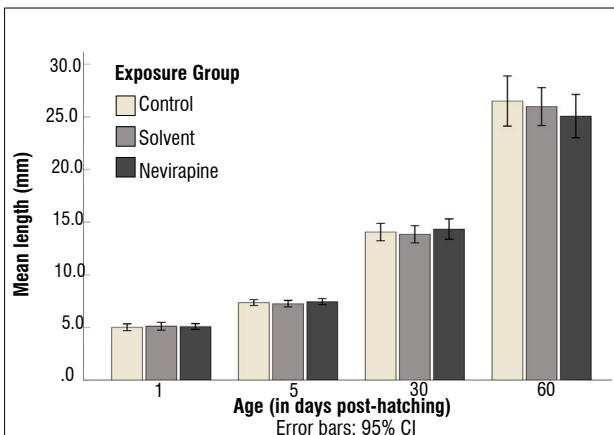
Figure 2: Mean length of fish in the three exposure groups on Days 1, 5, 30 and 60 post-hatching.

Figure 3: Mean body mass of fish in the three exposure groups on Days 1, 5, 30 and 60 post-hatching.

Figure 4: Scatterplot showing the strong positive correlations (upward position from left to right) between mean length and mass measurements of fish in all three exposure groups from Day 1 to Day 60.

Figure 5: Mean condition factor of fish in all three exposure groups from Day 1 to Day 60 post-hatching.

Table 1: Mean total length, mass and condition factor of control fish exposed to borehole water (Control) or the solvent dimethysulfoxide (DMSO) and of nevirapine-exposed fish

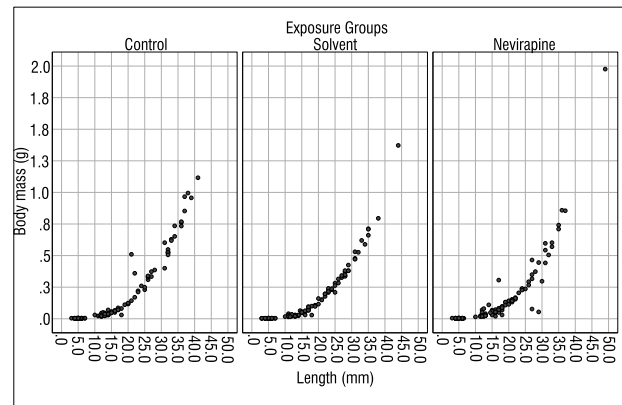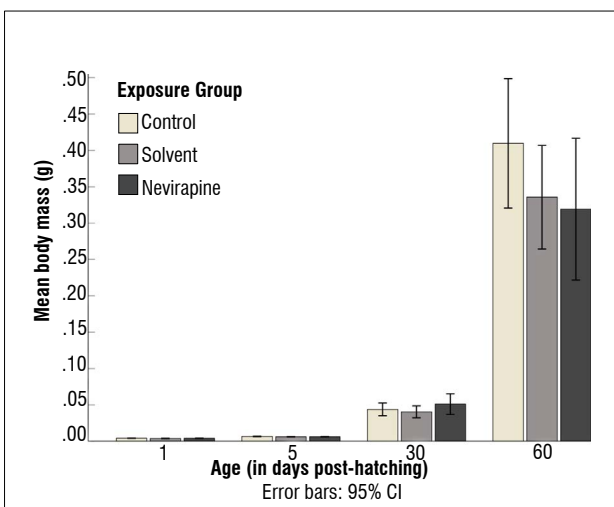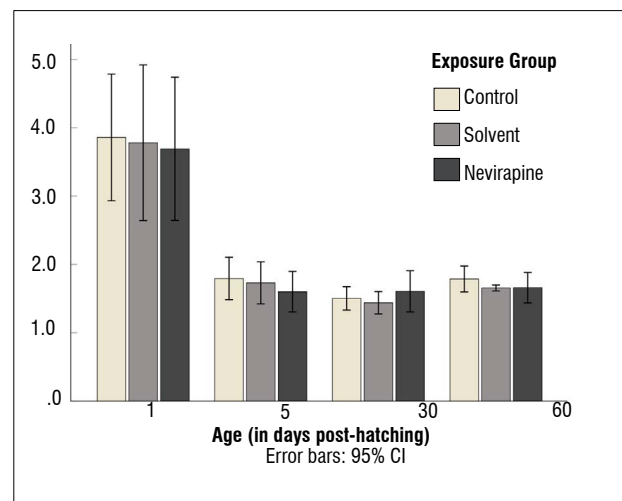| Age (days)		Control			DMSO			Nevirapine		
		Length (mm)	Mass (g)	Condition factor	Length (mm)	Mass (g)	Condition factor	Length (mm)	Mass (g)	Condition factor
1	N	35	35	35	35	35	35	35	35	35
	Mean	5.0	0.004	3.9	5.1	0.004	3.8	5.1	0.004	3.7
	s.d.	0.9	0.001	2.7	1.1	0.001	3.3	0.8	0.001	3.0
	Median	5.0	0.004	3.2	5.0	0.004	3.0	5.0	0.004	3.0
	Minimum	3.0	0.003	1.2	3.0	0.003	1.2	3.0	0.003	1.4
	Maximum	7.0	0.006	16.3	7.0	0.005	15.6	6.5	0.005	18.5
5	N	35	35	35	35	35	35	35	35	35
	Mean	7.4	0.006	1.8	7.3	0.006	1.7	7.5	0.006	1.6
	s.d.	0.8	0.001	0.9	0.9	0.001	0.9	0.8	0.001	0.9
	Median	7.5	0.006	1.5	7.5	0.006	1.4	7.8	0.006	1.3
	Minimum	5.2	0.004	0.8	5.0	0.004	0.8	5.0	0.004	0.8
	Maximum	9.0	0.009	5.0	9.0	0.008	4.8	9.0	0.009	4.8
30	N	35	35	35	35	35	35	35	35	35
	Mean	14.1	0.044	1.5	13.9	0.040	1.4	14.3	0.051	1.6
	s.d.	2.4	0.026	0.5	2.4	0.024	0.5	2.8	0.041	0.9
	Median	14.0	0.035	1.4	13.0	0.030	1.4	13.0	0.034	1.4
	Minimum	10.0	0.018	0.5	10.0	0.015	0.5	10.0	0.015	0.5
	Maximum	20.0	0.121	3.0	19.0	0.103	3.0	23.0	0.205	4.1
60	N	46	46	46	47	47	47	47	47	47
	Mean	26.5	0.410	1.8	26.0	0.336	1.7	25.1	0.319	1.7
	s.d.	8.0	0.300	0.6	6.1	0.243	0.1	7.0	0.332	0.8
	Median	26.0	0.350	1.7	25.0	0.282	1.7	24.0	0.227	1.6
	Minimum	13.0	0.032	1.2	15.0	0.051	1.3	14.5	0.048	0.2
	Maximum	41.0	1.116	5.5	44.0	1.372	2.0	49.0	1.976	6.2

Table 2: Specific growth rate (SGR) of fish in the three exposure groups from Day 1 to Day 30 and from Day 30 to Day 60 post-hatching

Exposure group	Repeat	SGR (%) Day 1 to 30	SGR (%) Day 30 to 60
Control	1	10.6	6.5
	2	8.1	7.9
	3	8.3	8.1
	4	6.6	7.5
	Mean ± s.d.	8.4 ± 1.6	7.5 ± 0.7
Solvent control	1	10.6	7.5
	2	8.2	6.0
	3	7.2	7.6
	4	7.6	7.5
	Mean ± s.d.	8.4 ± 1.5	7.2 ± 0.8
Nevirapine	1	11.8	6.9
	2	8.9	6.4
	3	8.6	5.4
	4	6.3	6.9
	Mean ± s.d.	8.9 ± 2.3	6.4 ± 0.7

Discussion

We used a laboratory chronic exposure to examine the possible effects of the highest environmental concentration of NVP previously reported in South African surface water on the growth of early life stages of *O. mossambicus*. Although our results showed no significant differences in length and mass of larvae and juveniles between those exposed to NVP and control fish, it was noted that 30- and 60-day-old juveniles showed variations in growth rate in the same container and in all groups. However, this finding is not abnormal in the *Oreochromis* genus; previous studies on various developmental stages (embryos, larvae and juveniles) of *O. niloticus* from several broods observed variations in development and growth rates among fish from the same brood – the authors of those studies suggested that the variations were probably caused by the density of fish and the inadequate quality of food.[33,34]

For the present study, the exposure started with the same number of embryos in each group for each repeat, and all larvae and early juveniles received the same treatment in terms of feeding and care. However, after hatching, there were mortality cases in all groups as shown by the survival rate presented in Figure 1. Even though the mortality rate was low with no significant differences in survival rate between groups, it may have caused a certain disturbance in the density of fish in the tanks concerned, leading to observed slight variations in growth parameters within groups. In addition, the length and mass measurements recorded in this study as well as the growth rate are comparable to those from a previous study on *O. mossambicus* from a natural environment in which 1-day-old larvae had a mean standard length of 4 mm[39] and *O. niloticus*

reared in captivity had lengths of 4.9 ± 0.1 mm for 1-day-old larvae and 17 ± 0.8 mm for 30-day-old juveniles.[33]

The strong positive correlation between length and mass measurements, and between fish age and both length and mass measurements, in all groups, confirmed that as the larval and juvenile stages of *O. mossambicus* were aging, they were also increasing both in length and in mass, as is the case for healthy fish.[28,40] If NVP in water had a negative impact on fish growth, the relationship between length and mass would have been affected, and the correlation would not have been strong and positive. In fisheries, healthy fish are characterised by strong positive mass–length, age–length and age–mass relationships; and length measurements are the most used in regressive models to predict either fish age or fish mass.[29] For the present study, it was also found that length measurements in all groups were good predictors of the age of fish.

The calculated CF showed that larvae and juvenile fish from all three groups had a high mean CF of about 4 for 1-day-old larvae; however, as the fish were growing, the CF decreased and approached 1 – the normal CF for healthy adult fish.[12] Froese[28] previously reported that fish early life stages have a normal CF above 1 because they increase more rapidly in length than they do in other dimensions. The mean CF of the fish in all the groups was within normal range for early life stages of fish; therefore, NVP had no significant negative impact on the CF.

The fish in all three groups had a high and rapid mean SGR from Day 1 to Day 30; from Day 30 to Day 60, the mean SGR decreased in all groups but was still high compared to the SGR recorded for *O. niloticus* in their natural environment.[40] *O. mossambicus* early life stages are known to have a fast and high initial growth rate to allow them to reach large sizes quickly in order to avoid predation.[41] According to Sparre and Venema[29], in normal conditions, the growth rate of a healthy fish decreases as the fish gets older, becoming zero when the fish gets to its maximal growth. Thus, even though the NVP-exposed juveniles showed a reduced mean SGR between the 30th and the 60th day of their life, statistical analysis showed no significant differences between the three groups. It is therefore evident that, despite the observed variations in the rate of development, the fish in all the three groups showed comparable growth rates from the first day of their life. Therefore, NVP did not significantly affect the growth rate of the fish. The variations observed in growth rate of the fish in the same group have also been observed in previous studies on *Oreochromis* and are believed to be mainly caused by fish density, food quality and variation in temperature of the water.[33,34,40,41] As this study was conducted in a controlled-environment room, and the fish in all groups were fed at the same time and with the same amount of food, the only factor that could have caused the variations in growth rate would be the slight variations in fish density, as explained above.

Conclusion

This study was the first in which the effects of NVP on larval and juvenile stages of *O. mossambicus* were assessed. It can be concluded that the highest concentration (1.48 μg/L) of NVP detected in South African waters at the start of this study, did not have significant detrimental effects on *O. mossambicus* early juvenile growth in terms of length and body mass during the 2 first months of life. However, a slightly reduced growth rate was noted between the 30th and the 60th day in the NVP-exposed fish compared to the control fish. Future studies should further investigate this finding by extending the exposure to adult stages.

Acknowledgements

Authors' contributions

U.M.C.N.: Conceptualisation; methodology; data collection; sample analysis; data analysis; validation; data curation; writing – the initial draft. G.M.W.: Conceptualisation; project leadership; project management; funding acquisition; writing – revisions; student supervision. I.E.J.B.: Writing – revisions; student supervision; funding acquisition.

References

1. Boxall ABA. The environmental side effects of medication. EMBO Rep. 2004;5(12):1110–1116. https://doi.org/10.1038/sj.embor.7400307

2. Corcoran J, Winter MJ, Tyler CR. Pharmaceuticals in the aquatic environment: A critical review of the evidence for health effects in fish. Crit Rev Toxicol. 2010;40(4):287–304. https://doi.org/10.3109/10408440903373590

3. World Health Organization (WHO). Pharmaceuticals in drinking water. WHO/HSE/WSH/11.05. Geneva: World Health Organization; 2011. Available from: http://www.who.int/water_sanitation_health/publications/2011/pharmaceuticals_20110601.pdf

4. Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB, et al. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams 1999–2000: A national reconnaissance. Environ Sci Technol. 2002;36(6):1202–1211. https://doi.org/10.1021/es011055j

5. Kidd AK, Blanchfield JP, Mills HK, Palace PV, Evans ER, Lazorchak MJ, et al. Collapse of a fish population after exposure to a synthetic estrogen. Proc Natl Acad Sci USA. 2007;104(21):8897–8901. https://doi.org/10.1073/pnas.0609568104

6. Lin T, YanQiu C, Chen W. Impact of toxicological properties of sulphonamides on the growth of zebrafish embryos in the water. Environ Toxicol Pharmacol. 2013;36:1068–1076. https://doi.org/10.1016/j.etap.2013.09.009

7. Triebskorn R, Casper H, Scheil V, Schwaiger J. Ultrastructural effects of pharmaceuticals (carbamazepine, clofibric acid, metoprolol, diclofenac) in rainbow trout (*Oncorhynchus mykiss*) and common carp (*Cyprinus carpio*). Anal Bioanal Chem. 2007;387:1405–1416. https://doi.org/10.1007/s00216-006-1033-x

8. Robson L, Barnhoorn IEJ, Wagenaar I. The potential effects of efavirenz on *Oreochromis mossambicus* after acute exposure. Environ Toxicol Pharmacol. 2017;56:225–232. https://doi.org/10.1016/j.etap.2017.09.017

9. Aus der Beek T, Weber FA, Bergmann A, Hickmann S, Ebert I, Hein A, et al. Pharmaceuticals in the environment – Global occurrences and perspectives. Environ Toxicol Chem. 2016;35(4):823–835. https://doi.org/10.1002/etc.3339

10. Human Sciences Research Council (HSRC). The fifth South African national HIV prevalence, incidence, behaviour and communication survey, 2017: HIV impact assessment summary report. Cape Town: HSRC Press; 2018. Available from: http://www.hsrc.ac.za/uploads/pageContent/9225/SABSSMV_Impact_Assessment_Summary_ZA_ADS_cleared1%20(002).pdf

11. World Health Organization (WHO). Global update on HIV treatment 2013: Results, impact and opportunities: WHO report in partnership with UNICEF and UNAIDS. Geneva: World Health Organization; 2013. Available from: http://www.who.int/hiv/pub/progressreports/update2013/en/

12. Prasse C, Schlusener MP, Schulz R, Ternes TA. Antiviral drugs in wastewater and surface waters: A new pharmaceutical class of environmental relevance? Environ Sci Technol. 2010;44(5):1728–1735. https://doi.org/10.1021/es903216p

13. Swati J, Raj VK, Prabhat P, Sangeeta V. A review on fate of antiviral drugs in environment and detection techniques. Int J Environ Sci. 2011;1(7):1526–1541. https://doi.org/10.6088/ijessi.00107020012

14. Madikizela LM, Tavengwa NT, Chimuka L. Review: Status of pharmaceuticals in African water bodies: Occurrence, removal and analytical methods. J Environ Manage. 2017;193:211–220. https://doi.org/10.1016/j.jenvman.2017.02.022

15. Wood TP, Basson AE, Duvenage C. The chlorination behaviour and environmental fate of the antiretroviral nevirapine in South African surface water. Water Res. 2016;104:349–360. https://doi.org/10.1016/j.watres.2016.08.038

16. Wood TP, Duvenage CSJ, Rohwer E. The occurrence of anti-retroviral compounds used for HIV treatment in South African surface waters. Environ Pollut. 2015;199:235–243. https://doi.org/10.1016/j.envpol.2015.01.030

17. Swanepoel C, Bouwman H, Pieters R, Bezuidenhout C. Presence, concentrations and potential implications of HIV-anti-retrovirals in selected water resources in South Africa. WRC report no. 2144/1/14. Pretoria: Water Research Commission; 2015. https://doi.org/10.13140/RG.2.2.20637.51688

18. Vankova M. Biodegradability analysis of pharmaceuticals used in developing countries; screening with OxiTop®-C 110 [thesis]. Tampere, Finland: Tampere University of Technology; 2011.

19. Antunes MMA, Sidarus M, Novais AD, Harjivan GS, Santos PP, Da Silva JLF, et al. Oxidation of 2-hydroxynevirapine, a phenolic metabolite of the anti-HIV drug nevirapine: Evidence for an unusual pyridine ring contraction. Molecules. 2012;17:2616–2627. https://doi.org/10.3390/molecules17032616

20. Rivero A, Mira JA, Pineda JA. Liver toxicity induced by non-nucleoside reverse transcriptase inhibitors. J Antimicrob Chemother. 2007;59:342–346. https://doi.org/10.1093/jac/dkl524

21. Brausch JM, Connors KA, Brooks BW, Rand GM. Human pharmaceuticals in the aquatic environment: A review of recent toxicological studies and considerations for toxicity testing. Rev Environ Contam Toxicol. 2012;218:1–99. https://doi.org/10.1007/978-1-4614-3137-4_1

22. Stancova V, Plhalova L, Bartoskova M, Zivna D, Prokes M, Marsalek P, et al. Effects of mixture of pharmaceuticals on early life stages of tench (*Tinca tinca*). Biomed Res Int. 2014;2014: 253–468. http://dx.doi.org.10.1155/2014/253468

23. Memmert U, Peither A, Burri R, Weber K, Schmidt T, Sumpter JP, et al. Diclofenac: New data on chronic toxicity and bio-concentration in fish. Environ Toxicol Chem. 2013;32(2):442–452. https://doi.org/10.1002/etc.2085

24. Schoenfuss HL, Furlong ET, Phillips PJ, Scott TM, Kolpin DW, Cetkovic-Cvrlje M, et al. Complex mixtures, complex responses: Assessing pharmaceutical mixtures using field and laboratory approaches. Environ Toxicol Chem. 2016;35(4):953–965. https://doi.org/10.1002/etc.3147

25. Gonzalez-Doncel M, Garcia-Maurino JE, San Segundo L, Beltran EM, Sastre S, Torija CF. Embryonic exposure of medaka (*Oryzias latipes*) to propylparaben: Effects on early development and post-hatching growth. Environ Pollut. 2014;134:360–369. https://doi.org/10.1016/j.envpol.2013.09.022

26. Holden KK, Bruton MN. A life-history approach to the early ontogeny of the Mozambique tilapia *Oreochromis mossambicus* (Pisces, Cichlidae). Afr Zool. 1992;27:173–191. https://doi.org/10.1080/02541858.1992.11448279

27. Sadauskas-Henrique H, Sakuragui MM, Paulino MG, Fernandes MN. Using condition factor and blood variable biomarkers in fish to assess water quality. Environ Monit Assess. 2011;181:29–42. https://doi.org/10.1007/s10661-010-1810-z

28. Froese R. Cube law, condition factor and weight-length relationships: History, meta-analysis and recommendations. J Appl Ichthyol. 2006;22:241–253. https://doi.org/10.1111/j.1439-0426.2006.00805.x

29. Sparre P, Venema SC. Introduction to tropical fish stock assessment. Part 1: Manual. FAO Fisheries Technical Paper no. 306.1, Rev. 2. Rome: FAO; 1998. Available from: http://www.fao.org/docrep/W5449E/w5449e00.htm

30. Pena R, Dumas S. Development and allometric growth patterns during early larval stages of the spotted sand bass *Paralabrax maculatofasciatus* (Percoidei: Serranidae). In: Clemmesen C, Malzahn AM, Peck MA, Schnack D, editors. Advances in early life history study of fish. Barcelona: Scientia Marina; 2009. p. 183–189. Available from: http://hdl:10013/epic.36540

31. Sumpter JP. Pharmaceuticals in the environment: Moving from a problem to a solution. In: Kummerer K, Hempel M, editors. Green and sustainable pharmacy. Berlin: Springer-Verlag; 2010. p. 11–22. Available from: https://www.springer.com/us/book/9783642051982

32. Kruger T. Effects of zinc, copper and cadmium on *Oreochromis mossambicus* free embryos and randomly selected mosquito larvae as biological indicators during acute toxicity testing [thesis]. Johannesburg: Rand Afrikaans University; 2002.

33. Fujimura K, Okada N. Development of the embryo, larva and early juvenile of the Nile tilapia *Oreochromis niloticus*. Developmental staging system. Develop Growth Differ. 2007;49:301–324. https://doi.org/10.1111/j.1440-169X.2007.00926.x

34. Morrison CM, Miyake T, Wright JR Jr. Histological study of the development of the embryo and early larvae of *Oreochromis niloticus* (Pisces: Cichlidae). J Morphol. 2001;247:172–195. https://doi.org/10.1002/1097-4687(200102)247:2<172::AID-JMOR1011>3.0.CO;2-H

35. Carlander KD. Handbook of freshwater fishery biology. 3rd ed. Ames, IA: Iowa State University Press; 1969. Available from: https://UnivofPretoria.on.worldcat.org/oclc/934888916

36. Bagenal TB, Tesch FW. Age and growth. In: Bagenal TB, editor. Method assessment of fish production in freshwater. 3rd ed. Oxford: Blackwell Scientific Publication; 1978. p. 101–136.

37. Organisation for Economic Co-operation and Development (OECD). Test no. 210: Fsh, early-life stage toxicity test. OECD guidelines for the testing of chemicals, section 2. Paris: OECD Publishing; 2013. https://doi.org/10.1787/9789264203785-en

38. Ferrer I, Thurman EM. Analysis of 100 pharmaceuticals and their degradates in water samples by liquid chromatography/quadrupole time-of-flight mass spectrometry. J Chromatogr A. 2012;1259:148–157. https://doi.org/10.1016/j.chroma.2012.03.059

39. Tachihara K, Obara E. Morphological development of embryos and juveniles in the Mozambique tilapia, *Oreochromis mossambicus* as a direct developmental fish under rearing conditions. Suisan Zoshoku. 2003;51(3):295–306.

40. Makori AJ, Abuon PO, Kapio R, Anyona DN, Dida GO. Effects of water physico-chemical parameters on tilapia (*Oreochromis niloticus*) growth in earthen ponds in Teso North Sub-County, Busia County. Fish Aquat Sci. 2017;20, Art. #30, 10 pages. https://doi.org/10.1186/s41240-017-0075-7

41. Weyl OLF, Hecht T. The biology of *Tilapia rendalli* and *Oreochromis mossambicus* (Pisces: Cichlidae) in a subtropical lake in Mozambique. Afr Zool. 1998;33(3):178–188. https://doi.org/10.1080/02541858.1998.11448469

AUTHORS:
Kathleen A. Grant[1,2,3]
Ettienne J. Myburgh[4,5]
Elizabeth Murray[6,7,8]
Fredrieka M. Pienaar[9]
Martin Kidd[10]
Colleen A. Wright[1,11]
Maritha J. Kotze[12,13]

AFFILIATIONS:
[1]Division of Anatomical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[2]Department of Biomedical Sciences, Faculty of Health and Wellness, Cape Peninsula University of Technology, Cape Town, South Africa
[3]DST-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa
[4]Private practice, Panorama Medi-Clinic, Cape Town, South Africa
[5]Division of Surgery, Stellenbosch University, Cape Town, South Africa
[6]Rondebosch Medical Centre, Cape Town, South Africa
[7]Medi-Clinic Constantiaberg, Cape Town, South Africa
[8]University of Cape Town Private Academic Hospital, Cape Town, South Africa
[9]CancerCare, Panorama Medi-Clinic, Cape Town, South Africa
[10]Centre for Statistical Consultation, Department of Statistics and Actuarial Sciences, Stellenbosch University, Stellenbosch, South Africa
[11]Lancet Laboratories, Johannesburg, South Africa
[12]Division of Chemical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[13]South African National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa

CORRESPONDENCE TO:
Kathleen Grant

EMAIL:
grantk@cput.ac.za

DATES:
Received: 17 Aug. 2018
Revised: 22 Feb. 2019
Accepted: 27 Feb. 2019
Published: 27 Mar. 2019

Reclassification of early stage breast cancer into treatment groups by combining the use of immunohistochemistry and microarray analysis

Immunohistochemistry (IHC) is routinely used to approximate breast cancer intrinsic subtypes, which were initially discovered by microarray analysis. However, IHC assessment of oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2) status, is a poor surrogate of molecular subtype. Therefore, MammaPrint/BluePrint (MP/BP) microarray gene expression profiling is increasingly used to stratify breast cancer patients into different treatment groups. In this study, ER/PR status, as reported by standard IHC and single-gene mRNA analysis using TargetPrint, was compared with molecular subtyping to evaluate the combined use of MP/BP in South African breast cancer patients. Pathological information of 74 ER/PR positive, HER2 negative tumours from 73 patients who underwent microarray testing, were extracted from a central breast cancer genomics database. The IHC level was standardised by multiplying the intensity score (0–3) by the reported proportion of positively stained nuclei, giving a score of 0–300. Comparison between mRNA levels and IHC determination of ER/PR status demonstrated a significant correlation ($p<0.001$) for both receptors (ER: 0.34 and PR: 0.54). Concordance was shown in 61 (82%) cases and discordance in 13 (18%) of the 74 tumours tested. Further stratification by MP/BP identified 49 (66.2%) Luminal A, 21 (28.4%) Luminal B and 4 (5.4%) Basal-like tumours. Neither IHC nor TargetPrint could substitute BP subtyping, which measures the functional integrity of ER and can identify patients with false-positive tumours who are resistant to hormone therapy. These findings support the implementation of a pathology-supported genetic testing approach combining IHC and microarray gene profiling for definitive prognostic and predictive treatment decision-making in patients with early stage breast cancer.

Significance:

- Single-gene genomic oestrogen and progesterone receptor reporting adds limited additional information to the molecular stratification of breast cancer tumours and does not supersede the immunohistochemistry results.

- Neither single-gene genomic mRNA nor immunohistochemistry reporting of oestrogen and progesterone receptor status can replace the combined use of MammaPrint/BluePrint genomic molecular subtyping.

- Reliable distinction between Luminal A and B type tumours is not possible using immunohistochemistry or single-gene genomic mRNA assessment of oestrogen/progesterone and HER2 receptor status.

- Combining immunohistochemistry and microarray gene profiling enables the identification of endocrine treatment resistant hormone-positive tumours lacking $ER\alpha$ function (Basal-like), despite positive expression at the protein and single-gene RNA level.

Introduction

Breast cancer defines a broad spectrum of histological lesions that are considered highly heterogeneous in presentation, morphological characteristics, prognosis and therapeutic outcome.[1] Microarray-based gene expression profiling led to the discovery of intrinsic molecular subtypes underlying the variability in biological behaviour and response to treatment amongst breast cancer patients.[2,3]

Five distinct subtypes were described by Perou et al.[2], although the normal-like subtype was subsequently considered to represent normal breast tissue within tumour. Luminal tumours expressed oestrogen receptor (ER) and responded to endocrine therapy. While Luminal A tumours show little benefit from the addition of chemotherapy, Luminal B tumours display some genetic similarities to Basal-like tumours in that they have a higher risk of being hormone resistant and show additional benefit from chemotherapy as demonstrated by the significant pathological complete response rate after neo-adjuvant chemotherapy. Human epidermal growth factor receptor-2 (HER2)-enriched and Basal-like subtypes are considered more aggressive with an unfavourable prognosis, although paradoxically exhibiting greater chemosensitivity compared to the Luminal subtype. Basal-like breast cancers are inherently resistant to endocrine therapy, and tumours subtyped as HER2 enriched respond to anti-HER2 therapy in addition to chemotherapy. Some tumours reported as HER2 positive are subtyped as Luminal B and retain some responsiveness to endocrine treatment in addition to chemotherapy and HER2 targeted treatments.[4] Borley et al.[5] demonstrated that the HER2 gene copy number provides additional information for stratifying breast cancer patients into different treatment groups, because HER2-positive patients with a low degree of HER amplification were shown to derive less benefit from trastuzumab (the chemotherapy agent more commonly known as Herceptin®).

Numerous studies using standard pathology have been performed to identify, with some accuracy, treatable molecular subtypes. Suggestions for incorporating markers such as the epidermal growth factor receptor (EGFR), proliferation marker Ki67, tumour suppressor gene protein p53, transmembrane tyrosine kinase receptor CD117(c-kit) and cytokeratin 5/6 into a standard immunohistochemistry (IHC) panel for breast cancer[3,6] have not been adopted because of poor standardisation. High-quality assessed Ki67 is considered most useful when the indication for

ARTICLE INCLUDES:
☒ Peer review
☐ Supplementary material

DATA AVAILABILITY:
☐ Open data set
☐ All data included
☒ On request from authors
☐ Not available
☐ Not applicable

KEYWORDS:
oestrogen receptor; microarray; molecular subtype; BluePrint; MammaPrint

adding adjuvant chemotherapy to endocrine treatment is uncertain, but molecular classification can help to identify a larger group of early-stage breast cancer patients with low risk of recurrence.[7] Some studies suggested that loss of progesterone receptor (PR) expression might be indicative of the Luminal B subtype[8], but this association has not been universally reported. ER, PR and HER2 status have been incorporated into standard pathology reporting of breast cancers with reproducible prognostic and predictive value[9,10]. It is only with the advent of genetic tumour profiling that accurate molecular subtyping became part of daily clinical practice. A prospective study performed by Whitworth et al.[11] proved that the combined use of the 70-gene MammaPrint (MP) and 80-gene BluePrint (BP) assays in microarray analysis of mRNA expression may be more accurate than standard IHC to guide treatment decisions. Notably, 22% of over 400 breast cancer patients studied were reclassified into a different subgroup compared with conventional assessment and showed an improved distribution of response rates in the relevant treatment groups. Similar findings were reported by Yao et al.[12] Previous studies have shown that mRNA reporting of ER, PR and HER2 using microarrays is highly comparable to IHC testing.[13-15] However, others have cautioned against the preferential use of hormone receptor reports using RNA-based reverse transcription polymerase chain reaction (RT-PCR) technology, highlighting discordance with the IHC results and the potential of denying patients who were ER or PR positive on IHC the benefit of endocrine therapy.[16,17] Application of different methodologies for the same purpose therefore requires careful consideration.

Microarray-based tumour profiling using the 70-gene MP profile has been available in South Africa since 2007 and, from 2009, local referral criteria were introduced for reimbursement by medical aid providers.[18] Initially, analysis was performed on fresh tissue only, but since 2012 the use of formalin fixed paraffin embedded (FFPE) tissue became available and has become the only method used. A central database was established by using an ethics approved protocol for comparative effectiveness studies on data of MP tests requested in southern Africa. In addition to ER/PR mRNA reporting by TargetPrint (TP) from 2009, BP – which determines the tumour molecular subtype[19] – has also become part of the MP service from 2011. BP provides a comprehensive multigene expression analysis of the tumour molecular subtypes, which may not be sufficiently reflected by single-gene IHC or mRNA testing.[20,21] Although our data on HER2 expression indicated a 100% correlation between fluorescence in-situ hybridisation (FISH) and microarray testing using TP[22], discrepancies in ER/PR reporting between IHC and RNA-based RT-PCR techniques[23,24] warranted evaluation in this study of the value and potential clinical impact of ER and PR status as reported by TP using RNA-based microarray analysis.

Our aim was to evaluate the combined use of MP/BP in clinical practice using a pathology-supported genetic testing approach incorporating ER, PR and HER2 status as part of the above-mentioned referral criteria, called the microarray pre-screen algorithm (MPA). ER and PR status as assessed by IHC was compared with that reported by TP in order to determine the correlation between the two techniques. We used the BP result to identify the molecular subtype and most probable response to therapy and correlated this result to the IHC and TP results. This study is the first to correlate IHC and mRNA hormone receptor status in South African breast cancer patients, using microarrays performed on FFPE specimens in the context of molecular breast cancer subtyping. As a consequence of the MPA employed in southern Africa, which generally excludes patients with ER/PR negative or HER2 positive tumours from testing, our series was limited to ER/PR positive, HER2 negative tumours only. South African patients reclassified as HER2 positive using TP and reflex FISH in a recent study[22] were also excluded from the current analysis.

Methods

This study was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Ethical approval was granted by the Health and Research Ethics Committee (HREC) of Stellenbosch University (reference number N09/06/166).

Study population

The records of 128 patients with 131 tumours were available in a central breast cancer genomic database, which comprised data of all patients referred for MP in southern Africa between 2007 and 2014. The database is maintained centrally under a rigorous quality control programme to ensure the integrity of the data. We extracted anonymised pathological data of 74 HER2 negative tumours from 73 patients who had TP and BP testing on FFPE tissue. The tumour pathology included tumour type, grade and size; ER, PR and HER2 status; MP risk status; TP results for ER, PR and HER2; and the molecular subtype as determined by BP.

Immunohistochemistry testing

Standard pathology reporting of hormone and HER2 receptor status using IHC to measure protein expression levels varies amongst different laboratories. To standardise the data for statistical analysis of hormone receptor status, estimation of the semi-quantitative expression of ER and PR was performed using the intensity score (0–3), multiplied by the reported proportion of positively stained nuclei, thereby calculating a final ER and PR score (0–300).

Microarray-based gene expression profiling

Microarray-based gene expression profiling (MP, TP and BP) was performed on 74 FFPE tissue samples obtained from 73 breast cancer patients, using a pathology-supported genetic testing strategy.[18] An experienced pathologist evaluated tumour suitability for genomic analysis based on confirmation of a minimum tumour cell content of 30% in accordance with compliancy criteria laid out by the US Food and Drug Administration. These samples were transported under an export permit to the Netherlands where tumour assessment was performed at the centralised Agendia Laboratory in accordance with standard testing protocols.[25] mRNA expression for ER and PR is reported on a continuous exponential scale from -1 to 1 and values of less than 0 are considered to be negative.

Comparative analysis

Quantitative analysis was performed by comparing the level of mRNA for ER and PR as reported by TP with the IHC score. ER and PR were considered to be positive when the IHC score was >10 or the mRNA score was >0. Qualitative analysis was performed to allow in-depth evaluation of the relationship between IHC and mRNA compared to tumour subtyping, with the aim of determining the clinical implications of individual versus combined assessment of pathology and microarray-based genetic testing.

Statistical analysis

Statistical analysis was performed using the Statistica v.13 software package. Observer agreement measures for IHC testing to determine hormone receptor status and microarray-based mRNA readout assessment were calculated from two-way contingency table analysis.[26] The relationship between protein expression (IHC) and mRNA (TP) levels was assessed using Spearman rank correlation analysis. A possible association between loss of PR expression in ER-positive cases (ER+/PR-) and the high-risk Luminal B subtype as determined by microarray analysis was further assessed. Results corresponding to a $p < 0.05$ were considered statistically significant.

Results

Description of tumour pathology in relation to molecular subtype

All tumours were ER and/or PR positive on IHC. Analysis of the pathological characteristics of the 74 HER2 negative tumours subjected to TP (ER, PR and HER2 status) and BP using FFPE specimens is presented in Table 1. Molecular subtyping using the BP microarray test stratified 49 (66.2%) tumours as Luminal A, 21 (28.4%) as Luminal B and 4 (5.4%) as Basal-like. In the patient with multi-focal pathology, one tumour was designated Luminal A and the other Luminal B.

Table 1: Comparison of tumour morphology and grade between breast cancer subtypes according to the BluePrint microarray in 74 tumour specimens from 73 South African breast cancer patients

	Total	Luminal A	Luminal B	Basal-like
	n (%)	n (%)	n (%)	n (%)
Total (n)	74 (100)	49 (66.2)	21 (28.4)	4 (5.4)
Tumour morphology				
Ductal	62 (83.8)	41 (67.2)	17 (27.9)	4 (6.5)
Lobular, pleomorphic	3 (4.1)	1 (33.3)	2 (66.7)	0 (0.0)
Lobular, classic	9 (12.2)	7 (77.8)	2 (22.2)	0 (0.0)
Tumour grade				
1	17 (23)	14 (82.4)	3 (17.6)	0 (0.0)
2	35 (47.2)	24 (68.6)	10 (28.6)	1 (2.8)
3	15 (20.3)	6 (40.0)	6 (40.0)	3 (20)
n/a	7 (9.5)	5 (71.4)	2 (28.6)	0 (0.0)

Comparative analysis of ER and PR status

A quantitative comparison between mRNA levels and IHC determination of ER and PR is shown in Figures 1 and 2. Spearman R-values were 0.34 for ER and 0.54 for PR ($p < 0.001$), indicating the correlation between mRNA and IHC. Table 2 indicates the mean values for IHC and mRNA

scores for both ER and PR according to tumour subtype. Significant differences were observed amongst tumour subtypes for all variables except for mRNA expression of PR. Fisher's Least Significant Difference Test confirmed significant differences between Basal type and Luminal types with no difference seen between Luminal A and B subtypes. The sum of ER and PR scores, which resulted in maximum values of 600 and -2 to +2 for IHC and mRNA, respectively, indicated significant differences between Basal and Luminal tumours. Comparably, no differences were shown between Luminal A and B molecular subtypes. The combined score also amplified the differences between the groups, as illustrated in Figure 3.

Figure 1: Scatterplot showing the relationship between oestrogen receptor (ER) immunohistochemistry (IHC) score and mRNA score; Spearman $R = 0.34$ ($p = 0.00$).

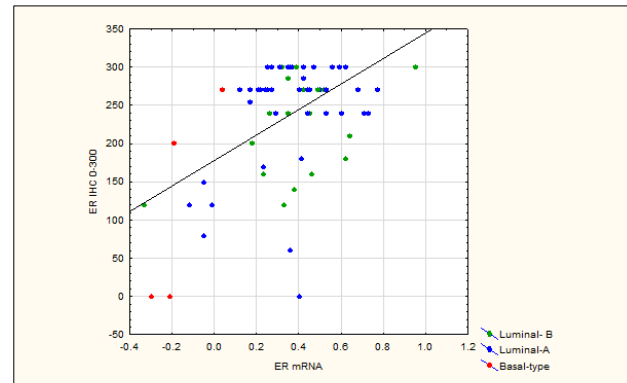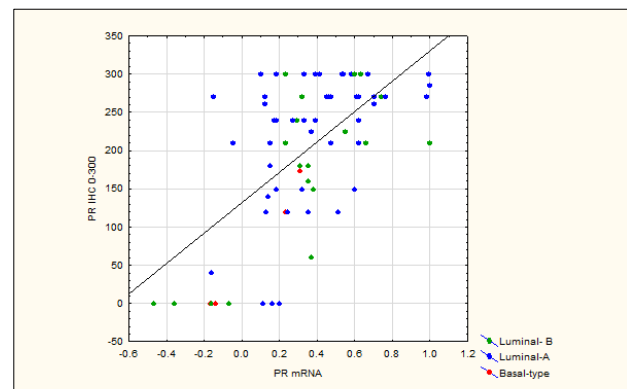**Figure 2:** Scatterplot showing the relationship between progesterone receptor (PR) immunohistochemistry (IHC) score and mRNA score; Spearman $R = 0.54$ ($p = 0.00$).
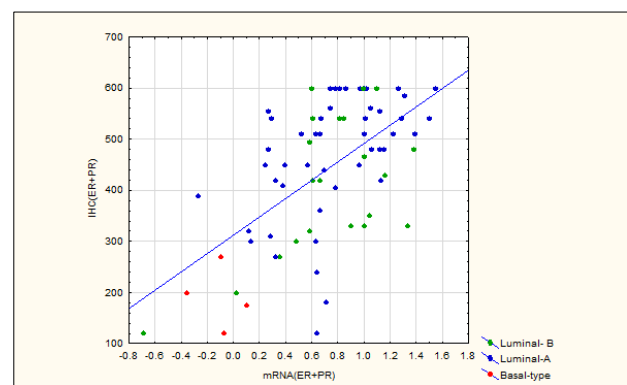

Figure 3: Combined oestrogen (ER) and progesterone receptor (PR) scores for immunohistochemistry (IHC) (0–600) and TargetPrint (-2 to +2), in relation to tumour molecular subtype determined using BluePrint.

Table 2: Mean values of oestrogen receptor (ER), progesterone receptor (PR) and combined scores for immunohistochemistry (IHC) and RNA in relation to tumour subtype

Quantitative score	Luminal A (n=49)	Luminal B (n=21)	Basal (n=4)
ER IHC score 0–300	249.3878	232.1429	117.5*
PR IHC score 0–300	222.4898	181.1905	73.5*
ER RNA score -1 to 1	0.381	0.399	-0.165*
PR RNA score -2 to 2	0.391	0.332	0.058
Combined ER and PR IHC score 0–600	471.877	413.333	191.0*
Combined ER and PR RNA score -2 to 2	0.772857	0.773	-0.1075*

*$p<0.01$. In all these cases, Basal type was found to differ significantly from both Luminal A and B subtypes.

A qualitative analysis was performed to evaluate the relationship between IHC and mRNA in relation to tumour subtyping. Table 3 provides a summary of the data. Of the 74 tumours, ER and PR status was concordant in 61 (82%) and discordant in 13 (18%). In the concordant group, 39 (64%) were Luminal A, 17 (31%) were Luminal B and 3 (5%) were Basal breast cancer subtypes. In the discordant group (n=13), three tumours were ER and PR negative on mRNA, with one being Luminal A, one being Luminal B and the other being a Basal-subtype. Of these 13 discordant tumours, 10 (77%) were Luminal A, 2 (15%) were Luminal B and 1 (8%) was a Basal-type.

Table 3: Tumour classification according to molecular subtyping (BluePrint) assessed in relation to ER/PR status in IHC/TargetPrint concordant (n=61) as well as discordant (n=13) cases

Qualitative IHC ER/PR expression n (%)	TargetPrint ER/PR mRNA expression	BluePrint Molecular subtyping Gene expression profiling n (%)		
		Luminal A	Luminal B	Basal-like
Concordance 61 (82.4)				
ER+/PR+ 56 (91.8)		39 (69.6)	17 (30.4)	
ER+/PR- 3 (4.9)		0	2 (67.7)	1 (33.3)
ER-/PR+ 2 (3.2)		0	0	2 (100)
Discordance 13 (17.6)				
ER+/PR+ 7 (9.5)	(3)ER-/PR+	3 (42.8)		
	(2)ER+/PR-	2 (28.6)		
	(1)ER-/PR-	1 (14.3)		
	(1)ER+/PR-		1(14.3)	
ER+/PR- 5 (6.8)	(3)ER+/PR+	3 (60)		
	(2)ER-/PR-		1 (20)	1 (20)
ER-/PR+ 1 (1.4)	(1)ER+/PR+	1(100)		

ER/PR, oestrogen/progesterone receptor; IHC, immunohistochemistry

The relationship between IHC and mRNA status versus tumour subtype is indicated in Figure 4. Importantly, none of the ER+/PR+ tumours determined on IHC were reported as Basal subtype breast cancers. When either ER or PR expression was lost, on both IHC and mRNA, there was a 57% (4 out of 7) risk of a Basal subtype. No predictive pattern in hormone receptor expression determined by IHC could be identified to distinguish Luminal A from Luminal B tumours identified through combined use of MP/BP.
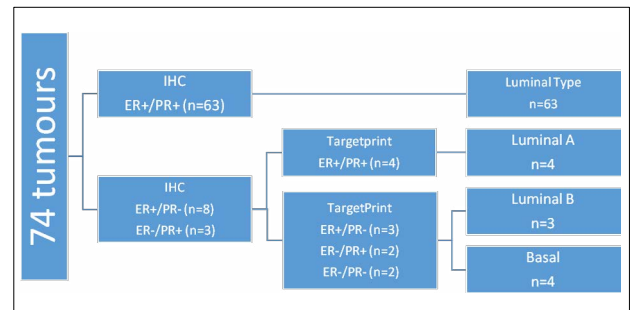

Figure 4: Flow diagram showing the relationship of immunohistochemistry (IHC) and TargetPrint results on molecular subtype determined by combined use of MammaPrint and the BluePrint functional subtype classifier.

Discussion

Ongoing validation of emerging genomic technologies against current standards in breast cancer pathology is an important research focus. In South Africa, an increased level of confidence based on growing clinical experience supported the incorporation of MP/BP[19,27] into routine clinical practice. Integration of these results into treatment decision may lead to a change in therapy in one of every two early-stage breast cancer patients treated in South Africa.[28] In the present study, the molecular subtype based on BP was used as an indicator of the expected response to therapy. The Luminal type of breast cancer is expected to respond to endocrine therapy, whereas the Basal subtype is inherently resistant to endocrine therapy. IHC and TP results were evaluated for their ability to predict a non-Luminal subtype.

In contrast to previous results from our research group showing a 100% correlation in HER2 status between FISH and TP, irrespective of whether fresh or FFPE specimens were used[22], our current qualitative results show 18% discordance between the IHC and TP determination of ER and PR status. More importantly, three (4%) of the ER+/PR+ tumours as determined by IHC/FISH were ER-/PR- on TP regardless of being a Luminal-type. If the TP results were interpreted without the IHC and/or BP results, the results could have a major implication on treatment decision-making for the patients. It is therefore important to realise that determination of mRNA status for ER and PR using single-gene mRNA analysis does not necessarily translate to protein expression or reflect the presence of a functioning receptor protein.[29] Because molecular subtyping performed through BP is enriched for several genes involved in ER function, the combined score of 80 genes included in the BP microarray profile provides a better indication of an intact ER-mediated kinase pathway and subsequent response to endocrine therapy.[20]

Reasons for the apparent discrepancy between mRNA expression and receptor protein levels based on IHC can be found in our quantitative analyses which indicated that ER and PR protein expression levels varied substantially in relation to mRNA levels. In Figure 1, there were eight tumours with mRNA scores of <0, indicating ER mRNA negativity, but with a mean IHC score of 156 (0–300), and only two had an IHC score <10. As mRNA levels increased above 0, there was only one Luminal A tumour with an IHC score of 0. For PR there seemed to be better correlation of low mRNA levels to the IHC score, with 6 out of 9 tumours reported with a mRNA score <0 showing an IHC score of 0. Similar results were reported in a study using quantitative immunofluorescence for measuring mRNA in situ by Bordeaux et al.[30] These authors found a nonlinear

relationship between ERα protein expression and mRNA levels on tissue sections visualised using RNA scope probes. In a similar finding to this study, protein expressions varied considerably at very low levels of RNA. In their analysis, mRNA did not show any prognostic value but had some predictive value above and beyond that of the ERα protein expression.

The presence of a splice variant in the ER gene might also influence the apparent ER protein expression, as previously described by Groenendijk et al.[31] Several ER splicing variants have been reported in the literature, resulting in one or more exons being omitted from the ER mRNA. Advanced techniques such as exome sequencing might be particularly effective in identifying these abnormalities.[32] Antibodies used to detect ER protein during routine IHC assessment rely on epitope recognition encoded by the first exon of the ER gene; therefore, antibody binding at these sites will produce a positive ER result, despite the lack of normal ER functionality as a result of a splice variant. Similarly, RNA methods based on single-gene identification of ER status such as TP (microarray, previously provided as a separate readout with MP) and Oncotype DX (RT-PCR, included in 21-gene assay) will not routinely detect such variants. The ability of microarray analysis to detect functional ERα activity, could explain why a patient with the ERΔE7 splice variant as noted by Groenendijk et al.[31] was classified as high risk by MP, but low risk by Oncotype DX (Recurrence Score of 8). Comprehensive genomic evaluation using multi-gene tests such as MP/BP is needed to identify these hormone-resistant tumours. This justifies recent discontinuation of TP and introduction of the European Conformity (CE)-marked next-generation sequencing based MP/BP® (Agendia, Inc.).

Numerous technical factors have an influence on the accuracy of standard IHC reporting, including the effect of cold ischaemic time on ER epitope availability, resulting in false low levels of protein expression.[33,34] Different samples used for initial IHC versus subsequent receptor status determination using TP, could contribute to discordant results as some authors have reported high false positive and negative rates when testing was performed on tissue obtained from core needle biopsies as opposed to resected specimens.[35] Sample differences might also play a role as some of the patients had the initial receptor status reported on core needle biopsies whereas TP was performed on the resection specimen. In most units, however, this is an accepted practice with good correlation, although it should be avoided in ER and PR negative tumours.[36]

Variations in individual ER and PR levels measured by IHC or mRNA levels did not discriminate between Luminal A and B tumours. Basal tumours had significantly lower scores for ER and PR on IHC as well as mRNA levels for ER. Utilising a summative ER/PR score for both IHC and mRNA allowed better discrimination between Basal tumours and Luminal tumours. When this combined score was employed, as illustrated in Figure 3, four out of six tumours with an IHC ≤ 270 and mRNA ≤0.1 were Basal while the remaining two were Luminal B breast cancer subtypes. This was also reflected in our qualitative analysis. All the tumours which were ER+/PR+ using IHC were Luminal A or B subtypes and the additional results provided by TP did not add any further information to assist in distinguishing Luminal A from Luminal B tumours identified by BP. In cases where ER or PR was negative on IHC, TP did add some additional value although the numbers in this group were limited ($n=11$). In this group, all four tumours for which TP showed an ER+/PR+ profile, were Luminal A and would potentially gain little or no benefit from additional chemotherapy. The remainder were either Luminal B or Basal subtypes, suggesting the addition of chemotherapy to the treatment plan.

Interpretation of the results is limited by a relatively small sample size as well as pre-selection of hormone-positive and HER2 negative patients in compliance with the MPA developed for reimbursement purposes in South Africa.[18] This precluded evaluation of the effect TP might have on reclassification of IHC ER-/PR- tumours[22], possibly classifying some of these as hormone receptor positive or Luminal type supported by BP results. The main strength of the study lies in the reclassification of a subset of patients into the Luminal B and Basal-like subtypes who require more aggressive treatment compared to patients with Luminal A type tumours. Because the 80-gene BP profile is enriched in ER-target genes and measures the functional integrity of ER, it has the potential to identify a subgroup of breast cancer patients who are ER-positive

by IHC and/or single-gene mRNA expression analysis but would fail to respond to hormone treatment.[31]

In conclusion, our results show that in contrast to the added value of TP as a second opinion for HER2 status[22], single-gene microarray readout of ER/PR status provided little additional information beyond that obtained from standard IHC results and performed poorly in predicting molecular tumour subtype. If genomic hormone receptor status alone would be used in clinical decision-making, it is possible that some patients might be erroneously denied endocrine therapy. However, in tumours for which ER or PR expression was lost on both IHC and TP, the molecular subtype determined by BP was less likely to be Luminal A, thus indicating the potential benefit of adjuvant chemotherapy. Basal tumours can potentially be identified by utilising the sum of the ER and PR results in both IHC and mRNA and selecting tumours for which IHC≤270 and mRNA≤0.1, or the flow diagram in Figure 4 can be employed. Accurate distinction between Luminal A and B molecular subtypes, and identification of the Basal-type despite apparent positive ER status on IHC and TP noted in at least one South African patient, were only possible by using the BP 80-gene profile. It was our consistent observation that loss of PR expression on IHC was not indicative of the Luminal B subtype. TP provided limited additional information compared to IHC, which justifies recent discontinuation of single-gene mRNA microarray testing of the ER, PR and HER2 genes now incorporated into the 80-gene BP profile.

The clinical relevance of a pathology-supported genetic testing approach to breast cancer management, combining microarray-based analysis as ancillary to existing clinico-pathological risk stratification and prognostication tools, is supported by the results presented in this study. The routine implementation of genomic profiling alongside standard pathology tests may increase clinician confidence in treatment decision-making and ultimately optimise individualised management of early-stage breast cancer patients by identifying molecular subgroups more accurately.

Acknowledgements

Competing interests

M.J.K. is a director and shareholder of Gknowmix (Pty) Ltd. which developed a database tool for research translation under the auspices of the South African Medical Research Council. The authors have no other relevant affiliations or financial involvement with any organisation or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Authors' contributions

K.A.G. was responsible for data collection, analysis and interpretation, and wrote the first draft of the manuscript. E.J.M. contributed to data acquisition, clinical interpretation and improvement of the database used in this study and revised the manuscript critically prior to submission for publication. E.M. contributed to data acquisition and improvement of the

References

1. Polyak K. Heterogeneity in breast cancer. J Clin Invest. 2011;121(10):3786–3788. https://doi.org/10.1172/JCI60534

2. Perou CM, Sørlie T, Eisen MB, Van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747–752. https://doi.org/10.1038/35021093

3. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res. 2008;14(5):1368–1376. https://doi.org/10.1158/1078-0432.CCR-07-1658

4. Tran B, Bedard PL. Luminal-B breast cancer and novel therapeutic targets. Breast Cancer Res. 2011;13(6), Art. #221, 10 pages. https://doi.org/10.1186/bcr2904

5. Borley A, Mercer T, Morgan M, Dutton P, Barrett-Lee P, Brunelli M, et al. Impact of HER2 copy number in IHC2+/FISH-amplified breast cancer on outcome of adjuvant trastuzumab treatment in a large UK cancer network. Br J Cancer. 2014;15;110(8):2139–2143. https://doi:10.1038/bjc.2014.147

6. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the Basal-like subtype of invasive breast carcinoma. Clin Cancer Res. 2004;10(16):5367–5374. https://doi.org/10.1158/1078-0432.CCR-04-0220

7. Viale G, De Snoo FA, Slaets L, Bogaerts J, Van 't Veer L, Rutgers EJ, et al. Immunohistochemical versus molecular (BluePrint and MammaPrint) subtyping of breast carcinoma. Outcome results from the EORTC 10041/BIG 3-04 MINDACT trial. Breast Cancer Res Treat. 2018;167(1):123–131. https://doi.org/10.1007/s10549-017-4509-9

8. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Progesterone receptor loss identifies Luminal B breast cancer subgroups at higher risk of relapse. Ann Oncol. 2013;24(3):661–668. https://doi.org/10.1093/annonc/mds430

9. Viale G. The current state of breast cancer classification. Ann Oncol. 2012;10:207–210. https://doi.org/10.1093/annonc/mds326

10. Phillips T, Murray G, Wakamiya K, Askaa J, Huang D, Welcher R, et al. Development of standard estrogen and progesterone receptor immunohistochemical assays for selection of patients for anti-hormonal therapy. Appl Immunohistochem Mol Morphol. 2007;15(3):325–331. https://doi.org/10.1097/01.pai.0000213135.16783.bc

11. Whitworth P, Stork-Sloots L, De Snoo FA, Richards P, Rotkis M, Beatty J, et al. Chemosensitivity predicted by BluePrint 80-gene functional subtype and MammaPrint in the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST). Ann Surg Oncol. 2014;21(10):3261–3267. https://doi.org/10.1245/s10434-014-3908-y

12. Yao K, Goldschmidt R, Turk M, Wesseling J, Stork-Sloots L, De Snoo F, et al. Molecular subtyping improves diagnostic stratification of patients with primary breast cancer into prognostically defined risk groups. Breast Cancer Res Treat. 2015;154:81–88. https://doi.org/10.1007/s10549-015-3587-9

13. Roepman P, Horlings HM, Krijgsman O, Kok M, Bueno-de-Mesquita JM, Bender R, et al. Microarray-based determination of estrogen receptor, progesterone receptor, and HER2 receptor status in breast cancer. Clin Cancer Res. 2009;15(22):7003–7011. https://doi.org/10.1158/1078-0432.CCR-09-0449

14. Gevensleben H, Göhring UJ, Büttner R, Heukamp LC, Kunz G, Dimpfl T, et al. Comparison of MammaPrint and TargetPrint results with clinical parameters in German patients with early stage breast cancer. Int J Mol Med. 2010;26(6):837–843. https://doi.org/10.3892/ijmm_00000532

15. Viale G, Slaets L, Bogaerts J, Rutgers E, Van 't Veer L, Piccart-Gebhart MJ, et al. High concordance of protein (by IHC), gene (by FISH; HER2 only), and microarray readout (by TargetPrint) of ER, PgR, and HER2: Results from the EORTC 10041/BIG 03-04 MINDACT trial. Ann Oncol. 2014;25(4):816–823. https://doi.org/10.1093/annonc/mdu026

16. Kraus JA, Dabbs DJ, Beriwal S, Bhargava R. Semi-quantitative immunohistochemical assay versus OncotypeDx® qRT-PCR assay for estrogen and progesterone receptors: An independent quality assurance study. Mod Pathol. 2012;25(6):869–876. https://doi.org/10.1038/modpathol.2011.219

17. Park MM, Ebel JJ, Zhao W, Zynger DL. ER and PR immunohistochemistry and HER2 FISH versus Oncotype DX: Implications for breast cancer treatment. Breast J. 2014;20(1):37–45. https://doi.org/10.1111/tbj.12223

18. Grant KA, Apffelstaedt JP, Wright C, Myburgh E, Pienaar R, De Klerk M, et al: MammaPrint Prescreen Algorithm (MPA) reduces chemotherapy in patients with early stage breast cancer. S Afr Med J. 2013;103(8):522–526. https://doi.org/10.7196/SAMJ.7223

19. Krijgsman O, Roepman P, Zwart W, Carroll JS, Tian S, De Snoo FA, et al. A diagnostic gene profile for molecular subtyping of breast cancer associated with treatment response. Breast Cancer Res Treat. 2012;133(1):37–47. https://doi.org/10.1007/s10549-011-1683-z

20. Glück S, De Snoo F, Peeters J, Stork-Sloots L, Somlo G. Molecular subtyping of early-stage breast cancer identifies a group of patients who do not benefit from neoadjuvant chemotherapy. Breast Cancer Res Treat. 2013;139(3):759–767. http://doi:10.1007/s10549-013- 2572-4

21. Vieira A F, Schmitt F. An update on breast cancer multigene prognostic tests – emergent clinical biomarkers. Front Med. 2018;5, Art. #248, 12 pages. https://doi.org/10.3389/fmed.2018.00248

22. Grant KA, Pienaar FM, Brundyn K, Swart G, Gericke GS, Myburgh EJ, et al. Incorporating microarray assessment of HER2 status in clinical practice supports individualised therapy in early-stage breast cancer. The Breast. 2015;24(2):137–142. https://doi.org/10.1016/j.breast.2014.12.006

23. Khoury T, Yan L, Liu S, Bshara W. Oncotype DX RT-qPCR assay for ER and PR correlation with IHC: A study of 3 different clones. Appl Immunohistochem Mol Morphol. 2015;23(3):178–187. https://doi.org/10.1097/PAI.0000000000000078

24. Hanna MG, Bleiweiss IJ, Nayak A, Jaffer S. Correlation of Oncotype DX Recurrence Score with histomorphology and immunohistochemistry in over 500 patients. Int J Breast Cancer. 2017;2017, Art. #1257078, 6 pages. https://doi.org/10.1155/2017/1257078

25. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics. 2006,7, Art. #278, 10 pages. https://doi.org/10.1186/1471-2164-7-278

26. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159–174.

27. Myburgh EJ, Langenhoven L, Grant KA, Van der Merwe L, Kotze MJ. Clinical overestimation of HER2 positivity in early estrogen and progesterone receptor-positive breast cancer and the value of molecular subtyping using BluePrint. J Glob Oncol. 2016;3(4):314–322. https://doi.org/10.1200/JGO.2016.006072

28. Pohl H, Kotze MJ, Grant KA, Van der Merwe L, Pienaar FM, Apffelstaedt JP, et al. Impact of MammaPrint on clinical decision-making in South African patients with early-stage breast cancer. Breast J. 2016;16,22(4):442–446. https://doi.org/10.1111/tbj.12605

29. Itoh M, Iwamoto T, Matsuoka J, Nogami T, Motoki T, Shien T, et al. Estrogen receptor (ER) mRNA expression and molecular subtype distribution in ER-negative/progesterone receptor-positive breast cancers. Breast Cancer Res Treat. 2014;143(2):403–409. https://doi.org/10.1007/s10549-013-2763-z

30. Bordeaux JM, Cheng H, Welsh AW, Haffty BG, Lannin DR, Wu X, et al. Quantitative in situ measurement of estrogen receptor mRNA predicts response to tamoxifen. PLoS ONE. 2012;7(5), e36559, 8 pages. https://doi.org/10.1371/journal.pone.0036559

31. Groenendijk FH, Zwart W, Floore A, Akbari S, Bernards R. Estrogen receptor splice variants as a potential source of false-positive estrogen receptor status in breast cancer diagnostics. Breast Cancer Res Treat. 2013;140(3):475–484. https://doi.org/10.1007/s10549-013-2648-1

32. Taneri B, Asilmaz E, Gaasterland T. Biomedical impact of splicing mutations revealed through exome sequencing. Mol Med. 2012;30(18):314–319.

33. Welsh AW, Moeder CB, Kumar S, Gershkovich P, Alarid ET, Harigopal M, et al. Standardization of estrogen receptor measurement in breast cancer suggests false-negative results are a function of threshold intensity rather than percentage of positive cells. J Clin Oncol. 2011;29(22):2978–2984. https://doi.org/10.1200/JCO.2010.32.9706

34. Gown AM. Current issues in ER and HER2 testing by IHC in breast cancer. Mod Pathol. 2008;21(Suppl 2):S8–S15. https://doi.org/10.1038/modpathol.2008.34

35. Seferina SC, Nap M, Van den Berkmortel F, Wals J, Voogd AC, Tjan-Heijnen VC. Reliability of receptor assessment on core needle biopsy in breast cancer patients. Tumour Biol. 2013;34(2):987–994. https://doi.org/10.1007/s13277-012-0635-5

36. Dekker TJ, Smit VT, Hooijer GK; Reliability of core needle biopsy for determining ER and HER2 status in breast cancer. Ann Oncol. 2013;24(4):931–937. https://doi.org/10.1093/annonc/mds599

AUTHORS:
Lizette Moolman[1]
Sam M. Ferreira[2]
Angela Gaylard[1]
Dave Zimmerman[3]
Graham I.H. Kerley[4]

AFFILIATIONS:
[1]Scientific Services, South African National Parks, Knysna, South Africa
[2]Scientific Services, South African National Parks, Skukuza, South Africa
[3]Veterinary Wildlife Services, South African National Parks, Port Elizabeth, South Africa
[4]Centre for African Conservation Ecology, Department of Zoology, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Lizette Moolman

EMAIL:
Lizette.moolman@sanparks.org

The decline of the Knysna elephants: Pattern and hypotheses

Understanding and identifying drivers of local population declines are important in mitigating future risks and optimising conservation efforts. The Knysna elephants have, after being afforded protection since the early 1900s, declined to near extinction today. We propose three hypotheses as to why the Knysna elephant population declined. The refugee hypothesis suggests that anthropogenic activities forced the elephants to take refuge in the forest and that the low-quality food acted as the primary driver of decline. The illegal killing hypothesis suggests that the elephants adapted to the forest and its immediate fynbos habitat, with the decline being a consequence of illegal kills. The stochastic founder population hypothesis postulates that the population size and structure left it vulnerable to demographic stochasticity. We critically reviewed available evidence for these hypotheses and found that, although the historical elephant range decline most likely resulted through the refugee hypothesis, the weak demographic and life-history information limits elimination of either of the other hypotheses. We touch on the implications for decision-makers and draw attention to information requirements.

Significance:
- We highlight the knowledge and management challenges which exist for small, threatened populations of which long-term demographic data are sparse.
- We provide the first unbiased evaluation of multiple drivers that may have caused the decline of the Knysna elephants.

Introduction

African elephants, *Loxodonta africana*, are declining across the continent, largely because they are poached for their ivory.[1] In contrast to continental trends, elephant populations in South Africa have increased in recent years.[2] The Knysna elephants, however, are an anomaly. They are the most southern group of elephants in Africa, the only remaining free-ranging elephants in South Africa and represent one of four relict populations in the country.[3] This small population failed to flourish after official protection was afforded in 1908, and its chances for persistence are of concern. Here we review the history of this population and develop and explore hypotheses that may explain why this small population failed to recover after protection.

Based on an estimated 3000 elephants that roamed the Cape Floristic Region in pre-colonial times[4], it is likely that about 1000 elephants occupied the Outeniqua–Tsitsikamma area[5]. Between 1856 and 1886, Knysna experienced a marked influx of people and a boom in development which increased human–elephant conflict, often at a cost to the elephants.[6] During the late 1800s, an estimated 400 to 500 elephants lived in the area[7], but by 1900, only 30–50 individuals were left[3].

During the early 1900s, attitudes shifted from regarding the elephants as a nuisance towards favouring them as local assets.[6] Since then, the question of how many elephants are left in the Knysna forest has prevailed among conservationists, scientists and the public. Numerous survey attempts followed, often accompanied by suggestions on how to recover the population.[7-12]

Management of this population is challenging if the cause of the decline is not clear.[13] A key question is thus, why has the population declined even after official protection since 1908?[9,11,14-17] We propose three hypotheses for the decline. The refugee species hypothesis suggests that human disturbance and encroachment displaced the local elephants from other more optimal habitats into poor-quality habitats of the forests and its surrounding fynbos.[18] The decline then resulted from a limitation of good-quality food.[17] The illegal killing hypothesis suggests that the decline resulted from illegal killing.[11,19] The stochastic founder population hypothesis suggests that the founder population's small size and its structure lead to demographic stochasticity[13] that significantly constrained recovery.

Here we review evidence for and perceptions around these three hypotheses. In addition, we discuss information gaps and the management implications of our findings for decision-making on the future of the Knysna elephants.

The refugee hypothesis

Refugee populations are separated from optimal habitats and are confined to suboptimal habitats. This in turn leads to decreased fitness and smaller populations.[18] The Knysna elephants are candidates for refugee status based on their decline in range, fitness and density.[18]

Researchers have suggested that human settlement and agricultural development taking place in the more open areas confined the southern Cape elephant population's historical range to the southern Cape forests.[17,20,21] As early as 1755, travellers in the southern Cape noted the ongoing shooting of the elephants, which subsequently resulted in the elephants seeking refuge in the forests in the Tsitsikamma area.[22] This record implies that these elephants chose the forests as they provided safety from human disturbances, rather than for nutritional and other life-history needs.

The effects of human disturbances on elephant ranging, movement and distribution patterns and behaviour elsewhere illustrate how elephants avoid such disturbances[23-26], which lead to declines in elephant ranges. Such populations

remain fragmented within their original distribution range, often in suboptimal habitat that may lead to decreases in fitness.[18]

Here we review available evidence relevant to the refugee concept for the Knysna elephant population. We focus on the historical range decline followed by habitat and resource quality as limits of persistence for a refugee elephant population. We also use a once-off reintroduction of elephants as a case study to evaluate the refugee hypothesis.

Safety – historical range decline

The refugee concept suggests that the range of the Knysna elephant population declined from them living in historically open areas to being confined to the forest and its immediate surrounding fynbos habitat. Genetic evidence illustrates that the Knysna elephants once belonged to a single South African population[27], which ranged widely across South Africa from the Cape Peninsula to Limpopo[3,20]. This large historical population occurred in eight different biomes – Fynbos, Succulent Karoo, Desert, Nama-Karoo, Grassland, Savanna, Albany Thicket and Forest.[28] Today, the Knysna elephants have access to only two habitat types – the afro-temperate forest and immediate surrounding fynbos. Their modern-day range, about 185 km² in size, spans the fragmented afro-temperate forest which occurs mostly on the footslopes of the Outeniqua mountain range, interspersed with fynbos and commercial pine plantations.

Historically, elephants occupied open areas of varied habitat types outside of the southern Cape afro-temperate forests. Elephants were observed between Mossel Bay and George in 1497 by Vasco da Gama, in 1601 by Paulus van Caerden[29] and between 1773 and 1776 by other travellers[30]. Local farmers and residents regularly reported elephant sightings east of George in the late 1800s[6] with records at Wittedrift (north of Plettenberg Bay) dated 1782 and 1816[29]. Further away, in the Little Karoo, local newspapers[31] reported a 'troop' of elephants barring the road to the Uitenhage Ostrich Farming Company in 1883. Further east, in the Langkloof, Sparrman recorded elephants close to Humansdorp in 1775, as did Rev C.I. la Trobe in 1816.[29]

Increases in human settlement and population growth, the growing ivory trade, and crop protection exterminated most of South Africa's elephant population between 1652 and 1870.[3] Demand for ivory as well as habitat transformation were the main driving forces for the southern Cape elephants' demise in the 19th century.[32] The expulsion of elephants from open areas outside the southern Cape forests[6,15,17,33] started taking place in the late 1700s[29]. The pressures and dangers that existed for the elephants on the perimeters and in neighbouring areas to the Outeniqua (Knysna) and Tsitsikamma forests were ongoing during the 18th century. Anders Sparrman wrote in 1775 of the continuous shooting of elephants in the George and Knysna areas[29] while Le Vaillant recorded elephant pitfalls at Kaaimansgat, east of George, in 1782[30]. This small population survived most likely because hunters had difficulty in finding them in the forest terrain.[29] These historical accounts suggest that elephants primarily moved into the forest areas to avoid human disturbances.

Food quality

A second element of the refugee hypothesis is that confinement of Knysna elephants to suboptimal habitats affected their fitness and that ultimately led to their decline. The range of modern-day Knysna elephants covers afro-temperate forest, fynbos and commercial pine plantations growing on nutrient-poor soils derived from sandstone.[34] Several studies have attempted to evaluate the link between low nutrients of forest plants and the Knysna elephant decline[15,17,21,35], but with no conclusive support or rejection of the suboptimal habitat concept.

Key studies focused on Knysna elephant faecal and browse mineral contents[15,21,35] and faecal N/C ratios[17]. These studies hypothesised nutrient and mineral deficiencies as the primary drivers of the decline in elephant numbers. However, these assessments were lacking knowledge of what constitutes diet deficiencies that affect fitness for wildlife as most of this information is available mainly for domestic and laboratory animals.[36] Koen and colleagues[15,21,35] used nutrient limits set

for domestic livestock for the Knysna elephant diet assessment, but acknowledged that livestock nutrient limits are unreliable standards for assessing elephant diet deficiencies.

Because of their simple digestive systems, which have relatively low digestive efficiency as a result of fast passage, and their low metabolic rate per unit of body mass, elephants can tolerate lower quality forage than smaller herbivores can, but require a higher abundance of plants.[37] Olff[38] illustrated that, on a global scale, the occurrence of elephant and Cape buffalo was independent of plant-available nutrients and dependent on plant-available moisture. In east and southern Africa, megaherbivore abundance increased with rainfall, but was independent of soil nutrient status and it was suggested that megaherbivores are therefore limited by food quantity and not quality.[39]

A second approach to quantify potential Knysna elephant diet deficiencies and consequences for fitness focused on comparing browse and faecal nutrient and mineral contents with those of fit elephant populations such as those of the Addo Elephant National Park (AENP) and Kruger National Park[15,17,21], both of which have a higher soil nutrient status. Alternatively, more insightful comparisons would be between the Knysna population and healthy, growing populations occurring in other areas of low habitat quality or soil nutrient status e.g. in the Kalahari-sand region of Hwange National Park in Zimbabwe[40] or Tembe Elephant Park in KwaZulu-Natal, South Africa[41]. Sodium concentrations are extremely low throughout elephant ranges[40,42] which suggests that elephants may require much lower concentrations of sodium to maintain condition and survival than commonly perceived[43].

Finally, a key mechanism associated with the low N/C ratio in the Knysna elephant faeces, results from the lack of C4 grasses in the Knysna elephants' forest range.[17] In areas where African elephant populations have access to woodlands, forests and grassland habitats, they prefer grass species during the rainy season and browse species during the dry season.[44,45] Isotope records in historical and prehistoric elephant bones originating from areas in the Western Cape, however, illustrate that these elephants utilised much less C4 grass than elephants in the Kruger National Park, suggesting that C4 grasses cannot be considered a limiting factor in elephant distribution.[46] This finding implies that the lack of C4 grasses in the Knysna elephant range may not be an indicator of suboptimal habitat.

The failure of an introduction of some elephants sourced from Kruger National Park to the Knysna forest was linked to the low quality of the local food[17], which would support the suboptimal habitat concept and refugee hypothesis. This link was based on a prediction that the introduced elephants would bond and remain with the native elephants in the forest. The Kruger elephants, however, rejected the forest habitat and spent most of their time on neighbouring farmland.[16,17] It was suggested that they did so in search of more nutritious food.[17] However, more recent knowledge on elephant behaviour and translocations suggests otherwise (Box 1).

Calf mortality may reflect the outcomes of nutritional stress. If the habitat quality in the Knysna forest and fynbos is inadequate as suggested[15,17,21,35], it may have contributed to the high calf mortality observed. Female elephants under nutritional stress have been shown to have calves which have lower survivorship.[51] A high calf mortality was reported for the Knysna population in the 1900s.[7-10,52,53] Phillips[7] noted for the Knysna population that several calves died between 1922 and 1925 and dead calves were found in 1937, 1942[9] and in 1968[16]. Calf mortality of the Knysna population was estimated to be between 60% and 80%, compared with 7.5% for the AENP population.[53]

Historical records support the refugee hypothesis by illustrating how people forced the southern Cape's elephants out of open areas and into the forest and its surrounding fynbos. A link between the Knysna elephant decline and habitat quality could, however, not be made with the information available.

Box 1: Failed elephant introduction to the Knysna forests

Three cull-orphaned elephant cows aged 7 to 9 years from the Kruger National Park were introduced to the Knysna forest in 1994 to supplement the Knysna elephant population.[16,17] They were released after spending 2 months in an enclosure, and within 1 month the youngest cow had died of stress-related pneumonia.[16,17] The two remaining cows moved to more open fynbos areas within days after their release. Although they made contact with a native elephant cow, after about 7 weeks of following her, they separated and moved back to more open areas, adjacent to the forest area in which the native cow remained. Over the course of 5 years they increasingly moved into open and adjacent farmland areas[17]. In 1999, they were relocated to a private reserve because the responsible authorities were unable to keep them off farmlands.

Whilst still in Knysna, the introduced elephants' faecal N/C ratios were compared to those of the native elephants.[17] The native elephants had low dietary N/C ratios compared to the introduced elephants and it was concluded that this could be the reason for the Knysna elephants' poor reproductive performance as well as the Kruger Park orphans' rejection of the forest habitat. This conclusion was based on available knowledge of elephant social behaviour in natural populations that would predict that the introduced cows would bond and remain with the native cow. It was therefore speculated that the introduced cows separated from the older native cow in search for more nutritious habitats outside of the native elephant forest range that matched their source habitat, in support of the suboptimal habitat concept and refugee hypothesis.

Recent findings, however, illustrate how animals move away from release sites and choose to settle in habitats familiar to their natal habitat[47], in some cases even when the new chosen habitat is of lower quality than the release site[48]. This behaviour is referred to as natal habitat preference induction (NHPI) in which an individual's experience with a natal habitat shapes its future habitat preference.[47,48] NHPI implies that reintroduction and translocation operations may be more likely to succeed if animals are released in habitats comparable to those in which they were raised.[47,48] Elephants have displayed similar behaviours by rejecting their release sites and moving away to settle in a habitat further away, and in some cases even travel back to the source habitat.[49,50] Pinter-Wollman[49] reported female elephants leaving their social group behind at the release site and walking back home, with only their youngest calves in tow.

Rejection of release sights may be perpetuated at locations in which animals have had aversive experiences, as they may associate local cues with negative experiences from their time in captivity, and therefore are more likely to reject habitats containing those cues after release.[47] The cows who were introduced to Knysna experienced a traumatic second immobilisation and collaring operation while in their forest enclosure[16] and the death of one of their group's members whilst in the forest, straight after release. Their rejection of the Knysna forest release site, and subsequent separation from the unrelated native cow, may therefore not have been as unpredicted as suggested[17], and may have been driven by factors other than food quality.

The illegal killing hypothesis

The illegal killing hypothesis suggests that elephants were adapted to the forest and fynbos habitats, and their decline resulted from illegal killing.[11,19] Predictions of forest adaptation focused on attempts to describe the Knysna elephants as a different sub-species.[11,54] Lydekker[54] assigned the Knysna elephant as *Elephas africanus toxotis*, based on comparisons of ear-shapes of specimens from Knysna and AENP, but later discovered that the museum specimen labelled as a Knysna elephant, was in fact a specimen from the Addo area[14]. Carter[11] speculated that although these elephants are *L. africana* and not a sub-species, their 'habits appeared to have become modified for existence in the forest'. This perception of forest adaptation remains, and media, blogging sites and local tourism and property agencies' brochures refer to Knysna's forest elephants. There is, however, no published evidence of genetic, behavioural or morphological adaptations of the Knysna elephants to forest habitats.

Official protection was realised when the Knysna elephants were declared Royal Game in 1908.[17] In 1920, authorities issued Major P.J. Pretorius with a permit to shoot one elephant for scientific purposes – a hunt which ultimately caused the death of five elephants.[16] In 1971, an old bull known as Aftand was shot by forestry officials because of his crop-raiding behaviour on smallholdings.[16] Apart from this shooting, there were no elephant shootings officially reported after 1908.[17]

Although the elephants roamed mostly on forestry land, forestry officials had no authority to protect the elephants when they moved off forestry land, even after they were declared Royal Game.[53] Speculations claimed that the Forestry Department harboured a negative attitude towards the elephants and that 'the total destruction of the elephants was advocated in some quarters' during the 1920s because of the considerable damage they caused to State Blackwood plantings.[8] Later, Carter[11,55] reported the killing of four elephants between 1940 and 1970 and found the remains of a young elephant that had apparently been shot.

Several advocates of the illegal killing hypothesis argued that protection required more effort than mere declaration of protection, such as physical protection on the ground or fencing the elephants.[10,11,14,52,53] The Knysna elephants' crop-raiding excursions[10,11,52,53] often made headlines[56-59]. A major concern was that landowners shot at the elephants to chase them off their land and that these injured animals died only later of their injuries.[9,11] In 1976, a local newspaper[60] reported several requests to the Knysna Department of Forestry to destroy a damage-causing elephant. In addition, indirect and long-lasting effects of traumatic events, such as poaching or attempted poaching, have been reported elsewhere[61] and include social fragmentation, potentially higher calf mortalities and lowered reproductive success.

Similarly, it was suggested that mere declaration of protection of the AENP elephants, with the proclamation of the park in 1931, did not lead to a recovery of the population as wandering elephants were shot on neighbouring farms.[62] Reduction in elephant mortality and population recovery was only realised by erecting a boundary elephant-proof fence 23 years after park proclamation.[53,62] Whitehouse and Hall-Martin[63] identified the loss of the founder population's sexually mature bulls, between 1932 and 1940 as another factor that caused the initial slow growth rate of AENP's elephant population. The remaining young bulls reached sexual maturity by 1946, after which the population started its recovery.[62,63]

The available evidence highlighted here does not allow for a robust evaluation of the illegal killing hypothesis. However, it is likely that some illegal killing contributed to the slow demise of the Knysna elephants.

The stochastic founder hypothesis

Demographic stochasticity is a fluctuation in a population's growth rate caused by chance independent events of individual mortality and reproduction.[13,64] With regard to extinction risk, small populations are particularly vulnerable to demographic stochasticity.[13,64] The stochastic founder hypothesis suggests that after protection, the small population size and structure of the Knysna elephants exposed them to demographic stochasticity.

Between 1920 and 1970, it was estimated that four cows of breeding age were present in the Knysna elephant population.[17] Forestry scientists argued that more than the observed two calves should therefore have been present if breeding was normal and suboptimal habitat did not reduce fertility.[17] However, available information on the Knysna elephant population size, structure and individual life histories is vague or lacking. Between 1920 and 1970, a number of surveys recorded only one or two adult cows[8,9,52], while Carter's[11] survey concluded the presence of three adult cows (Table 1). The assumption of four sexually mature cows present during this time[17] may therefore be an overestimate. In addition, even if there were four adult cows between 1920 and 1970, not all of them may have been fertile.

The differences between the Knysna and AENP elephants' founder populations lend credibility to the stochastic founder population hypothesis by illustrating how the Knysna population's small size and structure likely enhanced chance events that determined its fate. The Knysna elephant founder population consisted of 1–4 adult cows of unknown reproductive status, whereas the AENP founder population consisted of six sexually mature adult cows.[63] By 1935/1936, eight sexually mature cows were present and eight calves were born during this time.[63]

Another difference between the populations is that the AENP cows seemed to have existed as one social group[63] whereas the Knysna elephants were mostly sighted in small groups, containing only one adult cow and her offspring[8,11,19,52]. Such minimal group sizes may have contributed to the high calf mortality reported for the Knysna population.[7-10,52,53] Elephants live in matriarchal societies consisting of family units of related adult females and their calves[66,67] in which

Table 1: Knysna elephant population size estimates from the late 1800s to 2007

Year	Elephant number	Elephant group structure	Nature of estimate/survey methods	Source
1876	400–500			Phillips, 1925[7]
1902	30–50		Forestry Department records	Dommisse, 1951[8]
1904	20		Forestry Department records	Dommisse, 1951[8]
1908	20		Forestry Department records	Dommisse, 1951[8]
1910	17	15 'large' and 2 'small' elephants of unknown sex	Forestry Department records	Dommisse, 1951[8]
1914	13		Forestry Department records	Dommisse, 1951[8]
1916	10–12		Forestry Department records	Koen, 1984[35]
1918	15/16		Forestry Department records	Koen, 1984[35]
1920	7–13		Forestry Department record less 5 killed by Major Pretorius	Dommisse, 1951[8]
1921	12		Forestry Department records	Koen, 1984[35]
1925	12	6 bulls, 4 'large' cows and 2 calves of unknown sex		Phillips, 1925[7]
1931	13		Forestry Department records	Koen, 1984[35]
1951	4	2 adult bulls, 1 cow and 1 young bull	Bernard Carp Expedition	Woods, 1952[9]
1957	7	4 adult bulls, 1 old bull, 1 adult cow and 1 calf	Cape Department of Nature Conservation Expedition (Fraser Expedition)	Woods, 1958[65]
1958	3	2 adults of unknown sex and 1 female calf	Wildlife Society Expedition	Greig, 1982[14]
1968	7		Wildlife Society Expedition	Greig, 1982[14]
1970	11	3 old bulls, 2 young bulls, 3 cows, 1 adult of unknown sex, 2 calves	Wildlife Society Survey	Carter, 1970[11]
1970	13	3 old bulls, 2 adult bulls, 1 young bull, 1 old cow, 4 adult cows, 2 calves	Forestry Department records	Koen, 1984[35]
1974	6		Forestry Department records	Koen, 1984[35]
1976	4		Forestry Department records	Koen, 1984[35]
1977	6		Stroebel family records	Koen, 1984[35]
1979	4		Forestry Department records	Koen, 1984[35]
1981	3		Forestry Department records	Koen, 1984[35]
1989	4		Forestry Department records plus 1 (birth of calf)	Mackay, 1996[16]
2007	5	5 cows	Faecal DNA genotyping survey	Eggert et al., 2007[12]
2018	1	1 cow	Camera trap survey	Moolman et al., 2019[75]

allomothering enhances survivorship[68]. One can therefore speculate that if there were sexually mature bulls and cows present between 1920 and the late 1900s, and the cows were mostly split into small groups consisting of an adult cow and her calf, the lack of allomothering may have played a significant role in reduced calf survival. Additionally, small group sizes may cause higher stress levels in African elephant cows which subsequently could lead to lowered reproductive outputs.[61]

It is highly likely that demographic stochasticity played a role in the inability of the small founder Knysna elephant population to recover after protection had been afforded.

Synthesis

Historical accounts illustrate how the southern Cape elephants were decimated by humans in the more open areas outside the afro-temperate forests, with the survivors taking up refuge in the forests during the late 1700s and 1800s. The elephant range thus decreased from a diverse array of habitats to only two habitat types: the afro-temperate forest and the surrounding fynbos. Ever-increasing human development eventually confined these elephants to the forests and isolated them from previously available habitat types. Range restriction – being one of the predictions of the refugee hypothesis – was thus realised.

Why the small surviving group of elephants in the Knysna forest never recovered, even after being afforded protection through their declaration of Royal Game in 1908, is less clear. Most of the scientific undertakings attempted to answer this question by linking habitat quality and population fitness by using density estimations obtained during the 1900s. Apart from the unreliability of the Knysna elephant density figures[52], density can be a misleading indicator of habitat quality without demographic data to validate it[18,69]. The importance of demographic data in understanding processes that influence elephant populations has

been clearly illustrated.[70] In addition, habitat quality has been assessed in terms of food quality only, and other habitat aspects, for example, the availability of areas for socialising, that may be important to elephant, have been ignored. In light of this, we argue that the published works to date are too weak to link habitat quality to fitness, and the refugee hypothesis can thus not be eliminated.

Records of conflict between humans and the small surviving group of elephants on land neighbouring the Knysna forests suggest that landowners occasionally shot at elephants. In addition, some locals were of the opinion that the Forestry Department, in charge of the elephants' protection, harboured a negative attitude towards the elephants at the time. Besides the recorded Major Pretorius hunt, which saw five elephants killed in 1920, and the killing of a bull in 1971 by the Forestry Department, no records of killings are available. In addition, once again, demographic data are lacking. The illegal killing hypothesis can therefore not be eliminated.

The comparative records of the AENP and Knysna elephant founder populations' group structures reflect on potential stochastic effects. The small size and fragmented structure of the founder population may have played a significant role in the population decline, exacerbated by human disturbances such as illegal killings. We can thus not eliminate the stochastic effect hypothesis.

The results suggest that the failure of Knysna elephants to recover result from synergistic mechanisms.[71] After the decimation of the southern Cape population in the 1700 and 1800s, the surviving population experienced a range restriction and took up refuge in the forests. Recovery of this refugee population was challenged because the habitat was sub-optimal. Conflict with humans, including illegal killings, most likely imposed an additional stressor on recovery. It is likely that small population stochastic effects accentuated the consequences of illegal killing and suboptimal habitats.

The Knysna elephant population exhibits similarities to other wildlife populations identified as refugee populations.[18,72] Individuals of these refugee populations elsewhere had no access to historically optimal habitats and were subsequently managed and confined to suboptimal habitats. In these cases, population fitness was reduced with consequent declines in numbers. A major conservation risk is when authorities actively manage such populations under the false perception that the focal population occupies suitable habitats. Such perceptions restrict the development of appropriate mitigating management strategies.[18]

Determining the Knysna elephants' refugee status would therefore be a crucial and responsible consideration for their future short- and long-term management planning. The challenge, however, is the limited demographical data available as illustrated throughout this review. To overcome this challenge, we propose that an alternative approach be to investigate how the southern Cape elephants used the landscape before human barriers denied them the choice and access to more suitable habitat. The estuarine lake system, which has a higher soil nutrient status compared to the forest and fynbos habitats[73] and which occurs along the southern Cape coast only 15–30 km south of the elephants' current range, could potentially be a more suitable habitat. These lake systems became cut-off from the current elephant range by farmlands, a national highway and towns. A scientific investigation into the historical southern Cape elephant population's seasonal habitat use patterns and potential diet shifts could be undertaken by analysing the stable isotopes of elements such as carbon, nitrogen and oxygen along ivory growth trajectories[74] and the Knysna elephant population refugee status can thereby be robustly tested.

Management implications

The synergistic mechanisms of decline that we highlight have three implications for management. If elephant presence is desirable in the Knysna forest, the recovery of elephants requires innovative elephant reintroductions from elsewhere, as the forests have been isolated from potential elephant source populations for a long time and constructed barriers will prevent natural colonisation. A primary challenge is how to reintroduce elephants given the failure of the previous reintroduction.[16,17]

However, reintroductions will overcome the predictions of the stochastic small population hypothesis. Furthermore, the persistence of elephants will require regular supplementary introductions to overcome the predictions of the suboptimal habitat element of the refugee hypothesis. Additionally, the availability of optimal habitat types, close to the Knysna elephant range, and the possibilities of range expansion, should be investigated. In addition, authorities will need to implement effective anti-poaching and law enforcement programmes to overcome the predictions of the illegal killing hypothesis.

Authors' contributions

L.M.: Conceptualisation, data collection, writing the initial draft, writing revisions, preparing the submitted draft. S.M.F.: Conceptualisation, writing revisions. A.G. and D.Z.: Writing revisions. G.I.H.K.: Conceptualisation, writing the initial draft, writing revisions, project leadership.

References

1. Robson AS, Trimble MJ, Purdon A, Young-Overton KD, Pimm SL, Van Aarde RJ. Savanna elephant numbers are only a quarter of their expected values. PLoS ONE. 2017;12(4), e0175942, 14 pages. https://doi.org/10.1371/journal.pone.0175942

2. Chase MJ, Schlossberg S, Griffin CR, Bouché PJ, Djene SW, Elkan PW, et al. Continent-wide survey reveals massive decline in African savannah elephants. PeerJ. 2016;4, e2354, 24 pages. https://doi.org/10.7717/peerj.2354

3. Hall-Martin AJ. Distribution and status of the African elephant *Loxodonta africana* in South Africa, 1652–1992. Koedoe. 1992;35(1):65–88. https://doi.org/10.4102/koedoe.v35i1.390

4. Kerley GIH, Pressey RL, Cowling RM, Boshoff AF, Sims-Castley R. Options for the conservation of large and medium sized mammals in the Cape Floristic Region hotspot, South Africa. Biol Conserv. 2003;112:169–190. https://doi.org/10.1016/S0006-3207(02)00426-3

5. Boshoff AF, Kerley GI, Cowling RM. Estimated spatial requirements of the medium-to large-sized mammals, according to broad habitat units, in the Cape Floristic Region, South Africa. Afr J Range For Sci. 2002;19:29–44. https://doi.org/10.2989/10220110209485772

6. Roche CJ. 'The Elephants at Knysna' and 'The Knysna Elephants': From Exploitation to conservation: Man and elephants at Knysna 1856-1920 [Hons thesis]. Cape Town: University of Cape Town; 1996.

7. Phillips JF. The Knysna elephant: A brief note on their history and habits. S Afr J Sci. 1925;22:287–293.

8. Dommisse EJ. The Knysna elephants, historical sketch of a world-famous herd. Afr Wildl. 1951;5:195–199.

9. Woods DH. Report on the Bernard Carp Knysna elephant expedition. Afr Wildl. 1952;6:6–7.

10. Hey D. The elephants of the Knysna forest. Afr Wildl. 1962;16:101–108.

11. Carter B. Knysna elephant survey, February 1969-January 1970. Port Elizabeth: Wild Life Protection and Conservation Society of South Africa, Eastern Province Branch; 1970.

12. Eggert LS, Patterson G, Maldonado JE. The Knysna elephants: A population study conducted using faecal DNA. Afr J Ecol. 2007;46:19–23. https://doi.org/10.1111/j.1365-2028.2007.00794.x

13. Caughley G. Directions in conservation biology. J Anim Ecol. 1994;63:215–244. https://doi.org/10.2307/5542

14. Greig JC. Are the Knysna elephants a distinct race? Afr Wildl. 1982;36:210–215.

15. Koen JH, Hall-Martin AJ, Erasmus T. Macro nutrients in plants available to the Knysna, Addo, and Kruger National Park elephants. S Afr J Wildl Res. 1988;18:69–71.

16. Mackay M. The Knysna elephants and their forest home. Knysna: Wildlife and Environment Society of South Africa; 1996.

17. Seydack AH, Vermeulen C, Huisamen J. Habitat quality and the decline of an African elephant population: Implications for conservation. S Afr J Wildl Res. 2000;30:34–42.

18. Kerley GIH, Kowalczyk R, Cromsigt JP. Conservation implications of the refugee species concept and the European bison: King of the forest or refugee in a marginal habitat? Ecography. 2012;35:519–529. https://doi.org/10.1111/j.1600-0587.2011.07146.x

19. Carter B. The elephants of Knysna. Cape Town: Purnell; 1971.

20. Fairall N. The Knysna elephants – a non-issue. Afr Wildl. 1982;36:197.

21. Koen JH. Trace elements and some other nutrients in the diet of the Knysna elephants. S Afr J Wildl Res. 1988;18:109–110.

22. Smithers RHN. The mammals of the southern African subregion. Pretoria: University of Pretoria; 1983.

23. Granados A, Weladji RB, Loomis MR. Movement and occurrence of two elephant herds in a human-dominated landscape, the Bénoué Wildlife Conservation Area, Cameroon. Trop Conserv Sci. 2012;5:150–162. https://doi.org/10.1177/194008291200500205

24. Jachowski DS, Slotow R, Millspaugh JJ. Physiological stress and refuge behavior by African elephants. PLoS ONE. 2012;7, e31818, 11 pages. https://doi.org/10.1371/journal.pone.0031818

25. Gunn J, Hawkins D, Barnes RF, Mofulu F, Grant RA, Norton GW. The influence of lunar cycles on crop-raiding elephants; evidence for risk avoidance. Afr J Ecol. 2013;52:129–137. https://doi.org/10.1111/aje.12091

26. Murai M, Ruffler H, Berlemont A, Campbell G, Esono F, Agbor A, et al. Priority areas for large mammal conservation in Equatorial Guinea. PLoS ONE. 2013;8, e75024, 13 pages. https://doi.org/10.1371/journal.pone.0075024

27. Essop MF, Hall-Martin AJ, Harley EH. Mitochondrial DNA analysis of two southern African elephant populations. Koedoe. 1996;39:85–88. https://doi.org/10.4102/koedoe.v39i1.284

28. Boshoff AF, Landman M, Kerley GIH. Filling the gaps on the maps: Historical distribution patterns of some larger mammals in part of southern Africa. Trans R Soc S Afr. 2016;71:23–87. https://doi.org/10.1080/0035919X.2015.1084066

29. Skead CJ. Historical incidence of the larger land mammals in the broader Western and Northern Cape. 2nd ed. Port Elizabeth: Centre for African Conservation Ecology, Nelson Mandela Metropolitan University; 2011.

30. Rookmaaker LC. The zoological exploration of southern Africa 1650–1790. Rotterdam: AA Balkema; 1989.

31. Elephant again. The George and Knysna Herald. 1883 May 23; no. 74:3 (col. 3).

32. Carruthers J, Boshoff AF, Slotow R, Biggs H, Avery G, Matthews W. The elephant in South Africa: History and distribution. In: Scholes RJ, Mennell KG, editors. Elephant management: A scientific assessment for South Africa. Johannesburg: Wits University Press, 2008; p. 23–83.

33. Skinner JD, Chimimba CT. The mammals of the southern African sub-region. Cambridge: Cambridge University Press; 2005. https://doi.org/10.1017/CBO9781107340992

34. Van Daalen JC. Distinguishing features of forest species on nutrient-poor soils in the Southern Cape. Bothalia. 1984;15:29–39. https://doi.org/10.4102/abc.v15i1/2.1123

35. Koen JH. A study of the distribution, population composition, movements and feeding of the Knysna elephants Loxodonta africana africana (Blumebach 1797). Knysna: Forestry Department, 1984.

36. Robbins CT. Wildlife feeding and nutrition. New York: Academic Press; 1983.

37. Illius AW, Gordon IJ. Modelling the nutritional ecology of ungulate herbivores: Evolution of body size and competitive interactions. Oecologia. 1992;89:428–434. https://doi.org/10.1007/BF00317422

38. Olff H, Ritchie ME, Prins HH. Global environmental controls of diversity in large herbivores. Nature. 2002;415:901–904. https://doi.org/10.1038/415901a

39. Fritz H, Duncan P, Gordon IJ, Illius AW. Megaherbivores influence trophic guilds structure in African ungulate communities. Oecologia. 2002;131:620–625. https://doi.org/10.1007/s00442-002-0919-3

40. Holdø RM, Dudley JP, McDowell LR. Geophagy in the African elephant in relation to availability of dietary sodium. J Mammal. 2002;83:652–664. https://doi.org/10.1644/1545-1542(2002)083<0652:GITAEI>2.0.CO;2

41. Matthews WS, Van Wyk AE, Van Rooyen N, Botha GA. Vegetation of the Tembe Elephant Park, Maputaland, South Africa. S Afr J Bot. 2001;67:573–594. https://doi.org/10.1016/S0254-6299(15)31188-1

42. Weir JS. Spatial distribution of elephants in an African National Park in relation to environmental sodium. Oikos. 1972;23:1–13. https://doi.org/10.2307/3543921

43. Rode KD, Chiyo PI, Chapman CA, McDowell LR. Nutritional ecology of elephants in Kibale National Park, Uganda, and its relationship with crop-raiding behaviour. J Trop Ecol. 2006;22:441–449. https://doi.org/10.1017/S0266467406003233

44. Kabigumila J. Feeding habits of elephants in Ngorongoro Crater, Tanzania. Afr J Ecol. 1993;31:156–164. https://doi.org/10.1111/j.1365-2028.1993.tb00528.x

45. De Boer WF, Ntumi CP, Correia AU, Mafuca JM. Diet and distribution of elephant in the Maputo Elephant Reserve, Mozambique. Afr J Ecol. 2000;38:188–201. https://doi.org/10.1046/j.1365-2028.2000.00243.x

46. Radloff FGT. The ecology of large herbivores native to the coastal lowlands of the Fynbos Biome in the Western Cape, South Africa [PhD thesis]. Stellenbosch: Stellenbosch University; 2008.

47. Stamps JA, Swaisgood RR. Someplace like home: Experience, habitat selection and conservation biology. Appl Anim Behav Sci. 2007;102:392–409. https://doi.org/10.1016/j.applanim.2006.05.038

48. Davis JM, Stamps JA. The effect of natal experience on habitat preferences. Trends Ecol Evol. 2004;19:411–416. https://doi.org/10.1016/j.tree.2004.04.006

49. Pinter-Wollman N. Spatial behaviour of translocated African elephants (Loxodonta africana) in a novel environment: using behaviour to inform conservation actions. Behaviour. 2009;146:1171–1192. https://doi.org/10.1163/156853909X413105

50. Fernando P, Leimgruber P, Prasad T, Pastorini J. Problem-elephant translocation: translocating the problem and the elephant? PLoS ONE. 2012;7, e50917, 9 pages. https://doi.org/10.1371/journal.pone.0050917

51. Lee PC, Moss CJ. Early maternal investment in male and female African elephant calves. Behav Ecol Sociobiol. 1986;18:353–361. https://doi.org/10.1007/BF00299666

52. Fraser A. Report: Survey of elephant population of the Knysna Forest. Knysna: South African National Park; 1957.

53. Hall-Martin A. Elephant survivors. Oryx. 1980;15:355–362. https://doi.org/10.1017/S0030605300028830

54. Lydekker R. The ears as a race-character in the African elephant. Proc Zool Soc Lond. 1907;77:380–403. https://doi.org/10.1111/j.1096-3642.1907.tb01824.x

55. Carter B. Counting the Knysna elephants. Afr Wildlife. 1969;23:279–286.

56. Elephants again. The George and Knysna Herald. 1957 February 01; no. 3830:6 (col. 7).

57. Olifante eis eerste slagoffer [Elephants claim first victim]. The George and Knysna Herald. 1963 June 21; no. 3070:4 (col. 5). Afrikaans.

58. No elephant hunt at Knysna. The George and Knysna Herald. 1964 September 11;1 (col. 3).

59. How to get rid of an elephant. The George and Knysna Herald. 1970 July 17;1 (col. 5).

60. Request to kill Knysna elephant. EP Herald. 1967 October 20.

61. Gobush KS, Mutayoba BM, Wasser SK. Long-term impacts of poaching on relatedness, stress physiology, and reproductive output of adult female African elephants. Conserv Biol. 2008;22:1590–1599. https://doi.org/10.1111/j.1523-1739.2008.01035.x

62. Whitehouse AM. Tusklessness in the elephant population of the Addo Elephant National Park, South Africa. J Zool. 2002;257:249–254. https://doi.org/10.1017/S0952836902000845

63. Whitehouse AM, Hall-Martin AJ. Elephants in Addo Elephant National Park, South Africa: Reconstruction of the population's history. Oryx. 2000;34:46–55. https://doi.org/10.1017/S0030605300030891

64. Lee AM, Sæther BE, Engen S. Demographic stochasticity, allee effects, and extinction: The influence of mating system and sex ratio. Am Nat. 2011;177:301–313. https://doi.org/10.1086/658344

65. Woods DH. The Knysna elephants. Afr Wildlife. 1958;12:118–124.

66. Moss CJ. The demography of an African elephant (*Loxodonta africana*) population in Amboseli, Kenya. J Zool. 2001;255:145–156. https://doi.org/10.1017/S0952836901001212

67. Archie EA, Moss CJ, Alberts SC. The ties that bind: Genetic relatedness predicts the fission and fusion of social groups in wild African elephants. Proc R Soc Lond B Biol Sci. 2006;273:513–522. https://doi:10.1098/rspb.2005.3361

68. Lee PC. Allomothering among African elephants. Anim Behav. 1987;35:278–291. https://doi.org/10.1016/S0003-3472(87)80234-8

69. Van Horne B. Density as a misleading indicator of habitat quality. J Wildl Manage. 1983;47:893–901. https://doi.org/10.2307/3808148

70. Wittemyer G, Daballen D, Douglas-Hamilton I. Comparative demography of an at-risk African elephant population. PLoS ONE. 2013;8, e53726, 10 pages. https://doi.org/10.1371/journal.pone.0053726

71. Brook BW, Sodhi NS, Bradshaw CJ. Synergies among extinction drivers under global change. Trends Ecol Evol. 2008;23:453–460. https://doi.org/10.1016/j.tree.2008.03.011

72. Lea JM, Kerley GI, Hrabar H, Barry TJ, Shultz S. Recognition and management of ecological refugees: A case study of the Cape mountain zebra. Biol Conserv. 2016;203:207–215. https://doi.org/10.1016/j.biocon.2016.09.017

73. Whitfield AK, Lubke RA. Estuarine and coastal lake ecology. In: Lubke RA, De Moor I, editors. Field guide to the eastern and southern Cape coasts. Cape Town: University of Cape Town Press; 1998. p. 216–232.

74. Codron J. Annals of ivory: Perspectives on African elephant *Loxodonta africana* (Blumenbach 1797) feeding ecology from a multi-decadal record [PhD thesis]. Cape Town: University of Cape Town; 2008.

75. Moolman L, De Morney MA, Ferreira SM, Ganswindt A, Poole JH, Kerley GI. And then there was one: A camera trap survey of the declining population of African elephants in Knysna, South Africa. Afr J Wildl Res. 2019;49(1):16–26.

AUTHORS:
Brent Abrahams[1] iD
Nadia Sitas[2] iD
Karen J. Esler[1] iD

AFFILIATIONS:
[1]Centre for Invasion Biology, Department of Conservation Ecology and Entomology, Stellenbosch University, Stellenbosch, South Africa
[2]Natural Resources and the Environment, Council for Scientific and Industrial Research, Stellenbosch, South Africa

CORRESPONDENCE TO:
Brent Abrahams

EMAIL:
babrahams@sun.ac.za

Exploring integrative research in the context of invasive alien plant management

Addressing complex challenges facing social-ecological systems (SES) requires the integration of knowledge from a diversity of disciplines and stakeholders. This requirement has resulted in the establishment of many integrative research programmes, both globally and locally, aimed at co-producing knowledge relevant to solving SES challenges. However, despite the increase in integrative projects, there has been little research on the nature and extent to which these projects acknowledge and integrate information from diverse disciplines or knowledge types. In this study, we explored the extent to which the integration of different disciplines has occurred, using a case study of the South African invasive species management programme, Working for Water (WfW). Here we provide an overview of the research produced under the auspices of WfW, and how it came to be. Additionally, we assess the extent to which research associated with the programme addressed the research priorities and how these priorities relate to one another. Findings show that WfW-associated research is primarily focused on the ecological processes and impacts associated with invasive alien plants and biological control. Social science and applied research are, however, under-represented, infrequent in nature and inadequately address the research priorities set in the programme. To address these shortcomings, we recommend the development of a detailed research strategy and action plan conducive to integrative research and transdisciplinary collaboration, and relevant to solving complex SES challenges such as those associated with biological invasions.

Significance:
- We provide a reference point by which we can assess research progress and guide integration of diverse knowledge systems.
- The results can help guide research decision-making as it relates to invasive species management.

Introduction

Humans have altered the structure and function of many ecosystems, with negative impacts upon the production and flow of ecosystem services and associated impacts on human well-being.[1] Accordingly, there have been strong calls for integrated solutions that address both societal and environmental needs, such as those associated with water and food security, biodiversity loss and land degradation.[1-5] In response to these needs, several integrative research programmes have been formed, which range from local studies[5-7] to global initiatives.[4,8,9] Many researchers agree that these sustainability challenges require new approaches to knowledge production to ensure that decision-making processes are more transformative.[10-13] Consequently, calls for an integrated approach that incorporates diverse knowledge to address complex social-ecological problems have increased markedly in recent years.[10-13]

Research projects focusing on social-ecological systems (SES) attempt to foster interdisciplinary or transdisciplinary approaches to research planning and practice, and in doing so to co-produce solutions to sustainability challenges, thereby addressing both societal and environmental needs.[3,13] These programmes and projects recognise the importance of integrated research and that the co-production of knowledge is necessary for the generation of new understandings of SES, including: insights into changes in ecosystem services and their societal implications, ecosystem-based research strategies, and exploring new ways of conducting integrative research. Turner et al.[13] reviewed several SES projects undertaken over the last few decades from which lessons promoting integrated project success and challenges facing such projects were synthesised. Turner et al.[13] argue that to effectively address SES challenges, integrative research is required to account for a plurality of perspectives and sources of knowledge.[13-15] Such research should integrate diverse knowledge streams and systems.[13] These knowledge systems consist of actors, practices and institutions that combine the production, transfer and use of knowledge to address challenges.

Integrated science directed towards SES challenges involves expertise from diverse disciplines, and non-disciplinary experts (e.g. local or indigenous knowledge holders) collaborating to unravel the impacts and dynamics of sustainability challenges.[13] In attempting to bridge multiple knowledge systems, this approach attempts to rethink interactions between nature and society and science and democracy, across multiple domains and scales.[3,15] Evidence stemmed from scientific research has long been seen as a legitimate way to influence policy and depoliticise questions that should rightfully be subject to public deliberation.[16-18] However, without grounding such research in social processes, it is unlikely that the research will have the desired outcome.[10] Repeated calls have therefore been made for a new research paradigm that involves greater responsiveness to societal needs in choosing priorities.[13] This paradigm emphasises the need for setting explicit goals for producing practicable knowledge and implementing integrative SES research approaches, as well as for coordination with policymakers and agencies to incorporate evidence into policy processes and for more emphasis on long-term, place-based monitoring and analysis of SES.[13] Research produced in such a context would include that of an integrative, inter- or transdisciplinary nature.

Here, we explore the extent to which research produced under the long-running South African invasive alien plant (IAP) management programme Working for Water (WfW) has integrated various disciplinary insights towards achieving its mandates. This SES programme seeks to promote conservation in parallel with poverty alleviation through invasive alien plant control projects.[5] The need for IAP control and removal was drawn from a science-based realisation that if IAPs are left uncontrolled, they will have a significant negative impact on water resources.[5] Launched in 1995, this multifaceted flagship programme aimed to restore water lost to IAPs, and conserve biological diversity, ecological integrity and catchment stability while simultaneously empowering individuals through employment creation and community building.[19]

Despite the importance of scientific evidence in justifying the establishment of the WfW programme[20], external evaluations (and key research publications emerging over this time) in 1997 and 2003 were critical of WfW's lack of a dedicated research plan.[19,21] The reviews emphasised the need for a multidisciplinary, action-oriented approach to research, highlighting that for the programme to successfully achieve its mandate and meet its objectives, an improved understanding of all the aspects affecting the programme's activities would be required.[19,21-24] Consequently, WfW allocated funding amounting to approximately ZAR15 million per annum (of the ZAR397 million annual budget) over the period 2001–2003 to conduct research, the findings of which were presented at an inaugural research symposium in 2003, and published in a special issue of the *South African Journal of Science*.[23,25] This special issue provided the basis for developing the detailed research strategy and action plan (RSAP) adopted by WfW in 2005.[26] The research strategy highlighted and reiterated three key points throughout its formulation[26]:

> *(1) the WfW research programme is only justifiable if the research conducted is directed towards enhancing the efficiency and effectiveness of the overall programme; (2) any research conducted under its auspices would be held to the same standards of total accountability, commitment to transformation, and social responsibility that underpins the programme as a whole; and (3) the programme's research effort would be an integral part of the overall adaptive management approach that informed all aspects of the programme's implementation.*

The WfW research strategy also called for all research conducted under its auspices to be peer reviewed and published.[26] The national strategy for dealing with IAP[27] reiterates this point and proposes that the number of publications in journals indexed in Clarivate Analytic's Web of Science, and their citation counts, be used as indicators to evaluate research effort and its visibility. Ensuring that research is peer reviewed was stated as being crucial to the overall research management process.[26]

Against this background, we analysed the extent to which peer-reviewed research outputs of this large SES programme demonstrate the integration of diverse disciplinary insights. We selected the WfW as a case study because of its longevity and because it could provide us with insight into what research has been produced under one of the largest global invasive species programmes, the design of which is comparable to a large SES programme.[13] In order to add to the body of literature and conceptual understanding of the role of research in informing and shaping complex SES programmes, in this study we made use of qualitative content analysis to: (1) provide an overview of the research produced under the auspices of the programme since 1995; and (2) assess the extent to which the research output reflects the integration of diverse disciplines and knowledge types, and aligns with the aims, mission, mandate and research priorities set by the programme.

Materials and methods

We conducted a content analysis of WfW-associated research articles, published from 1995 to 2015, in journals listed in Web of Science. The full-texts were collected from the Stellenbosch University library services database. WfW-related research articles were selected based on the presence of the term 'Working for Water', its synonyms or alternatives (i.e. 'working-for-water', 'working for water programme', 'working for water programme', 'WfW', 'WfWP') in a paper's abstract, title and/or keywords, or where WfW was acknowledged in funding texts and is searchable in Web of Science. We are aware that WfW and IAP management related issues may be addressed in grey literature and in publications that are not on the Web of Science. However, here we focused on this subset of peer-reviewed literature, firstly, because the South African Department of Environmental Affairs proposed Web of Science indexed publications as an indicator for the evaluation and assessment of research associated with biological invasions[27], and, secondly, because of the logistical difficulties of locating, identifying and comprehensively covering the diversity of peer review and grey literature relating to WfW.

While the Web of Science has been used in many systematic reviews, it does not provide full coverage of scientific outputs[28-30], with natural and physical science disciplines being better represented than the social sciences and humanities.[28,31] Thus, it is possible that the use of papers in Web of Science indexed journals as a research indicator for WfW could bias results in favour of the natural and physical sciences. Nevertheless, despite its shortcomings in social science coverage, Web of Science has the best historical coverage, its functionality and sophistication exceed that of other databases[28,32], and it remains one of the more reliable tools for evaluating research.

The resulting 255 articles were subjected to qualitative content analysis – a method used for the analysis of written, verbal and visual communications to describe and quantify phenomena.[33] We used both inductive and deductive approaches to classify the papers. For the inductive approach, we used the content of the papers to decide on categories of research, while in the deductive approach, we assigned papers to pre-determined categories of research.[33] Combining approaches can enable confirmation or corroboration of findings through triangulation, enrich data and/or initiate new modes of thinking by addressing ambiguities emerging from the two data sources.[34]

Specifically, the latent content, or the underlying meaning of the articles, was coded inductively by reading each article in its entirety and making an overall assessment of its primary emphasis into basic themes.[35,36] Through the abstraction process, broader research categories were formulated.[33] The emergent categories are therefore exhaustive and mutually exclusive, with no single article grouped under more than one category. In cases in which articles contained materials relating to more than one category, each article was categorised according to its primary emphasis, as determined by the title and/or abstract content. Articles were then further categorised using a more deductive approach according to research areas and priorities obtained from the 2005 WfW research strategy[26], using the definitions and descriptions provided within the research strategy. In the cases in which articles covered more than one topic in detail (primary or secondary focus), they were assigned to multiple categories. In this instance, categories refer to the research areas and priorities obtained and adapted from the 2005 WfW RSAP.[26] Because publications may cover multiple categories, data were treated as multiple response data and analysed accordingly using IBM SPSS 23. A correspondence analysis was performed to examine the relationship between these categories and visualised using XLStat 2016, and included supplementary variables relating to funding. Correspondence analysis is a method that is used to describe and visualise relationships between several variables and categories.[37]

Results

Information sources

A total of 255 academic peer-reviewed publications were identified from Web of Science, of which 229 were original articles and 26 were review articles. There were no publications for the years 1995 (the year the programme was initiated) and 1996 (Figure 1). A sum of 217 publications were funded either in part or in full by WfW. Approximately 50% of all the publications were published in 5 of the 75 journals recorded, namely (in descending order with the number of publications

in parentheses): *African Entomology* (41), *South African Journal of Science* (27), *Diversity and Distributions* (20), *South African Journal of Botany* (20) and *Biocontrol* (18). The first four journals each published a related special issue (Figure 2). Of the journals represented in the data set, 11 are South African and together comprise 106 publications

(41.6%). Of these 11 journals, 2 are listed on the Social Science Citation Index and 9 on the Science Citation Index Expanded, with these indexes accounting for 2 and 104 publications, respectively. Most of these publications were published in the latter half (from 2004 onwards) of the programme's operation.
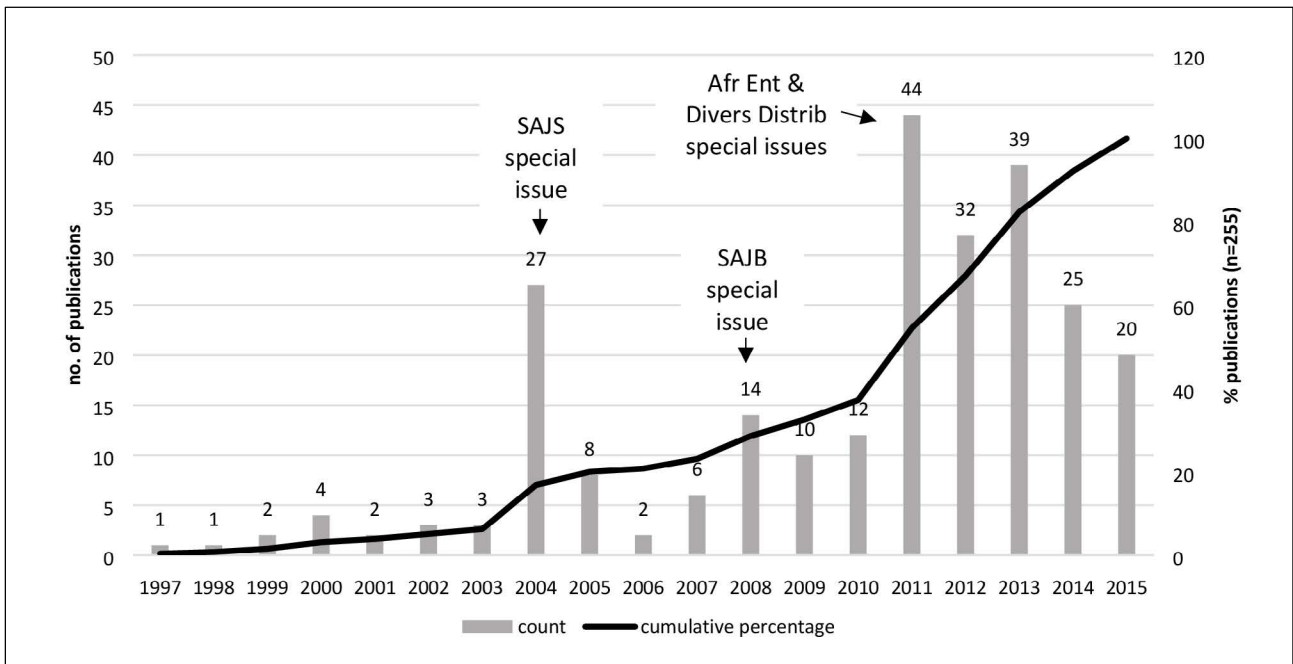
SAJS, South African Journal of Science*; SAJB,* South African Journal of Botany*; Afr Ent,* African Entomology*; Divers Distrib,* Diversity and Distributions

Figure 1: Working for Water related research output over time. The bars show the number of papers produced over the course of the programme and the black line shows the cumulative percentage of publications.
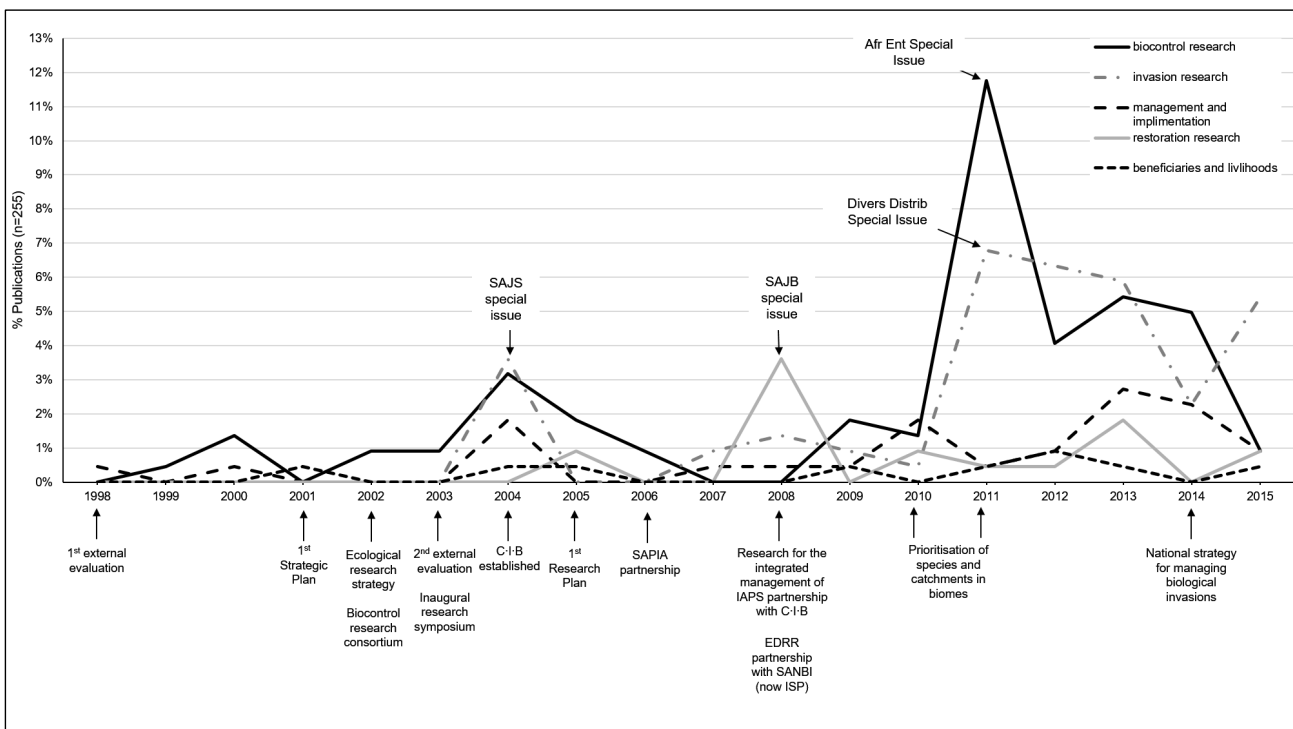
C•I•B, Centre for Invasion Biology*; SAPIA,* Southern African Plant Invaders Atlas*; EDRR,* Early Detection and Rapid Response*; ISP,* Invasive Species Programme*; SANBI,* South African National Biodiversity Institute*; SAJS,* South African Journal of Science*; SAJB,* South African Journal of Botany

Figure 2: Timeline of events in relation to research output associated with Working for Water research themes.

Inductive content analysis

Through the inductive content analysis of peer-reviewed publications (*n*=255), we identified eight research categories of which biological control (34.5%) and invasion research (29.4%) were found to be the primary focus of research produced under the auspices of the WfW programme (Table 1 and Figure 2). The largest proportion (68.2%) of biological control research was published from 2011 to 2015, peaking in 2011 with the publication of an *African Entomology* special issue reviewing biological control efforts against IAPs in South Africa. Biological control research focused on the testing of biological control agents for high priority IAPs that are already established and have caused extensive damage to ecosystems and included studies on host range and specificity, risk assessments relating to the suitability of biocontrol agent release, and ecological impacts of biocontrol agents on target and non-target species.

Invasion research covered a range of topics, such as species introduction (pathways and risk assessment), IAP establishment, expansion and spread (determinants of success) and their impacts (biodiversity patterns and process, ecosystem functioning). Of this research category, 78.7% was published from 2011 to 2015 (Figure 2). Restoration research is also a significant component of the ecological research cluster and is most notably illustrated by the *South African Journal of Botany* Volume 74 special issue of 2008 on riparian restoration and management (Figure 2). The *South African Journal of Botany* special issue covers a series of topics relating to post IAP clearing rehabilitation, and active and passive restoration, with a predominant focus on riparian ecosystems. A few publications (3.9%; Figure 2) had hydrological research as their primary focus and for those that did, IAP water use, water yields and catchment experiments were the topics discussed.

Research primarily focused on management, economic and social aspects of the WfW programme constituted less than 30% of the total publications (Table 1). Management and implementation research (11.4%) consisted largely of reviews of management efforts, and the challenges, limitations and trade-offs associated with IAP management and WfW in particular. A relatively small proportion of this research focused on operational aspects of management, such as communication, planning, monitoring and evaluation (Table 2). Economic research (7.8%) covered economic aspects of the programme including feasibility studies, cost–benefit analyses and, to a lesser extent, valuations and pricing estimation (Table 1). Social research dealing with the employment of beneficiaries

and community livelihoods constituted a relatively small proportion of total research produced under the auspices of the programme (3.5%) (Table 1).

Deductive content analysis

Categorising publications according to the research areas and priorities obtained from the WfW RSAP (Table 2) indicates that the research conducted was largely biophysical in nature and ecologically focused. The WfW-related social, economic and hydrological research was largely intermittent and infrequent, despite the research priorities for these themes having been set by WfW (Table 2).

The correspondence analysis based on data coded using the deductive approach and presented in Tables 2 and 3 shows that research objectives are clustered around themes that are simultaneously addressed (Figure 3). Figure 3a, the plot of dimensions 1 and 2, reveals three clusters of research: (1) ecology, economics, management and social, (2) hydrology and (3) biological control. Although it is suggested from Figure 3a that ecological, management, economic and social research objectives cluster (i.e. are often addressed simultaneously in publications), this grouping is teased apart in Figure 3b (which plots dimensions 2 and 3), showing that ecological and management research objectives cluster together and are associated with the contribution of the DST-NRF Centre of Excellence for Invasion Biology (C•I•B) at Stellenbosch University, through its collaboration with WfW, 'Integrated management of invasive alien species'. This cluster is distinct from social and economic research clusters that have clearer overlap in their research objectives. This observation suggests that there is a limited degree of correspondence between economic, social and ecologically focused research.

Biological control and hydrological research also formed unique clusters (Figure 3a), with biological control forming a distinct cluster with minimal cross categorisation with research themes (Table 3). Biological control research is characterised by a strong focus on the development and testing of biocontrol agents and funding collaborations with the South African Agricultural Research Council, whereas hydrological research formed a relatively distinct but loosely clustered grouping associated with the South African Water Research Commission co-funding. The analysis also shows that biological control, despite featuring prominently in research produced under the aegis of WfW, not only shows a poor degree of correspondence to other reported topics such as socio-economic research and hydrology, but also to other ecologically focused research (Figure 3).

Table 1: Description of the categories of research conducted under the auspices of the Working for Water programme, including the percentage of publications in each category (based on inductive content analysis)

Research category	Description of research category	% Publications (*n*=255)
Biocontrol research	Deals with biocontrol, management pros and cons and suitability for release	34.51%
Invasion research	Focuses on the definitions, concepts, mechanisms, new introductions, distribution, abundance, demography and synergistic effects etc. associated with invasive alien plants (IAPs) (i.e. research relating to invasion dynamics)	29.41%
Management and implementation	Discusses the management of IAPs and the outcomes of management activities	11.37%
Restoration research	Discusses restoration and rehabilitation after invasion	7.84%
Economic research	Deals with economic aspects of the programme, including feasibility studies, cost–benefit analyses, valuations and pricing estimates	7.84%
Hydrological impacts	Discusses the hydrological impacts associated with invasions (i.e. IAP water use, surface water yield and river flow response)	3.92%
Beneficiaries and livelihoods	Discusses the human dimensions associated with IAP management, including job creation, poverty relief and livelihoods	3.53%
Ecology	Deals with ecological studies not directly related to IAPs	1.57%
Total		**100%**

Table 2: Working for Water programme related publications categorised by research areas and priorities obtained from the 2005 research strategy[26] (using deductive content analysis). Number and percentage of total publications assigned to research areas and research priorities are reported.

Working for Water research area	Publications[a]		% Publications (n=255)
	N	%	
Ecological research	124	36.7	48.6
Economic research	28	8.3	11.0
Hydrological research	12	3.6	4.7
Management research	68	20.1	26.7
Social research	15	4.4	5.9
Biological control research	91	26.9	35.7
Total	338	100.0	132.5
Research priorities			
Ecological research			
Vectors and pathways of invasion (invasion dynamics)	47	12.6	18.4
Prevention and tools (mapping, risk assessment, prediction models etc.)	28	7.5	11.0
Control options (management recommendations)	17	4.6	6.7
Post-clearing rehabilitation (riparian restoration etc.)	20	5.4	7.8
Ecological impacts	21	5.6	8.2
Other[b]	18	4.8	7.1
Hydrological research			
Water use by invasive alien plants	4	1.1	1.6
GIS-based prediction modelling (and other predictive modelling)	2	0.5	0.8
Catchment experiments	2	0.5	0.8
Other[b]	28	7.5	11.0
Economic research	4	1.1	1.6
Management research			
Communications and extension	4	1.1	1.6
Planning	1	0.3	0.4
Financial management	–	–	–
Field operations	–	–	–
Beneficiation (secondary industries)	5	1.3	2.0
Education	–	–	–
Organisational structure	–	–	–
Data management	–	–	–
Audit and monitoring	–	–	–
Human resources	–	–	–
Legislation	1	0.3	0.4
Other[b]	58	15.6	22.7
Social research			
HIV/Aids impact on programme	–	–	–
Employment, training of beneficiaries (and poverty alleviation)	3	0.8	1.2
Exit strategy and job opportunities	1	0.3	0.4
Occupational health and safety	1	0.3	0.4
Other[b]	10	2.7	3.9
Biological control research			
Development of biocontrol agents	78	21.0	30.6
Pre-emptive control	8	2.2	3.1
Other[b]	11	3.0	4.3
	372	100.0%	145.9%

[a]Publications may be assigned to multiple research areas and priorities and therefore may be counted more than once. Percentages are calculated both as a proportion of the total number of counts (N) and total publications (n).

[b]Other represents the number of publications that cover the research areas more generally, but do not fit into any of the research priority categories.

Table 3: The number of papers categorised according to research priorities under each research area, presented in Table 2. Supplementary variables represent funding organisations administering Working for Water funding: Centre for Invasion Biology (C•I•B), the Agricultural Research Council (ARC) and the Water Research Commission (WRC). Research funded through the 'Integrated management of invasive alien species' (Int mgt) collaboration between C•I•B and WfW is also presented. Research priorities to which no publications are assigned are excluded from this table.

Research priorities	Code[a]	Research areas						Supplementary variables			
		Ecology	Economics	Hydrology	Management	Social	Biocontrol	Int mgt	ARC	C•I•B	WRC
Ecology											
Vectors and pathways of invasion	**e1**	47	–	–	9	–	1	24	2	37	1
Prevention and tools	**e2**	28	–	1	14	–	1	12	–	19	1
Control options	**e3**	17	–	–	14	–	–	4	–	9	1
Restoration and rehabilitation	**e4**	20	–	–	13	–	–	6	–	9	2
Ecological impacts	**e5**	21	–	–	3	–	–	9	–	15	–
Other	**e**	18	7	3	7	5	2	3	–	10	–
Economics											
	ec	7	28	1	12	6	1	2	–	5	3
Hydrology											
Water use by invasive alien plants	**h1**	–	–	2	–	–	–	–	–	–	1
GIS-based prediction modelling	**h2**	–	–	2	1	–	–	–	–	–	1
Catchment experiments	**h3**	1	1	4	-	–	–	–	–	–	–
Other	**h**	3	–	4	1	–	–	1	–	1	1
Management											
Communications and extension	**m1**	–	–	–	4	–	–	–	–	2	–
Planning	**m2**	1	–	–	1	–	–	1	–	1	–
Beneficiations	**m5**	1	3	–	5	2	–	1	–	1	–
Legislation	**m11**	–	–	–	1	–	–	1	–	1	–
Other	**m**	45	10	2	57	9	3	13	2	27	3
Social											
Employment, training of beneficiaries	**s2**	–	–	–	1	3	–	–	–	2	–
Exit strategy and job opportunities	**s3**	–	–	–	1	1	–	–	–	–	–
Occupational health and safety	**s4**	–	–	–	–	1	–	–	–	–	–
Other	**s**	4	5	–	9	10	–	1	–	3	–
Biological control											
Development of biocontrol agents	**b1**	1	–	–	–	–	78	–	20	–	2
Pre-emptive control	**b2**	–	–	–	–	–	8	–	4	–	–
Other	**b**	3	1	–	3	–	11	–	2	1	–

[a]Items in the column labelled 'Code' serve as a category identifier in Figure 3.
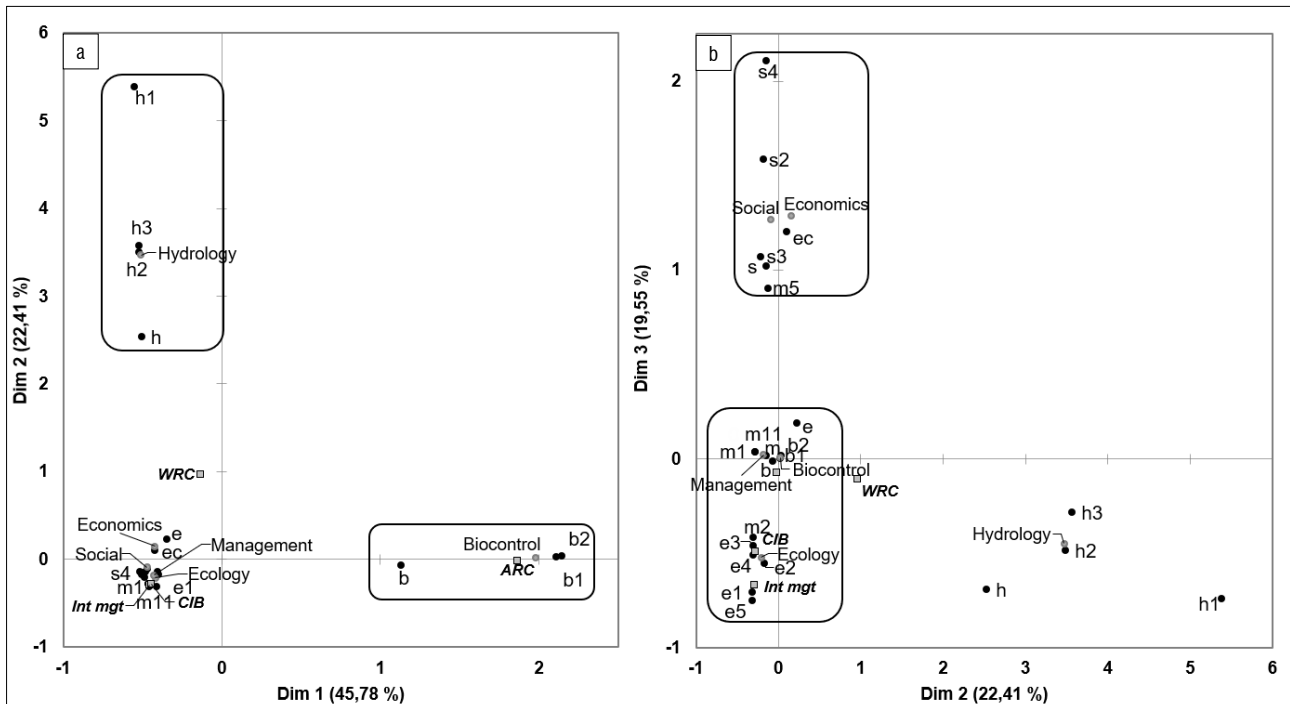
Figure 3: Correspondence analysis showing the relationship between research areas and priorities based on categories listed in Table 3. The two-dimensional solution (3a) explains 68.19% of the correspondence between research areas and priorities. The second solution (3b), Dim 2 and Dim 3, explains 41.95% of the correspondence.

Discussion

The integration of diverse disciplines and knowledge types can occur at multiple levels in the research process and may look different depending on the level of participation by stakeholders, on where in the research process participation occurs, or on how funding is allocated.[38] We found that research published under the aegis of WfW is suggestive of integration of multiple disciplines into the programme, but that it is biased towards the biophysical and natural sciences (including invasion biology and biological control), as opposed to the social and economic sciences. This bias is despite the acknowledgement by WfW in its RSAP, and by the national strategy for dealing with biological invasions (and related policy documents), that ecological sciences on their own are insufficient in addressing challenges associated with implementing WfW activities and achieving their mandate.[26,27] This finding is reflective of the broader literature – that is, that SES and associated challenges are defined and framed primarily in natural science terms and under-representative of the social sciences, despite SES challenges often being driven by social needs.[3,10,13,39] For example, Vaz et al.[40] show, in their review of global invasion science literature since 1950, that despite interdisciplinarity becoming increasingly prominent, collaborations between disciplines remain largely confined within sub-disciplines of ecology and the environmental sciences. Although contributions by the social sciences and humanities have increased, collaborations between social scientists and ecologists are minimal, and resulting integrative social-ecological studies therefore remain under-represented.[40]

The need for biophysical and natural science research

The over-representation of ecological research may be a concern for research funders in that more social and economic studies are needed to help inform and guide decision-making.[41] However, in the context of invasion biology and the management of invasions in the country, ecological research is still highly important and necessary.[42] Strong cases have been made for investing in ecologically focused research and research capacity building, towards addressing challenges associated with managing IAPs. This strength is evident from the numerous collaborations and partnerships that have been established since WfW's inception, and that have resulted in many research outputs, including journal special issues.[25,43] The first special issue was a *South African*

Journal of Science issue of 2004, funded by WfW with the primary purpose of expanding the basis of knowledge needed to make sound management decisions.[23] The special issue of 2008 (*South African Journal of Botany*) was a culmination of work from a project commissioned by WfW on targets for ecosystem repair in alien-invaded riparian zones, and also included some additional papers contributing to the theme. The primary aim of this ecosystem repair project was to develop guidelines and tools to improve management of alien invaded riparian systems.[43] The 2011 special issue of *Diversity and Distributions* on *Acacia* invasions has further contributed to the knowledge base needed to deal with biological invasions effectively, with contributions by several WfW co-funded authors.[44] Albeit limited, several papers in these special issues have demonstrated varying degrees of multi- or interdisciplinarity. Most papers were, however, largely monodisciplinary and ecologically focused.

While an over-representation of ecologically focused research is not inherently problematic, a lack of integration with economic or socially focused knowledge may be[10], particularly when SES challenges are to be addressed.[13] The prominence of ecologically focused research can be attributed to a handful of factors, amongst them WfW's prioritisation of ecologically focused research and their emphasis on building ecological research (including biological control) capacity[26,27], and the interests of the researchers who lead research initiatives under its auspices.[45] Furthermore, our findings show that biological control, despite featuring prominently amongst research produced under the aegis of WfW, shows a poor degree of correspondence to other reported topics recorded in this study (Figure 3). Investment in biological control research by WfW has long been part of their strategy for the control of plant invasions, and acknowledged by researchers and WfW management as a necessary component of effective integrative IAP control, particularly where mechanical and chemical control methods are insufficient.[27,46-48] If biological control is not more effectively integrated with other disciplines, the desired impact of such research on management operations will be limited.[10]

The need for better integration between social and natural science research

Integrated research that engages the social sciences is necessary to solve complex social-ecological challenges, but it does not occur automatically,

even when public funding encourages it. Integrative research involves more than simply aggregating several disciplines into a single research project.[38,41] Integration requires effective coordination and interaction between relevant stakeholders and enabling environments that encourage the incorporation of multiple knowledge systems and more pluralistic approaches towards achieving specific SES initiative mandates. Encouragingly, there have been efforts towards improving the integration (albeit limited with respect to socially focused research) of numerous disciplines, knowledge systems and stakeholders in the planning and research processes, and towards informing operations.[26,27,46-49]

The initial framing of the WfW programme brought together biologists, hydrologists and resource economists who built the case for the management of biological invasions in the broader context of ecosystem services and water security.[5] Further development of the RSAP, which makes a specific allowance for the establishment of a research advisory panel comprising experts in the field of IAP management[26], has contributed to improved linkages across disciplines, as well as between research and management.[5,26,42] However, despite social considerations being increasingly integrated into the planning and framing of strategies, socially focused research does not appear to be prioritised.[10,27,46,47,49-52] Our findings further suggest that WfW has not been comprehensive in addressing their socially focused research priorities (Table 2). The under-representation of the social sciences and operationally focused research is of concern (see Table 2). This under-representation could be attributed to a lack of explicit expertise devoted to social research, policy formation and implementation within the organisations leading the research. This gap was acknowledged by the WfW programme in their RSAP and subsequent policy documents, wherein an argument is made for engaging with institutions that conduct specialist social research, more particularly the Human Sciences Research Council of South Africa (the social research equivalent to the Council for Scientific and Industrial Research (CSIR) in South Africa) as well as other university faculties such as humanities and social sciences faculties, whose contribution to research is not clearly evident in this study. A concerted effort is necessary to address this imbalance. This effort would require building and strengthening of social and social-ecological focused research capacity, through training and the development of a research strategy that complements the current ecological focus. However, these initiatives would have implications for research spend.[41]

The WfW programme had an initial budget of ZAR25 million (in 1995) which rose sharply over the following 3 years, with the annual budget reaching ZAR250 million in 1997. The budget then stabilised between ZAR300 million and ZAR400 million until 2010. Thereafter, it increased to its current level of ZAR1979 million in the 2016/2017 financial year. The number of people who benefitted through employment from the programme changed over time, and stands at approximately 56 425 people for the 2016/2017 financial year, and 633 234 since the programme's inception.[27,52] Even though research funding has increased significantly from approximately ZAR45 million over 2001–2003 to a budget of approximately ZAR135 million in 2016/2017, research spend averaged approximately 4.8% of the total budget since 2000/2001 (Wannenburgh A 2018; written communication, August 6). Socially focused research spend, in particular, remains low.[25,52] Where integrative approaches to research are adopted, the role of funders like WfW becomes increasingly important.[41] This role includes framing calls for integrative research proposals and developing monitoring, learning and evaluation processes (for interdisciplinary proposals and funded projects).[41]

Challenges to the achievement of effective SES research

The ability of SES initiatives to achieve their mandates is significantly affected by capacity, skill and experience required for the effective implementation of operations, including the co-design (or co-generation) of research activities. Effective SES management programmes tend to engage and collaborate with organisations already undertaking integrative research in the sector, to guide or oversee research programmes and serve as intermediaries between stakeholders.[13] To this end, WfW has made considerable progress through its coordinated research and capacity-building partnerships with the C•I•B, South African National Biodiversity Institute and the CSIR.[45] There is, however, a need for a

mutually acceptable research strategy to be developed that leverages the potential of WfW's existing research partners, managers and relevant stakeholders' elsewhere.[45]

To be effective, SES research should be developed jointly by managers, researchers and practitioners from a wide range of backgrounds and implementation contexts. While we do not explicitly demonstrate if such engagement has occurred, several studies and policy documents suggest that managers and researchers working on issues relating to invasion do not engage extensively in developing SES research.[12,27,53,54] The 2005 WfW research strategy, while setting no clear goals, notes (among other things) the need to (1) ensure that the research capacity in WfW collaborates closely with the Monitoring & Evaluation Unit in the development of WfW's M&E programme; (2) ensure there is optimal two-way communication between research and management; and (3) set in place a procedure to review and update this strategy as and when necessary. However, it is unclear (from policy and other strategic documents) if these needs have been sufficiently addressed.[27] The RSAP, now 13 years old, has not been updated or revised. While several policy and strategy documents have listed research priorities and objectives, the RSAP remains the most explicit, publicly available research strategy and action plan employed by WfW.[27,48] A significant shift in the research priorities set by the WfW RSAP is not apparent when it is compared to later policy and other strategy documents. However, there have been calls to include the application of transdisciplinary research methods needed to ensure that socio-economic aspects of the programme are integrated in the problem and solution framing relating to invasive species.[27]

The potential benefits of achieving better integration

Integrative approaches to research such as those associated with transdisciplinarity have the potential to improve linkages between science, policy and practice through improved stakeholder engagement and grounding of research in social processes, making research more relevant and practicable for knowledge users.[10,55] However, these approaches do not guarantee that the scientific knowledge will be automatically integrated into policy or practice unless appropriate governance mechanisms are in place.[56-58] A lack of involvement of practitioner stakeholders during the formulation of research programmes, the lack of relevant and accessible information in an appropriate form and the potentially artificial distinctions between science and society, are just some of the barriers to producing an adequate evidence (knowledge) base for informing management action.[12,59] Co-production approaches strongly support integrated learning between researchers and practitioners and emphasise the fundamental role of communication, translation and mediation processes between researchers and practitioners. The supportive role of intermediary organisations in creating and enhancing potentially mutually beneficial activities facilitated by the approach, are key to its adoption and application.[12,60-63] Several publications analysed in this study demonstrate the need for increased engagement with WfW managers and conservation practitioners in the research. These studies are those aimed at initiating dialogue between researchers and managers[53]; understanding how scientific evidence informs practice[12,54]; identifying the drivers and challenges facing WfW[64]; and comparing stakeholder perceptions on the ecosystem services approach to IAP management.[65]

In the case of WfW and other large SES programmes, translating investment into action requires the effective coordination of multiple partner institutions, their mandates and resources, particularly where expectations and operational standards may differ.[66] Multi-organisational partnerships offer important means of governing and managing public or SES programmes. These partnerships are, however, subject to key challenges relating to the management of interactions between organisations, different modes of governance and benefit sharing. The benefits of these partnerships can, however, open the decision-making processes, and improve use and benefit sharing of scarce resources (e.g. finances, research capacity and skills).[66,67] The productivity (in terms of research output observed in this study) of the collaborative efforts between WfW and C•I•B demonstrates the value of research partnerships in building a research capacity, resource and benefit sharing.[42] The WfW–C•I•B collaboration has significantly impacted the

framing of IAP management related research. This work has continued to build on the successes of partnerships with the likes of CSIR, which has played a leading role in understanding the management of biological invasions, and the translation of that knowledge into policy. Both the CSIR and C•I•B – through their research capacity, expertise and long-standing relationships with WfW – have become embedded in the programme. Maintaining these collaborations and extending agreements to include (more proactively) more socially focused research drivers such as the Human Sciences Research Council (that already have formal mandates to conduct research in areas which overlap with WfW's priority research fields), will go a long way in improving the reach and relevance of research to a wider audience.[26]

Conclusion

Producing relevant and strategic research that meets the social-economic demands of society has become a recurrent theme in environmental policy documents both in South Africa and globally.[15,30,60,68] We show that while WfW has made significant progress in addressing its ecological research priorities, it has not been as comprehensive in addressing its socially and economically focused research priorities. There is a strong need to rectify the disciplinary imbalance of its research and draw upon diverse knowledge systems outside of academia. This rectification is key to improving the decision-making processes guiding IAP management and the social processes that govern them. Furthermore, comprehensive planning and adaptive management are essential to the effective implementation of integrative research programmes that convey responsibilities, timelines and relationships between project components. This coordination can ensure that the impacts of programme activities are meaningful, long lasting, and more effectively monitored, reflected upon, evaluated and updated.[10,60,63,64,68]

Appropriate governance structures are needed to support these integrative research programmes. Improving linkages between practitioners and researchers is essential, particularly as they relate to the framing and development of research and management priorities.[69] A detailed research strategy and action plan conducive to integrative research and transdisciplinary collaboration, and relevant to solving complex SES challenges such as those associated with biological invasions, needs to be developed. This strategy will, however, have significant implications for funding[41], including allowances for investment in liaison roles and less visible processes (such as warm-up activities, start-up support, team-building exercises, and network- and community-building) in funding agreements, as well as clear mechanisms for capturing organisational learning if integrative SES research investments are to achieve their desired outcomes.[70]

Acknowledgements

Authors' contributions

B.A.: Conceptualisation, methodology, data collection, data analysis, data curation, writing the initial draft. N.S.: Conceptualisation, writing revisions, student supervision. K.J.E.: Conceptualisation, writing revisions, student supervision, funding acquisition.

References

1. Reid W, Chen D, Goldfarb L, Hackman H, Lee Y, Mokhele K, et al. Earth system science for global sustainability: Grand challenges. Science. 2010;330:916–917. https://doi.org/10.1126/science.1196263

2. Cooke B, West S, Boonstra WJ. Dwelling in the biosphere: Exploring an embodied human–environment connection in resilience thinking. Sustain Sci. 2016;11(5):831–843. https://doi.org/10.1007/s11625-016-0367-3

3. Jerneck A, Olsson L, Ness B, Anderberg S, Baier M, Clark E, et al. Structuring sustainability science. Sustain Sci. 2011;6:69–82. https://doi.org/10.1007/s11625-010-0117-x

4. Reid W, Mooney H, Cropper A, Capistrano D, Carpenter S, Chopra K, et al. The Millennium Ecosystem Assessment: Ecosystems and human well-being: Synthesis. Washington DC: World Resources Institute; 2005.

5. Van Wilgen BW, Wannenburgh AM. Co-facilitating invasive species control, water conservation and poverty relief: Achievements and challenges in South Africa's Working for Water program. Curr Opin Env Sust. 2016;19:7–17. https://doi.org/10.1016/j.cosust.2015.08.012

6. Chapin FI, Knapp C, Brinkman T, Bronen R, Cochran P. Community-empowered adaption for self-reliance. Curr Opin Env Sust. 2016;19:67–75. https://doi.org/10.1016/j.cosust.2015.12.008

7. Esler K, Downsborough L, Roux D, Blignaut J, Milton S, Le Maitre D, et al. Interdisciplinary and multi-institutional higher learning: Reflecting on a South African case study investigating complex and dynamic environmental challenges. Curr Opin Env Sust. 2016;19:76–86. https://doi.org/10.1016/j.cosust.2015.12.002

8. Bridgewater P. The Man and the Biosphere program of UNESCO: Rambunctious child of the sixties, but was the promise fulfilled? Curr Opin Env Sust. 2016;19:1–6. https://doi.org/10.1016/j.cosust.2015.08.009

9. Reid W, Mooney H. The Millennium Ecosystem Assessment: Testing the limits of interdisciplinary and multi-scale science. Curr Opin Env Sust. 2016;19:40–46. https://doi.org/10.1016/j.cosust.2015.11.009

10. Cowling RM, Egoh B, Knight AT, O'Farrell PJ, Reyers B, Rouget M, et al. An operational model for mainstreaming ecosystem services for implementation. Proc Natl Acad Sci USA. 2008;105(28):9483–9488. https://doi.org/10.1073/pnas.0706559105

11. Lang DJ, Wiek A, Bergmann M, Stauffacher M, Martens P, Moll P, et al. Transdisciplinary research in sustainability science: Practice, principles, and challenges. Sustain Sci. 2012;7(suppl 1):25–43. https://doi.org/10.1007/s11625-011-0149-x

12. Ntshotsho P, Prozesky HE, Esler KJ, Reyers B. What drives the use of scientific evidence in decision making? The case of the South African Working for Water program. Biol Conserv. 2015;184:136–144. https://doi.org/10.1016/j.biocon.2015.01.021

13. Turner BL, Esler KJ, Bridgewater P, Tewksbury J, Sitas N, Abrahams B, et al. Socio-environmental systems (SES) research: What have we learned and how can we use this information in future research programs. Curr Opin Env Sust. 2016;19:160–168. https://doi.org/10.1016/j.cosust.2016.04.001

14. Cornell S, Berkhout F, Tuinstra W, Tabara JD, Jager J, Chabay I, et al. Opening up knowledge systems for better responses to global environmental change. Environ Sci Policy. 2013;28:60–70. https://doi.org/10.1016/j.envsci.2012.11.008

15. Tengo M, Brondizio ES, Elmqvist T, Malmer P, Spierenburg M. Connecting diverse knowledge systems for enhanced ecosystem governance: The multiple evidence base approach. Ambio. 2014;43:579–591. https://doi.org/10.1007/s13280-014-0501-3

16. Buscher B. Anti-politics as political strategy: Neoliberalism and transfrontier conservation in southern Africa. Dev Change. 2010;41:29–51. https://doi.org/10.1111/j.1467-7660.2009.01621.x

17. Elgert L. Politicizing sustainable development: The co-production of globalized evidence-based policy. Crit Policy Stud. 2010;3:375–390. https://doi.org/10.1080/19460171003619782

18. Clark WC, Tomich TP, Van Noordwijk M, Guston D, Catacutan D, Dickson NM, et al. Boundary work for sustainable development: Natural resource management at the Consultative Group on International Agricultural Research (CGIAR). Proc Natl Acad Sci USA. 2016;113:4615–4622. https://doi.org/10.1073/pnas.0900231108

19. Common Ground. Working for Water external evaluation synthesis report. Cape Town: Common Ground (Pty) Ltd; 2003.

20. Hobbs RJ. The Working for Water program in South Africa: The science behind the success. Divers Distrib. 2004;10(5–6):501–503. https://doi.org/10.1111/j.1366-9516.2004.00115.x

21. Jelinek J, Breen C. Working for Water program evaluation report. Florstadt: Ingenieurbüro für Landentwicklung; 1997.

22. Le Maitre DC, Richardson DM, Chapman RA. Alien plant invasions in South Africa: Driving forces and the human dimension. S Afr J Sci. 2004;100:103–112.

23. Macdonald I. Recent research on alien plant invasions and their management in South Africa: A review of the inaugural research symposium of the Working for Water programme. S Afr J Sci. 2004;100:21–26.

24. Richardson DM, Moran VC, Le Maitre DC, Rouget M, Foxcroft LC. Recent developments in the science and management of invasive alien plants: Connecting the dots of research knowledge, and linking disciplinary boxes. S Afr J Sci. 2004;101:126–128.

25. Van Wilgen BW. Scientific challenges in the field of invasive alien plant management. S Afr J Sci. 2004;100:19–20.

26. Working for Water. A new strategy and action plan for research within the Working for Water program [document on the Internet]. c2005 [cited 2018 Aug 23]. Available from: http://www.dwaf.gov.za/wfw/docs/WfWResearchStrat.pdf

27. Department of Environmental Affairs. A national strategy for dealing with biological invasions in South Africa. Pretoria: Department of Environmental Affairs; 2014.

28. Norris M, Oppenheim C. Comparing alternatives to the Web of Science for coverage of the social sciences' literature. J Informetr. 2007;1:161–169. https://doi.org/10.1016/j.joi.2006.12.001

29. Kahn M. A bibliometric analysis of South Africa's scientific outputs – some trends and implications. S Afr J Sci. 2011;107(1–2), Art. #406, 6 pages. https://doi.org/10.4102/sajs.v107i1/2.406

30. Chavarro D, Tang P, Rafols I. Interdisciplinarity and research on local issues: Evidence from a developing country. Res Eval. 2014;23:195–209. https://doi.org/10.1093/reseval/rvu012

31. Baneyx A. "Publish or perish" as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history. Arch Immunol Ther Ex. 2008;56:363–371. https://doi.org/10.1007/s00005-008-0043-0

32. Koier E, Horlings E. How accurately does output reflect the nature and design of transdisciplinary research programs? Res Eval. 2015;24:37–50. https://doi.org/10.1093/reseval/rvu027

33. Elo S, Kyngas H. The qualitative content analysis process. J Adv Nurs. 2008;62(1):107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

34. Johnson R, Onwuegbuzie A, Turner L. Toward a definition of mixed methods research. J Mix Methods Res. 2010;1:112–133. https://doi.org/10.1177/1558689806298224

35. Dey I. Qualitative data analysis: A user-friendly guide for social scientists. London: Routledge; 1993. https://doi.org/10.4324/9780203412497

36. Saldana J. The coding manual for qualitative researchers. London: Sage; 2009.

37. Clausen SE. Applied correspondence analysis: An introduction. Volume 121. Thousand Oaks, CA: Sage; 1998. https://doi.org/10.4135/9781412983426

38. Tress B, Tress G, Fry G. Defining concepts and the process of knowledge production in integrative research. In: Tress B, Tress G, Fry G, editors. Landscape research to landscape planning: Aspects of integration, education and application. Heidelberg: Springer; 2005. p. 13–26.

39. Buscher B, Wolmer W. Introduction: The politics of engagement between biodiversity conservation and the social sciences. Conserv Soc. 2007;5:1–21.

40. Vaz AS, Kueffer C, Kull CA, Richardson DM, Schindler S, Muñoz-Pajares AJ, et al. The progress of interdisciplinarity in invasion science. Ambio. 2017;46(4):428–442. https://doi.org/10.1007/s13280-017-0897-7

41. Lyall C, Fletcher I. Experiments in interdisciplinary capacity-building: The successes and challenges of large-scale interdisciplinary investments. Sci Public Policy. 2013;40:1–7. https://doi.org/10.1093/scipol/scs113

42. Van Wilgen BW, Davies SJ, Richardson DM. Invasion science for society: A decade of contributions from the Centre for Invasion Biology. S Afr J Sci. 2014;110(7/8), Art. #a0074, 12 pages. https://doi.org/10.1590/sajs.2014/a0074

43. Esler KJ, Holmes PM, Richardson DM, Witkowski ETF. Riparian vegetation management in landscapes invaded by alien plants: Insights from South Africa. S Afr J Bot. 2008;74(3):397–400. https://doi.org/10.1016/j.sajb.2008.01.168

44. Richardson DM, Carruthers J, Hui C, Impson FAC, Miller JT, Robertson MP, et al. Human-mediated introductions of Australian acacias – a global experiment in biogeography. Divers Distrib. 2011;17:771–787. https://doi.org/10.1111/j.1472-4642.2011.00824.x

45. Abrahams B, Sitas N, Esler KJ. Exploring the dynamics of research collaborations by mapping social networks in invasion science. J Environ Manage. 2019;229:27–37. https://doi.org/10.1016/j.jenvman.2018.06.051

46. Working for Water. The Working for Water 2004/5 to 2007/8 strategic plan. Draft 11. 2004. Unpublished report.

47. Working for Water (WfW). The Working for Water strategic plan 2008-2012. Pretoria: WfW; 2007.

48. Stafford W, Le Maitre DC, Forsyth G. S.M.A.R.T. Goals for the Natural Resource Management programmes. CSIR report CSIR/NRE/GES/ER/2016/0002/B. Final version: v2.5. 2016.

49. Working for Water. Working for Water historical figures [webpage on the Internet]. c2017 [cited 2019 Feb 11]. Available from: https://sites.google.com/site/wfwplanning/Home/WfW%20historical%20figures.xls?attredirects=0&d=1

50. Magadlela D, Mdzeke, N. Social benefits in the Working for Water programme as a public works initiative. S Afr J Sci. 2004;100:94–96.

51. Hough A, Prozesky HE. Beneficiaries' aspirations to permanent employment within the South African working for Water Programme. Soc Dyn. 2012;38(2):331–349. https://doi.org/10.1080/02533952.2012.719395

52. Natural Resource Management. 2016-17 NRM research summary. Pretoria: Department of Environmental Affairs; 2017. Available from: https://sites.google.com/site/nrmprogrammes/home/2016_17%20NRM%20research%20summary.xlsx?attredirects=0&d=1

53. Shaw JD, Wilson JRU, Richardson DM. Initiating dialogue between scientists and managers of biological invasions. Biol Invasions. 2010;12(12):4077–4083. https://doi.org/10.1007/s10530-010-9821-9

54. McConnachie M, Cowling RM. On the accuracy of conservation managers' beliefs and if they learn from evidence-based knowledge: A preliminary investigation. J Environ Manage. 2013;128:7–14. https://doi.org/10.1016/j.jenvman.2013.04.021

55. Newing H. Interdisciplinary training in environmental conservation: Definitions, progress and future directions. Environ Conserv. 2010;37(4):410–418. https://doi.org/10.1017/S0376892910000743

56. Ostrom E. A general framework for analyzing sustainability of social-ecological systems. Science. 2009;325(5939):419–422. https://doi.org/10.1126/science.1172133

57. Ebbesson J. The rule of law in governance of complex socio-ecological changes. Glob Env Change. 2010;20(3):414–422. https://doi.org/10.1016/j.gloenvcha.2009.10.009

58. Leemans R. The lessons learned from shifting from global-change research programmes to transdisciplinary sustainability science. Curr Opin Env Sust. 2016;19:103–110. https://doi.org/10.1016/j.cosust.2016.01.001

59. Mollinga PP. Boundary work and the complexity of natural resources management. Crop Sci. 2010;50:1–9. https://doi.org/10.2135/cropsci2009.10.0570

60. Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, et al. Knowledge systems for sustainable development. Proc Natl Acad Sci USA. 2003;100:8086–8091. https://doi.org/10.1073/pnas.1231332100

61. Reyers B, Nel JL, O'Farrell PJ, Sitas N, Nel DC. Navigating complexity through knowledge coproduction: Mainstreaming ecosystem services into disaster risk reduction. Proc Natl Acad Sci USA. 2015;112(24):7362–7368. https://doi.org/10.1073/pnas.1414374112

62. Esler KJ, Prozesky H, Sharma GP, McGeoch M. How wide is the "knowing-doing" gap in invasion biology? Biol Invasions. 2010;12:4065–4075. https://doi.org/10.1007/s10530-010-9812-x

63. Nel JL, Roux DJ, Driver A, Hill L, Maherry AC, Snaddon K, et al. Knowledge co-production and boundary work to promote implementation of conservation plans. Conserv Biol. 2016;30(1):176–188. https://doi.org/10.1111/cobi.12560

64. Roura-Pascual NN, Richardson DM, Krug RM, Brown A, Chapman RA, Forsyth GG, et al. Ecology and management of alien plant invasions in South African fynbos: Accommodating key complexities in objective decision making. Biol Conserv. 2009;142(8):1595–1604. https://doi.org/10.1016/j.biocon.2009.02.029

65. Urgenson LS, Prozesky HE, Esler KJ. Stakeholder perceptions of an ecosystem services approach to clearing invasive alien plants on private land. Ecol Soc. 2013;18(1), Art. #26, 13 pages. https://doi.org/10.5751/ES-05259-180126

66. Lowndes V, Skelcher C. The dynamics of multi-organizational partnerships: An analysis of changing modes of governance. Public Adm. 1998;76:313–333. https://doi.org/10.1111/1467-9299.00103

67. Folke C, Hahn T, Olsson P, Norberg J. Adaptive governance of social-ecological systems. Annu Rev Environ Resour. 2005;30:441–473. https://doi.org/10.1146/annurev.energy.30.050504.144511

68. Babbie E, Mouton J. The practice of social research (South Africa edition). Cape Town: Oxford University Press; 2001.

69. Moon K, Blackman D, Brewer TD, Sarre SD. Environmental governance for urgent and uncertain problems. Biol Invasions. 2017;19(3):785–797. https://doi.org/10.1007/s10530-016-1351-7

70. Bodin O, Crona B, Ernstson H. Social networks in natural resource management: What is there to learn from a structural perspective? Ecol Soc. 2006;11(2):r2. https://doi.org/10.5751/ES-01808-1102r02

71. Plummer R, Armitage D. A resilience-based framework for evaluating adaptive co-management: Linking ecology, economics and society in a complex world. Ecol Econ. 2007;61(1):62–74. https://doi.org/10.1016/j.ecolecon.2006.09.025

72. Ban NC, Mills M, Tam J, Hicks CC, Klain S, Stoeckl N, et al. A social–ecological approach to conservation planning: Embedding social considerations. Front Ecol Environ. 2013;11(4):194–202. https://doi.org/10.1890/110205

AUTHOR:
Bankole Falade[1]

AFFILIATION:
[1]South African Research Chair in Science Communication, Centre for Research on Evaluation, Science and Technology (CREST), Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Bankole Falade

EMAIL:
bankolefalade@gmail.com

Leveraging media informatics for the surveillance and understanding of disease outbreaks

Syndromic disease surveillance mechanisms can be enhanced by incorporating mass media informatics for disease discourse and aberration detection and social psychology for understanding risk perceptions and the drivers of uptake and resistance. Using computerised text analysis, the coverage of the outbreak of Zika virus in Brazil in 2017/2018 in four newspapers – *O Estado, O Globo*, the *Times* of London and the *New York Times* – was examined and patterns were compared with Google Trends. Quantitative indicators showed waves of attention to Zika peaked in the same period but local newspapers, *O Estado* and *O Globo,* indicated lower levels of anxiety in the run up to the Olympics when compared with foreign media. The unusual surge in attention to dengue in early 2015 was an early indication to sound the alarm for extensive clinical investigations. This, together with the flagging of Zika by *O Globo* almost a year before the global alarm, indicates the suitability of this method for surveillance and detection of aberrations. Media attention waves are also significantly associated with Google Trends, indicating empirical equivalence. Qualitative indicators show the extra motivation over Google, World Wide Web or Twitter searches by highlighting public perceptions. Findings show the absence of a stable body of scientific knowledge at the outbreak and an ensuing crisis of understanding. Local concerns were about the economic crisis, religious beliefs, poverty and crime – all inhibitors to containment – while the global alarm was amplified by risk to tourists and athletes, and political disputes mixed with religious beliefs.

Significance:

- This study contributes to research on the use of longitudinal media data as surrogate sources for syndromic disease surveillance.

- Mass media informatics provide empirical equivalence to Google Trends.

- Clinical and non-clinical factors contributed to public anxiety over disease epidemics.

- Lack of clinical knowledge at the onset of the crisis contributed to anxiety among scientists and the public.

Mass media and public health informatics

The news is a form of knowledge[1] and a feature article answers to the '5W's and H' principle of journalism: who, what, when, where, why and how. This principle is also true of social psychological approaches to knowledge[2] as the 'who', 'how', 'why', 'what' and 'what for' constitute the contexts of knowing and are central to communication, attribution, etc. The diagnosis of epidemiological[3] events also follows the same 5W's: what (diagnosis or health event), who (person), where (place), when (time) and why/how (causes, risk factors and modes of transmission). Thus, the conceptual approaches to mediated knowledge transfer by the media, the social psychological studies of knowledge in context and epidemiological studies of the distribution and determinants of health-related states or events are similar, in that they all address the '5W's and H'.

The threat of global pandemics posed by outbreaks of influenza H5N1, H1N1 and Severe Acute Respiratory Syndrome provoked interest in improving early warning systems and reinforced the need for combining data from different sources and experimentation with new types of media.[4,5] Rapid outbreak detection and the examination of a wide range of methods have also shown that no single detection approach is likely to be optimal.[6] Electronic information sources are proving to be valuable inputs and the Global Public Health Intelligence Network has shown that extensive monitoring and analysis of news media around the world can aid in early detection of disease threats.[7]

The news media tell the public what to think about by directing attention to problems and solutions and prioritising issues of importance.[8-10] Thus, they have become tools for understanding health problems in different contexts[11,12] and for understanding the psychosocial factors that propel epidemics[13]. Newspapers flagged opposition by religious authorities to vaccination in Nigeria in 2001, well ahead of the vaccine revolt of 2003–2004.[14]

Health informatics remains central to modern epidemiological practice and its advanced surveillance systems facilitate access to new data streams and the automation of processes for case and aberration detection.[6,7] Syndromic surveillance focuses on the early signals and an optimal system might be one that integrates data from multiple sources, clinical or surrogate, thereby potentially increasing investigators' confidence in the relevance of an alert from any single data source.[15]

News media informatics and the knowledge-in-context approach can form part of a complex of syndromic surveillance mechanisms for the early detection of related non-clinical discourses about health. Data mining of both current and archival media data can detect deviations from recognised patterns that may signal amplification of concern or the emergence of certain keywords that signify new entrants. It has been shown that changes in public anxiety about the H1N1 virus were detectable in people's online searches through analysing trends.[16]

The challenges of a novel and complementary data source and data analysis methodology notwithstanding[17], the method is being increasingly employed for passive monitoring. Twitter has been used to track levels of disease

activity and public concern and to improve surveillance[18,19]. Google has been used to monitor Ebola-related web search behaviour and dengue outbreaks.[20-22] Some authors have used a travel health website[23] to monitor anxiety over the Zika virus, others have examined the framing of the Zika crisis[24] in two Brazilian newspapers (*O Globo* and *Folha de Sao Paulo)* and the coverage of mosquito-borne diseases in Italian newspapers[25] and compared Zika with other diseases.

Newspaper coverage and Google Trends for the Zika virus in Brazilian, UK and US media were examined with the aim of contributing to a growing body of knowledge on the use of non-clinical sources for passive monitoring. Computer-assisted keyword analysis was used for disease discourse and aberration detection and to analyse trends indicative of waves of public concern. Knowledge crisis among scientists, perception crisis in the public and the psychosocial and economic inhibitors to containment were also examined.

Case study: The Zika virus disease

Alarm over the effect of the Zika virus on pregnancy in several parts of the Americas became a global phenomenon in the latter months of 2015 and early days of 2016. The virus had, however, been of concern in the Americas since 2014.[26,27] The first report of autochthonous transmission of Zika virus was reported in Chile in March 2014. In March 2015, the authorities in Brazil notified the World Health Organization (WHO) of an illness characterised by skin rash, but Zika was not suspected at that stage. In the same month, the Brazilian newspaper *O Globo* alerted to the circulation of Zika in some states and, in May, the Pan American Health Organisation (PAHO) issued an epidemiological alert on the circulation of Zika virus.[26,27]

In July 2015, Brazil reported neurological disorders associated with a history of infection including 49 confirmed cases of Guillain–Barré syndrome.[26] In October that year, the country again alerted WHO to an unusual increase in the number of children born with microcephaly – a neurological disease which manifests as unusually small heads in newborns. The following month, Brazil declared a national public health emergency over microcephaly and, shortly after, authorities in neighbouring countries including El Salvador, Guatemala and Mexico also reported the circulation of Zika virus. A continental transmission had begun with a corresponding global alarm. At this time, however, there was no confirmation of any link between Zika and microcephaly.[26]

The first report in the Western (UK and US) media sample for this study was in the London *Daily Telegraph* of 19 November 2015 titled 'The Zika virus: Answers about a growing menace'. The report said Brazil had declared a public health emergency on the confirmation of about 400 cases of newborns with small heads and the detection of the Zika virus in the amniotic fluid of two pregnant women whose foetuses were diagnosed with microcephaly. In January 2016, the US Centers for Disease Control (CDC) issued an alert for pregnant American women to avoid travel to the Americas, and on 1 February 2016, the WHO declared the disease a 'global health emergency'[26].

A mild virus of African origin turns invasive

Zika was discovered in monkeys in Uganda in 1947 and was later isolated in the *Aedes africanus* mosquito. The virus had no known severe effects in Africa and it is thought that the population has pre-existing immunity.[28] The disease moved from Africa to Asia and the first large outbreak in humans was in 2007 in the Pacific Island of Yap. Another outbreak followed in 2013 in French Polynesia, and, by 2015, Zika virus was spreading in Brazil. The *Aedes* species transmitting Zika in the Americas are the *Aedes aegypti* and *Aedes albopictus,* which also transmit yellow fever, dengue fever and chikungunya.[28-30]

Knowledge production by scientists about emerging and highly infectious diseases is often problematic at the onset of major outbreaks. This was the case with the Severe Acute Respiratory Syndrome[31], swine-origin Influenza A (H1N1) virus in humans[32] and other viral epidemics. In such situations of a crisis of knowledge, 'what' types of perceptions are expressed by the public, 'what' are the non-clinical discourses that may contribute to these public perceptions and 'why' and 'how' may these inhibit containment?

Researchers have examined science and society issues from two perspectives: the public understanding of science and risk perception[33]; however, both are functionally equivalent. The risk perception research, or the engineering approach, focuses on how scientists and the public perceive risk. While scientists focus on risk assessment using algorithms and normative rules, the public rely on intuitive risk judgements which are informed by images and associations, linked by experience to emotion and affect – thus when a decision is complex, they rely on readily available affective impressions.[34,35]

The public understanding of science perspective was initially focused on scientific literacy. This was criticised as a 'deficit model' which contrasts knowledgeable scientists against an ignorant public, but the field has transformed over the decades into a science and society paradigm which examines trust and expert deficit, the notion of 'public' and crisis of confidence. The new paradigm also acknowledges that the public can at times be inattentive, unmotivated and ignorant and public education becomes a priority.[36-39]

The Zika virus disease in Brazil was approached from the science and society[38] perspective of knowledge encounters and public sphere[40,41] and how the public perceive risk[35] from a virus described as the '*doença misteriosa*' or mystery disease. Scientific knowledge was problematic at the onset of the crisis, sometimes symptoms were confused with known diseases plausibly leading to a misdiagnosis or delayed diagnosis, and the public had to internalise the disease and its implications for health in a society with strong religious ties and ongoing political and economic crises.

Research objectives

The objective was to examine how disease surveillance and monitoring operations can leverage media reports and contribute to the growing body of evidence on the use of syndromic surveillance mechanisms.[18-22] The specific research questions were:

1. What are the quantitative indicators of public anxiety over the Zika virus? The focus here was on computerised text analysis for disease discourse and aberration detection, longitudinal media studies for public attention and patterns of public concern over time. Specifically:

    1.1 How early did the Zika virus discourse appear in the Brazilian media? How does the trend compare with those for similar diseases such as dengue and chikungunya? What are the levels of attention that indicate rise in public anxiety and an alarm for further clinical analysis? How do these indicators compare with Google Trends Brazil?

    1.2 What are the levels of attention in the media in the United Kingdom and USA? How do these indicators compare with Google Trends worldwide, and in the UK and USA?

2. What are the qualitative indicators[42] of the discourse about the Zika virus disease in the British and US press? The focus here being on psychosocial and economic drivers of perception, uptake or resistance.

Data and research methods

The analysis of public health surveillance data is typically performed to detect the presence of unusual patterns in comparison with historical data.[6] Mass media archives are also responsive to pattern recognition. Media is the common place through which communities cultivate shared and public notions about facts, values and the contingencies of human existence[42] over time, and thus there is stability in what is discussed and what is not. Unusual patterns and discourses then signal the appearance of a new phenomenon or the amplification of concerns over existing ones. However, being shared and collective does not denote consensus, rather it recognises the subcultural, class, generational, ideological, and even conflicts, among groups. Thus, studies of different newspapers within a country will reveal overarching community anxieties while also

retaining subcultural differences. However, local interests often cross national boundaries, and the focus of this study is on a disease epidemic ravaging a country but with the potential to globalise in our increasingly interconnected world.

Data for this study were obtained from *O Estado de São Paulo* (*O Estado*) and *O Globo*, two leading Brazilian newspapers; *The Times* of London and *The Telegraph* published in the UK and the US-based *USA Today* and *New York Times*. Coverage data were downloaded from the NEXIS database at various times between March 2017 and February 2018. Google Trends was used for online searches in Brazil, UK, USA and worldwide to compare with newspaper trends.

The US newspaper was chosen because of the presence of the mosquito carrier in some parts of the USA and the high risk of spread compared with the UK where the carriers are absent. The UK and US newspapers also provide indicators of when the media detected the events to report and the attention on two important events in Brazil (the Rio Carnival and the Olympics) that may have affected foreign visitors from these countries. The Brazilian newspapers were chosen to identify trends and aberrations to media attention in the local press within the review period.

For *O Globo* and *O Estado*, the NEXIS search was conducted using the keywords dengue or Zika or chikungunya between 1 January 2015 and 31 December 2016. Using these keywords, 503 articles for *O Globo* and 813 for *O Estado* were downloaded. For the UK and US newspapers, the keyword used was 'Zika'. There were no search results returned for 'Zika' between 1 January 2015 and 31 October 2015. Blogs, duplicates and articles with high levels of similarity were discarded. In total, 246 articles from *The Times* of London and 414 from the *New York Times* were used for the trend analysis. Google Trends[21] was used for online searches. For Google Brazil, the keywords used for the search were 'Zika', 'Dengue' and 'Chikungunya' to map the co-occurrence of searches about the three diseases from 1 January 2015 to 31 December 2016. The same dates were used for 'Zika' on Google worldwide, Google UK and Google USA. The NEXIS output is in the daily format of newspapers and downloads both headlines and text. The daily format was re-coded to weekly to match Google weekly outputs.

The QDA/Miner WordStat program, which runs in both English and Portuguese, was used to analyse the data.

Limitations of research

Secondary data were used and are limited in time to 2 years and to selected newspapers in Brazil, UK and USA. The study of the Brazilian media was limited to detecting aberrations and trends and comparing with Google searches. Other authors[24] have examined the framing of the debate in the Brazilian media. It would be interesting to see how a wider longitudinal study of newspapers worldwide could be used to trace the movement of the disease in the Americas and when the debates about its presence emerged in other Latin American newspapers. This would provide more data to support growing research on mass media as surrogate data for epidemiological studies.

Findings

Research question 1.1

This question addressed how early the Zika virus discourse appeared in the Brazilian media. How did the trend compare with similar diseases such as dengue and chikungunya? What were the levels of attention that indicate a rise in public anxiety and an alarm for further clinical analysis? How did these indicators compare with Google Trends Brazil?

Figure 1 shows the co-occurrence of Zika and dengue, both with similar symptoms, in May 2015 in *O Globo*. May 2015 is the earliest appearance of the word Zika in the two Brazilian newspapers; the articles indicate the level of unfamiliarity with the disease at that time. The first report, on 14 May 2015, titled '*O novo vírus do 'aedes'*' ('The new 'aedes' virus') said (translated) 'It [Zika] seems less serious, but it is possible to acquire the three viruses at the same time, and we do not know the clinical course of this co-infection with the Zika virus, for which we do not have immunity'. A 3 June 2015 follow-up article titled '*Rio tem o primeiro caso confirmado de Zika*' ('Rio has the first confirmed case of Zika') shows health officials still regarded Zika as benign and advised clinicians to treat as dengue when in doubt. Reference to both diseases did not occur again until November 2015 with Zika later reaching its peak in the second week of February 2016 when WHO declared a global emergency. The Kendall's tau_b correlation coefficient from 13 December 2015 to 31 December 2016 shows a significant association between Zika and dengue ($r=0.70$; $p<0.001$) but not between Zika and chikungunya ($r=0.54$; $p<0.08$).

Figure 1b shows that dengue and chikungunya had been in the media discourse in Brazil, in *O Estado*, from 1 January 2015 when the sampling commenced, with coverage of dengue rising rapidly thereafter with a major peak in the first week of May 2015. This peak shows an unusual pattern – an indication of rising public anxiety – which should raise an alarm for further clinical investigation. May 2015 was also the month Zika appeared in *O Globo* (Figure 1a), indicating it was flagged by one local newspaper and not the other, also showing the need for multiple media sources. The experience of doctors was with dengue and Zika was thought, at this time, to be benign. This may explain the comment by a Brazilian official reported in the *New York Times* of 31 December 2015 that 'Zika virus doesn't worry us' because it is a 'benign disease'. The appearance of Zika peaked at about the same time in the second week of February 2016 in both Brazilian newspapers. There is a significant association between Zika and dengue ($r=0.69$; $p<0.001$) and between Zika and chikungunya ($r=0.68$; $p<0.001$) between 13 December 2015 and 31 December 2016.

Figure 1c compares Google searches for dengue, Zika and chikungunya. The Google search number represents an index of popularity. It is used here as a variable indicating an index of appearances. The graph shows that those seeking online information had a much higher and sustained interest in dengue in the review period, except in early December 2015 when it was surpassed by Zika. The December surge coincides with similar surges in *O Estado* and *O Globo*. Searches for Zika were noticeable from the beginning of the review period in very low numbers, mostly one, with the highest peak recorded in the second week of February 2016, as with the newspapers. For Google and newspaper searches between 13 December 2015 and 31 December 2016, the correlation was significant for Zika in Google Brazil and *O Estado* ($r=0.65$; $p<0.001$) and *O Globo* ($r=0.59$; $p<0.001$). The correlation was also significant for dengue in Google Brazil and *O Estado* ($r=0.69$; $p<0.001$) and *O Globo* ($r=0.53$; $p=0.001$).

Coverage by both newspapers was compared over the period. The correlation for Zika in *O Globo* and *O Estado* was significant ($r=0.54$; $p<0.001$) as was the correlation for dengue ($r=0.61$; $p<0.001$).

Research question 1.2

This question addressed the levels of attention in the UK and US media. How did attention levels compare with Google Trends UK, USA and worldwide? For the UK, a comparison was made between Google UK, Google worldwide and *The Times* of London. Figure 2a shows that coverage in Google UK and *The Times* of London peaked at the same time, the third week of January 2016, a week before Google worldwide, and remained high until March 2016. The Rio Carnival was held in February 2016 and there is a noticeable peak in coverage in Google UK at this time. Coverage in the *Times of London* peaked again in the last week of July, just before the commencement of the Rio Olympics held from 5 August to 21 August 2016. The July to August surge in coverage in *The Times* of London, higher than the February peak, was not replicated in Brazilian newspapers[42], indicating differences in the perception of risk and anxiety. There was a significant association for the keyword Zika between *The Times* of London and Google UK ($r=0.37$; $p<0.001$) and between *The Times* of London and Google worldwide ($r=0.30$; $p<0.008$).

The *New York Times* data show a similar trend with those of *The Times* of London. The first peak in coverage in the *New York Times* was in late January 2016, coinciding with that in Google worldwide but a week after coverage in Google USA peaked. Coverage in the *New York Times* reached another height 2 weeks later, followed by several highs and lows mimicking the UK data in the run up to the Olympics with a noticeable
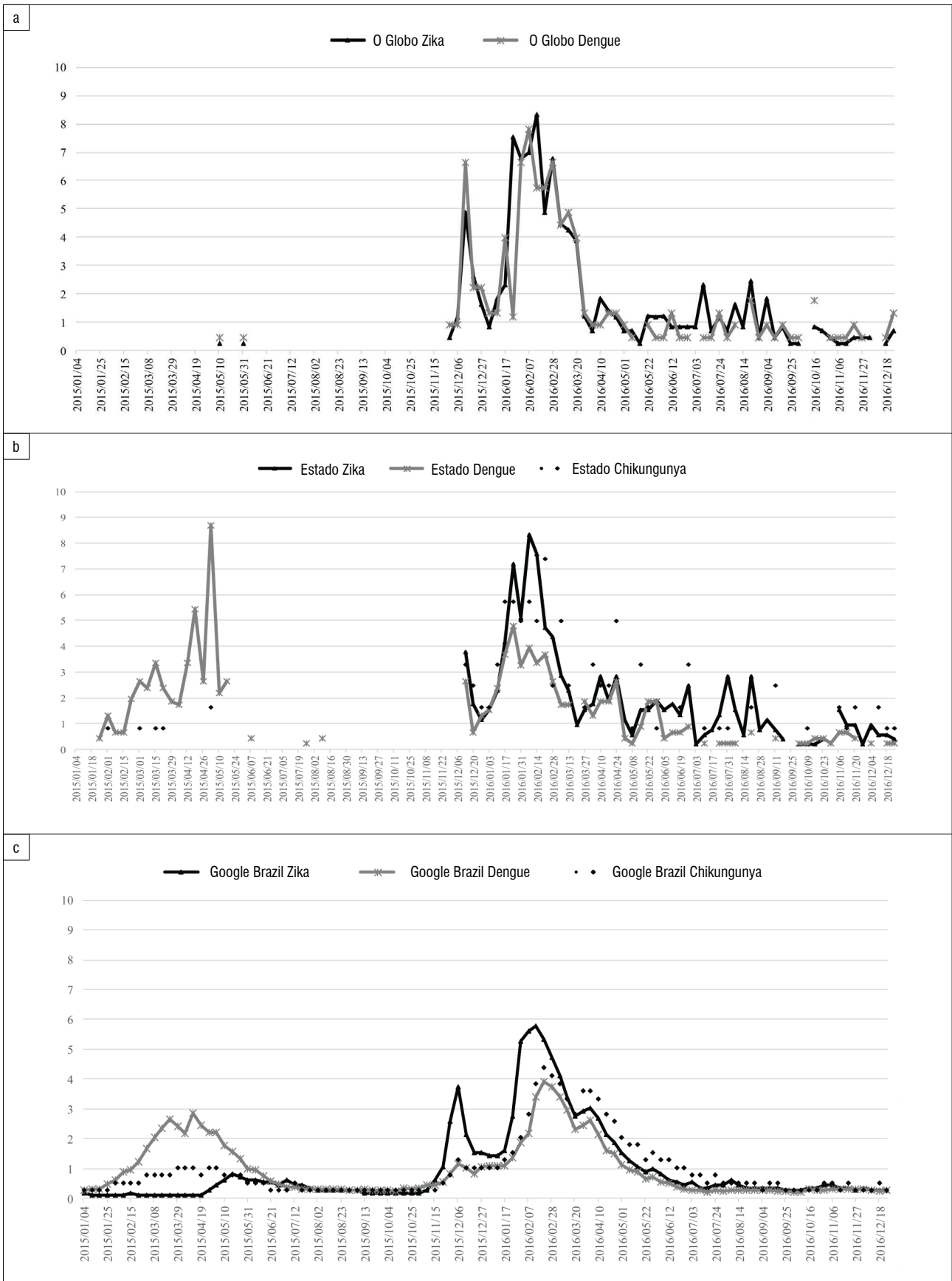
Figure 1: Coverage of (a) Zika and dengue in *O Globo* (*n*=503), (b) Zika, dengue and chikungunya in *O Estado* (*n*=813) and (c) Zika (*n*=1123), dengue (*n*=2564) and chikungunya (*n*=389) in Google Brazil. The *y*-axis represents the percentage of weekly coverage relative to the total for the period under review and the *x*-axis the week under review.

Figure 2: Comparison of Google Trends worldwide index with media coverage: (a) *Times* of London (*n*=246), Google UK (*n*=558) and Google worldwide (*n*=900) and (b) *New York Times* (*n*=414), Google USA (*n*=866) and Google worldwide (*n*=900). The *y*-axis represents the percentage of weekly coverage relative to the total for the period under review and the *x*-axis the week under review.
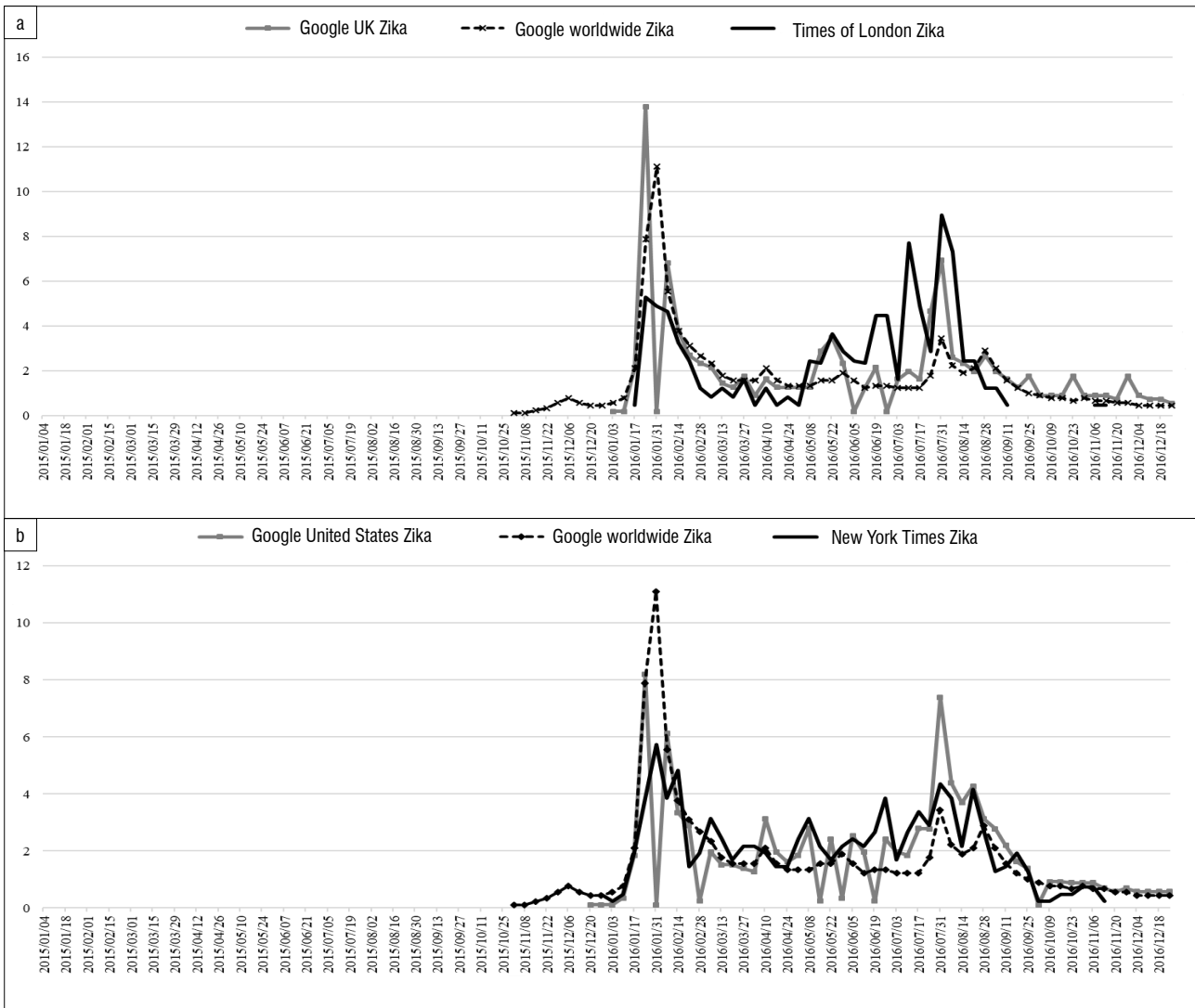
July to August surge. There was a significant association in coverage of Zika between the *New York Times* and Google USA (*r*=0.44; *p*<0.001) and between the *New York Times* and Google worldwide (*r*=0.52; *p*<0.001). The Kendall's tau_b correlation coefficient also shows significant correlations for the weekly appearance of Zika in *The Times* and the *New York Times* (*r*=0.55; *p*<0.001).

Debates about the Zika virus began in the foreign media – in the *Daily Telegraph* of 19 November 2015 – about 8 weeks before the first travel alert by the CDC on 15 January 2016. Reports later appeared in the *Sunday Times* of 13 December 2015, *USA Today* of 18 December 2015 and the *New York Times* of 29 December 2015. This is an indication of global anxiety ahead of the confirmation of a major outbreak. The disease had, however, been flagged in Brazil much earlier – in the second week of May 2015 – and public anxiety had risen much earlier, from February through April, over dengue, which presents similar symptoms to those of Zika.

Altogether, a collection of world newspapers on an ongoing basis produces a very similar trend chart to that for Google online searches and can thus form part of a routine surveillance system of non-clinical data. The process can also be automated for case and aberration detection[3,6,7] mapping the movement of invasive foreign species across continents. In addition to similarities with Google searches, the February peak is similar to Twitter data[43] in the same period. The February and August peaks are also similar, coinciding with visits to a UK travel health website.[23]

In addition, as will be shown in the next section, newspapers provide information on the scientific and non-scientific discourses about disease, public perceptions and how these may have influenced the uptake of containment information.

Research question 2

What are the qualitative indicators[42] of the discourse about the Zika virus disease in the UK and US press?

Crisis of knowledge, the known and unknowns

The first part of this section examines a crisis of knowledge among scientists during the early days of the outbreak while the second part reviews public perceptions of risk from the virus and psychosocial and economic inhibitors to containment – representations[34,35,40,41] and the crisis of mind. This section is limited to the US and UK media – from 'the outside looking in' (see Ribeiro et al.[24] for the framing of the debate in the Brazilian media).

Diagnosis

The inexperience with Zika disease meant existing facilities were inadequate, leading to several false positives and fears Zika may have been misclassified. Scientists were also not certain at the early stages of the outbreak about the causative agent. One doctor said[44]:

We don't know if it's only Zika or if it's a combination of Zika, dengue and chikungunya… Maybe a woman was infected by dengue a year before, and now is pregnant and gets Zika.

How the virus crosses the placenta to the foetal brain to cause damage – a condition referred to as microcephaly – was also unknown.[45] The extent of the dengue problem was captured by a news report which stated that Brazil registered more than 1.6 million cases of dengue, with 863 people dying of the disease the previous year.[46] There were also fears expressed in the media that the infection could lead to other serious diseases such as Guillain–Barré syndrome, which attacks the nervous system and can lead to fatal paralysis.

The government was also unsure about what instructions to issue for diagnosis. Health professionals were originally told to report suspected microcephaly cases when a baby's head at birth was 330 mm, or 13 inches, or less. The government then changed the threshold to 320 mm after concluding that there were many babies with small heads but without problems. Later, discussions began on lowering the circumference to 319 mm for boys and 315 mm for girls.[47]

Transmission

There were a lot of unknowns about the mode of transmission of the virus, whether it was a single route through mosquito bites, or whether there were multiple routes of transmission. Reports later emerged that the disease could be transmitted through sexual intercourse, blood and other body fluids such as saliva and could remain in semen for more than 2 months after infection. The uncertainty led to some taking extreme precautions:

We've also made the decision to have Greg's sperm frozen. We'd love to have more children and with research in its infancy, I wouldn't want to put myself in a situation which could have been prevented…[48]

Causative consensus

In April 2016, more than a year after the disease epidemic became a public concern in the Americas and over 4 months after the crisis became a global concern, the WHO agreed that there was a strong link between Zika and microcephaly in newborns. The announcement was not without debate though, as the CDC researchers, who also confirmed the link, sidestepped the more stringent Koch's Postulates, which require that scientists must re-isolate the suspected pathogen after infecting healthy people or animals with the suspected pathogen, and the Bradford Hill Criteria, a model for studying infectious diseases. Instead, a framework called Shepard's Criteria for Teratogenecity, which has a lower threshold, was adopted.[49]

Despite agreement on the cause, much remained unknown about the effects of Zika, including effects on other organs, the risk for pregnant women of having brain-damaged babies, and to what extent the risk varies according to when in pregnancy the infection occurs and other types of brain defects possible.

Risk perception, psychosocial and economic inhibitors

We perceive new issues and objects through our existing images of the world. Thus, how we initially react to a new phenomenon is determined largely by our shared experiences of similar objects, our culture and aspirations.[50] Invasive microbes have been in the news worldwide with their devastating consequences resulting in global panic. Previously unknown Severe Acute Respiratory Syndrome (SARS) spread from Guangdong Province in China to 37 countries including Canada, Singapore and Taiwan, with 774 reported deaths.[51] The Ebola virus[13] caused global panic when it resulted in the death of thousands in West Africa, where it was previously unknown. It is these shared images of fatalities with which the public perceive new viruses with the potential to globalise in new areas.

The arrival of Zika in a previously uninfected Latin America, its capacity to spread and the debilitating effect on the foetus caused a global panic. Brazil, a major tourist destination, was to host the annual Rio carnival and the 2016 summer Olympics, both of which would attract thousands of foreign visitors, which served to amplify anxiety over capacity of the disease to spread to new countries.

Local and international perceptions of risk

On the international scene, there was a high level of anxiety over the possible exposure to Zika of athletes and their family members during the Olympic Games. Some athletes withdrew from the Olympic Games, others followed safety measures. Over 100 leading scientists called for the Games to be postponed:

An unnecessary risk is posed when 500,000 foreign tourists from all countries attend the Games, potentially acquire that strain, and return home to places where it can become endemic… The Games should be moved or postponed.[52]

Despite one expert warning that the Olympics could become the 'the Olympics of brain damage', the International Olympic Committee rejected calls for its postponement. The Committee's response followed a statement from the WHO that there was 'no public health justification for postponing or cancelling the Games'[53].

Some athletes were not worried. One told a newspaper that the Games were more important than 'the bugs': 'It's the Olympics, it's the Olympics! … Mosquitoes? Like, whatever, I'm going. This is my shot.'[54] Others expressed concerns about safety.

The reaction of holidaymakers was of fear leading to several cancellations and anxiety among tourism firms and hotels over the financial implications. A travel consultant said she had 10 cancellations in a week, mostly for babymoons (vacations that some parents take before the impending arrival of their baby). Another travel consultant said she had 48 cancellations for trips to the Caribbean and Mexico:

I have a lot of clients who are pregnant or trying to get pregnant, so they're scared and are halting their travel plans altogether or rebooking to travel within the US instead.[55]

In Brazil, public apathy to the disease was evident during the annual Rio carnival where scantily dressed women danced on the streets in open defiance of the mosquitoes and warnings. In Salvador, a city hit hard by Zika, a lady said: 'Do I look worried? Ask me next week, after Carnival is over.' 'Carnival', the report added, 'is like anaesthesia' as it stops people from thinking about the problems we have…'

No crisis has ever diminished the magic and excitement of Carnival. During pessimistic and depressing times, it becomes more important ... It's escapism on steroids.[56]

An hotelier, however, pointed out that Brazilians have problems much bigger than Zika. He said: 'Most of my friends are more worried about finding jobs', and added: 'If you're not a pregnant woman, you don't need to worry.' He also noted that dengue fever, which killed more than 800 people in Brazil the previous year, was far more dangerous.[56]

The Church, politics and abortion

The Catholic Church is opposed to abortion and all forms of contraception, but these are options for women against the disease, reigniting the debate over church, state and choice in Brazil. Both are also bitterly contested in the United States Congress between Democrats and Republicans.

A Cardinal in Brazil said mothers must accept babies born with microcephaly 'as a mission', and that abortion was not an option, but Pope Francis was more open to contraception, arguing that avoiding pregnancy is 'not an absolute evil like abortion which is a crime and an absolute evil'[57].

Some Brazilians sought solace in God. A woman, when confronted with a diagnosis of possible microcephaly for her foetus said: 'It's God's will: he wanted us to have a baby like this.'[58]

There was, however, anxiety over the future of the infected babies. A woman with an infected 3-month-old son wondered about his future and his integration into society:

> What kind of life would he have? It gets me angry when someone on the bus looks at Arthur and asks, What's wrong with his head? … I tell them, Nothing's wrong, he's just different. But then I think to myself, Yes, something's wrong. My son will never be like the other boys.[59]

In the USA, a report noted that while the feud on Capitol Hill over responding to the virus appeared to be over how much money was needed, beneath the surface were issues that have long stirred partisan mistrust, including Republicans' fears about the use of taxpayers' money for abortion and possible increased use of contraception, and Democratic worries about protecting the environment from potentially dangerous pesticides.[60]

Political, economic and social crisis

It was also a period of severe political and economic crisis in Brazil, leading to cuts in public spending with implications for funding prevention strategies and health care. The epidemic also exposed the wide gap between the rich and the poor in Brazil.

> The epidemic mirrors the social inequality of Brazilian society as it is concentrated among young, poor, black and brown women, a vast majority of them living in the country's least-developed regions. The women at greatest risk of contracting Zika live in places where the mosquito is part of their everyday lives, where mosquito-borne diseases like dengue and chikungunya were already endemic.[61]

The same article indicated that the rich did not feel too concerned:

> I am a Brazilian woman. My friends who are planning to have children soon are worried about Zika. But they don't need to be too concerned. In our well-to-do neighbourhood in Brasília, the capital, there has not been a single case of a baby with the birth defects associated with the Zika epidemic.[61]

Violence among the gangs in the *favelas* (slums) also posed limitations to containment efforts as even the army, brought in to help clear mosquito breeding sites, were limited to safe areas.

> The army has been told not to go into areas controlled by gangs, which could provoke clashes. Health workers negotiate with residents' committees often linked to crime bosses, for safe entry.[62]

The blame game

Pharmaceutical companies were targets of blame, on the one hand for neglecting diseases which were not deemed profitable, and on the other hand for spreading diseases for profit. Some commentators also blamed urbanisation and air travel – offshoots of social and economic progress.

> The same conditions that drove cholera, overcrowding, corruption, poor hygiene, expanded transportation, are what drive pandemics today. In the age of cholera, shipping trade helped spread disease, today, it's air travel. If you want to understand the helter-skelter distribution of flu outbreaks, simply consult a map of airline routes.[63]

Rumours and conspiracy theories

Rumours are known to thrive during major health crises as the public attempts to find socio-political and economic origins of disease. Some people suggested that the plea by the government of El Salvador to delay having children was an effort to stem population growth. A lady said 'This government doesn't want anyone else … We don't fit already!'[64]

There were other interests opposed to genetic modification of insects and global powers. An article read:

> The Zika virus, some Brazilians are convinced, is the inadvertent creation of a biotech company that has been releasing genetically modified mosquitoes to combat dengue fever in Brazil. Others here and elsewhere see it as a plot by global elites to depopulate the earth and install a 'one-world government'.[65]

Conclusions

The detection of aberrations in patterns, the emergence of discourse about disease and public perceptions bring together three areas of research: epidemiology, media and social psychology. Combining multiple approaches allows epidemiologists to harness the strengths of individual research paradigms for a whole that has the potential to be greater than the sum of its parts.

The news media are fairly stable artefacts in cultural contexts; deviations or aberrations indicating new disease debates or a sudden upsurge in coverage of known diseases or the flagging of potential obstacles to scientific interventions should trigger the need for further clinical investigation and early interventions. As no single approach is optimal and extensive monitoring aids detection[6,7], media informatics, as this research has shown, has great potential as part of a web of disease surveillance mechanisms[16,18,21].

The high upsurge in media discourse about dengue in Brazil in 2015 was an early indication to sound the alarm for extensive clinical investigations on the causes of the sudden aberration. This shows that a sudden rise in public anxiety about a known disease was the indicator of the emergence of an alien species with similar symptoms. This finding affirms the suitability of longitudinal mass media data as part of a complex of syndromic surveillance mechanisms. The flagging of Zika by one newspaper and not another in the sample also shows the need to use multiple sources for surveillance.

Qualitative indicators show the extra motivation over Google, web or Twitter searches in that they further mirror the perception of the public about the disease and psychosocial factors that may enhance or inhibit uptake. The Zika virus epidemic was not a case of public deficit in knowledge of the disease in contrast to scientists[36,37] or of mistrust or expert deficit[36]. This was an instance of a virus, benign in its natural habitat, which crossed the continent and became virulent in another country. The virus also presents with the same symptoms of locally established diseases such as dengue and chikungunya. Zika was the '*doença misteriosa*', or mystery disease, which initially puzzled both scientists and the public.

The knowledge crisis, inadequate infrastructure to detect the presence of the disease and its capacity to globalise led to a global alarm and the declaration of a global health emergency. International concern focused on the risk to travellers for the Rio Carnival, the Olympics and tourism, all of which attract thousands of visitors to Brazil with the risk of being infected and transmitting the disease to their home countries. Local issues were about prevention, obstacles to containing the spread of the disease, the blame game, rumours and living with long-term effects on the children born to infected mothers.

The spread of the Zika virus has shown that in the age of international mobility, classification of a disease as benign is no longer sustainable and surveillance needs to be extended to uninfected areas, particularly where possible vectors of the disease can be found. Also in the age of massive computing power and web crawlers, monitoring public anxiety over diseases around the world in real time using a keyword approach is increasingly possible. Scientists also need to investigate why some diseases appear harmless in some communities, the source of immunity in these populations, the applicability of the immunity to vaccine production and the consequences if the disease appears in new areas – that the disease is not yet there does not preclude it from reaching there.

Acknowledgements

References

1. Park E. News as a form of knowledge. A chapter in the sociology of knowledge. Am J Sociol. 1940;45(5):669–686. https://doi.org/10.1086/218445

2. Jovchelovitch S. Knowledge in context: Representations, community and culture. New York: Routledge; 2007.

3. Centers for Disease Control and Prevention. Principles of epidemiology in public health practice: An introduction to applied epidemiology and biostatistics [webpage on the Internet]. No date [updated 2012 May 18; cited 2018 Jan 01]. Available from: https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section1.html

4. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: A chronology of innovation. J Med Internet Res. 2013;15(7), e147, 13 pages. https://doi.org/10.2196/jmir.2740

5. Aslam AA, Tsou MH, Spitzberg BH, An L, Gawron JM, Gupta KD, et al. The reliability of tweets as a supplementary method of seasonal influenza surveillance. J Med Internet Res. 2014;16(11), e250, 12 pages. https://doi.org/10.2196/jmir.3532

6. Groseclose SL, Buckeridge DL. Public health surveillance systems: Recent advances in their use and evaluation. Annu Rev Public Health. 2017;38:57–79. https://doi.org/10.1146/annurev-publhealth-031816-044348

7. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports. J Am Med Inform Assoc. 2008;15(2):150–157. https://doi.org/10.1197/jamia.M2544

8. McQuail D. The influence and effects of mass media. Mass Commun Soc. 1977:70–94.

9. McCombs ME, Shaw DL. The agenda-setting function of mass media. Public Opin Q. 1972;36(2):176–187. https://doi.org/10.1086/267990

10. Cohen BC. The press and foreign policy. Princeton, NJ: Princeton University Press; 1963.

11. Washer P, Joffe H. The "hospital superbug": Social representations of MRSA. Soc Sci Med. 2006;63(8):2141–2152. https://doi.org/10.1016/j.socscimed.2006.05.018

12. Joffe H, Haarhoff G. Representations of far-flung illnesses: The case of Ebola in Britain. Soc Sci Med. 2002;54(6):955–969. https://doi.org/10.1016/S0277-9536(01)00068-5

13. Falade BA, Coultas CJ. Scientific and non-scientific information in the uptake of health information: The case of Ebola. S Afr J Sci. 2017;113(7/8), Art. #2016-0359, 8 pages. https://doi.org/10.17159/sajs.2017/20160359

14. Falade B. Familiarizing science: A Western conspiracy and the vaccination revolt in Northern Nigeria. Pap Soc Represent. 2015;24:3.1–3.24.

15. Henning KJ. Overview of syndromic surveillance. What is syndromic surveillance? Morbidity Mortality Weekly Rep. 2004;53(Suppl):5–11 [article online]. No date [updated 2004 Sep 09; cited 2018 Nov 12]. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/su5301a3.htm

16. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: Blogs, newspaper articles, and Wikipedia visits. J Health Commun. 2012;27(2):179–185. https://doi.org/10.1080/10410236.2011.571759

17. Müller O, Junglas I, Vom Brocke J, Debortoli S. Utilizing big data analytics for information systems research: Challenges, promises and guidelines. Eur J Inf Syst. 2016;25(4):289–302. https://doi.org/10.1057/ejis.2016.2

18. Santillana, M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein, JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol. 2015;11(10), e1004513, 15 pages. https://doi.org/10.1371/journal.pcbi.1004513

19. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. PLoS One. 2011;6(5), e19467, 10 pages. https://doi.org/10.1371/journal.pone.0019467

20. Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, et al. Assessing Ebola-related web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes. Infect Dis Poverty. 2015;4(1), Art. #54, 13 pages. https://doi.org/10.1186/s40249-015-0090-9

21. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: A systematic review. PLoS One. 2014;9(10), e109583, 49 pages. https://doi.org/10.1371/journal.pone.0109583

22. Chan EH, Sahai V, Conrad C, Brownstein JS. Using web search query data to monitor dengue epidemics: A new model for neglected tropical disease surveillance. PLoS Negl Trop Dis. 2011;5(5), e1206, 6 pages. https://doi.org/10.1371/journal.pntd.0001206

23. Petersen J, Simons H, Patel D, Freedman J. Early detection of perceived risk among users of a UK travel health website compared with Internet search activity and media coverage during the 2015–2016 Zika virus outbreak: An observational study. BMJ Open. 2017;7(8), e015831, 6 pages. https://doi.org/10.1136/bmjopen-2017-015831

24. Ribeiro B, Hartley S, Nerlich B, Jaspal R. Media coverage of the Zika crisis in Brazil: The construction of a 'war' frame that masked social and gender inequalities. Soc Sci Med. 2018;200:137–144. https://doi.org/10.1016/j.socscimed.2018.01.023

25. Giardullo P. Spreading mosquitoes: A media analysis of Italian national newspaper coverage of mosquito-borne diseases and related interventions. In: Claeys C, editor. Mosquito management: Environmental issues and health concerns. Brussels: PIE Peter Lang; 2018. p. 84–111.

26. Pan American Health Organization/World Health Organization. Timeline of emergence of Zika virus in the Americas [webpage on the Internet]. c2016 [updated 2017 Jan 17; cited 2017 Sep 09]. Available from: http://www.paho.org/hq/index.php?option=com_content&view=article&id=11959%3Atimeline-of-emergence-of-zika-virus-in-the-americas&catid=8424%3Acontents&Itemid=41711&lang=en

27. Zanluca C, Melo VCAD, Mosimann ALP, Santos GIVD, Santos CNDD, Luz K. First report of autochthonous transmission of Zika virus in Brazil. Mem Inst Oswaldo Cruz. 2015;110(4):569–572. https://doi.org/10.1590/0074-02760150192

28. Hayes EB. Zika virus outside Africa. Emerg Infect Dis. 2009;15(9):1347–1350. https://doi.org/10.3201/eid1509.090442

29. Chen LH, Hamer DH. Zika virus: Rapid spread in the western hemisphere. Ann Intern Med. 2016;164(9):613–635.

30. World Health Organization. The history of Zika [webpage on the Internet]. No date [updated 2017 Feb 01; cited 2017 Dec 12]. Available from: https://www.who.int/emergencies/zika-virus/timeline/en/

31. Peiris JSM, Guan Y, Yuen KY. Severe acute respiratory syndrome. Nat Med. 2004;10(12s):S88–S97. https://doi.org/10.1038/nm1143

32. Goodwin R, Haque S, Neto F, Myers LB. Initial psychological responses to Influenza A, H1N1 ("Swine flu"). BMC Infect Dis. 2009;9(1), Art. #166, 6 pages. https://doi.org/10.1186/1471-2334-9-166

33. Bauer M. Atoms, bytes and genes – Public resistance and techno-scientific responses. New York: Routledge; 2014.

34. Slovic P. Perception of risk. Science. 1987;236(4799):280–285. https://doi.org/10.1126/science.3563507

35. Slovic P, Finucane ML, Peters E, MacGregor DG. Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk and rationality. Risk Anal. 2004;24(2):311–322. https://doi.org/10.1111/j.0272-4332.2004.00433.x

36. Wynne B. Rationality and ritual: The Windscale inquiry and nuclear decisions in Britain. Chalfont St. Giles: British Society for the History of Science; 1982.

37. Ziman J. The public understanding of science. Sci Technol Hum Val. 1991;16(1):99–105. https://doi.org/10.1177/016224399101600106

38. Bauer MW, Allum N, Miller S. What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. Public Underst Sci. 2007;16:79–95. https://doi.org/10.1177/0963662506071287

39. Einseidel EF. Understanding publics in the public understanding of science. In: Dierkes M, Grote V, editors. Between understanding and public trust: The public, science and technology. London: Gordon and Breach Publishing Group; 2000. p. 144–149.

40. Jovchelovitch S, Priego-Hernandez J. Cognitive polyphasia, knowledge encounters and public spheres. The Cambridge handbook of social representations. Cambridge: Cambridge University Press; 2015. p. 163–178. https://doi.org/10.1017/CBO9781107323650.014

41. Moscovici S. La Psychanalyse, son image et son public [Psychoanalysis: Its image and its public. Cambridge: Polity Press; 2008. French.

42. Gerbner G. Toward "cultural indicators": The analysis of mass mediated public message systems. AV Commun Rev. 1969;17(2):137–148.

43. Southwell BG, Dolina S, Jimenez-Magdaleno K, Squiers LB, Kelly BJ. Zika virus-related news coverage and online behavior, United States, Guatemala, and Brazil. Emerg Infect Dis. 2016;22(7):1320–1321. https://doi.org/10.3201/eid2207.160415

44. McNeil D. Influx of brain damage may be linked to virus. The New York Times. 2015 December 29;6.

45. McNeil D, Romeo S. Pregnant women warned on Latin American trips. The New York Times. 2016 January 16;3.

46. Romero SA. Struggle against mosquito-borne diseases. The New York Times. 2016 January 18;4.

47. Sreeharsha V. Guidelines may lead to over-reporting of defect in Brazil. The New York Times. 2016 February 04;10.

48. Hirst P. Rutherford freezes sperm over Zika fear. The Times of London. 2016 June 08;76.

49. Belluck P, McNeil D. Health officials say Zika damages infant brains. The New York Times. 2016 April 14;16.

50. Moscovici S. The phenomenon of social representations. In: Farr RM, Moscovici S, editors. Social representations. Cambridge: Cambridge University Press; 1984. p. 3–69.

51. Smith RD. Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. Soc Sci Med. 2006;63(12):3113–3123. https://doi.org/10.1016/j.socscimed.2006.08.004

52. Whipple T. Move Rio Olympics over virus; Zika threat too great, say 100 leading scientists. The Times of London. 2016 May 28;1–2.

53. Boswell J, Leake J. Olympic chiefs dismiss Zika fears for Rio. The Sunday Times. 2016 May 29;2.

54. Maccur J. Opting to skip Olympics because of Zika fears. The New York Times. 2016 June 04;7.

55. Vora S. How the Zika virus is affecting travel. The New York Times. 2016 January 29;In Transit. Available from: https://www.nytimes.com/2016/01/29/travel/how-the-zika-virus-is-affecting-travel.html

56. Jacobs A, Perpétua S. Zika? Partying Brazilians offer collective shrug. The New York Times. 2016 February 11;1.

57. Kington T. Pope ready to lift condom ban in fight against Zika. The New York Times. 2016 February 19;5.

58. Bowater D. How the Zika virus is striking at the heart of terrified families in Brazil. The Daily Telegraph, London. 2016 January 30;7.

59. Romeo S, Sreeharsha V. Tears and bewilderment at center of Zika crisis. The New York Times. 2016 January 30;1.

60. Herszenhorn D. Partisan mistrust adds complexity to congressional feuding over Zika. The New York Times. 2016 May 25;15.

61. Doniz D. Zika and a woman's right to choose. The New York Times. 2016 February 08;25.

62. Hider J. Fear and conspiracy theories hinder Brazil's Zika battle. The Times of London. 2016 February 22;2016: 34.

63. Senior J. Hosting the perfect pathogen party. The New York Times. February 25, 2016: 1.

64. Ahmed A. El Salvador's advice on Zika: Don't have babies. The New York Times. 2016 January 26;1.

65. Jacobs A. Conspiracy theories about Zika spread through Brazil with the virus. The New York Times. 2016 February 17;6.

AUTHORS:
Anita Wheeler[1,2] iD
Andrew T. Knight[3,4,5] iD
Mark Difford[5] iD
Susanne Vetter[1] iD

AFFILIATIONS:
[1]Department of Botany, Rhodes University, Grahamstown, South Africa
[2]CapeNature, Gouritz Corridor, Oudtshoorn, South Africa
[3]Department of Life Sciences, Imperial College London, Ascot, United Kingdom
[4]ARC Centre of Excellence in Environmental Decisions, University of Queensland, Brisbane, Queensland, Australia
[5]Department of Botany, Nelson Mandela University, Port Elizabeth, South Africa

CORRESPONDENCE TO:
Susanne Vetter

EMAIL:
s.vetter@ru.ac.za

Ostrich farmer characteristics predict conservation opportunity

Environmental sustainability rests on human choice and action. Understanding these may assist in determining the factors that predict or influence an individual's behaviour towards the environment. In South Africa, approximately 80% of the most threatened vegetation types are in the hands of the private agricultural community. In the Little Karoo, which is situated in the Succulent Karoo biodiversity hotspot, unsustainable land-use practices including ostrich flock breeding threaten this region's lowland biodiversity. We interviewed ostrich farmers in the Oudtshoorn Basin to quantify latent variables thought to represent components of conservation opportunity: environmental attitude, conservation knowledge, conservation behaviour, and willingness to collaborate with agricultural, environmental and conservation organisations. Three groups of land managers were identified: (1) younger land managers (<31 years' farming experience) with bigger farms (≥2050 ha) who had above-average scores for all four indicators, (2) older farmers (≥31 years' farming experience) who had above-average scores for environmental attitude and conservation knowledge, average scores for environmental attitude, but low willingness to collaborate, and (3) a large group of younger farmers (<31 years' farming experience) with smaller properties (<2050 ha) who had low to average scores for all four indicators. Farmers in the first two groups represent the best opportunities for conservation, although different strategies would have to be employed to engage them given the current low willingness to collaborate among older farmers. Land managers were more willing to collaborate with agricultural than conservation organisations, pointing to a need to involve agricultural organisations in championing more environmentally sustainable ostrich breeding practices.

Significance:

- Achievement of biodiversity conservation targets requires stewardship in production landscapes outside protected areas, which necessitates identification of farmers who present conservation opportunity, i.e. who are willing and able to participate in conservation.

- Plant biodiversity in the Little Karoo has been severely degraded through ostrich flock breeding, but ostrich farmers consider their practices to be ecologically sustainable.

- In the Little Karoo, land managers with more years of farming experience, and younger farmers with larger properties, represented the greatest opportunity for interventions to promote more biodiversity-friendly ostrich farming practices.

Introduction

Changing environmental behaviour

Sustainable development is driven by human choice and action. Changes in individual behaviour can contribute significantly to reduce human impacts on the environment.[1] Environmental impacts by agriculture are caused by the behaviour of a relatively low number of people, but with a high per capita impact. Changing farmer behaviour thus has the potential to have a substantial effect on the global environment, including a reduction in biodiversity losses. This approach requires an understanding of how farmers' attitudes and behaviours are interrelated, and what interventions are therefore most likely to be effective in changing farmers' behaviour.

To influence and change environmental behaviour, it is necessary to understand the factors that determine behaviour and how they interrelate. Several studies in the agricultural context[2-4] have examined the relationship between land managers' use of conservation practices that impact the environment and the personality characteristics of environmental attitude and moral reasoning about the environment. These studies suggest that farmers tend to be less concerned about the environment than other groups, because of the nature-exploitative character of most farming activities.

An individual's enduring disposition toward the environment is referred to as an environmental attitude[4] and can be a direct predictor of behavioural intention, which in turn is a key determinant of behaviour.[5,6] There are many environmental attitude measures available based on different conceptual frameworks, but there seems to be a consensus that environmental attitude is multidimensional and organised in a hierarchical fashion.[7] Commonly measured dimensions related to environmental attitude are conservation knowledge, conservation behaviour, willingness to participate and willingness to collaborate.[8]

Identifying human and social factors that directly influence environmental and conservation behaviour can support implementation of more biodiversity-friendly land management.[8,9] Studies of land manager characteristics that affect the adoption of conservation practices have, until recently, remained limited.[10,11] The last 10 years have seen an increase in research on factors that lead farmers to adopt (or fail to adopt) sustainable farming practices (including in Mexico[12], Italy[13], Brazil[14] and Ethiopia[11]) and water management (e.g. in Australia[15] and Finland[16]). Recent research has also examined factors that promote or limit farmers' adoption of practices that conserve biodiversity (e.g. in Ireland[17], Brazil[14] and Slovenia[18]). Given that meeting biodiversity conservation targets relies

to a large extent on private landholders in production landscapes[9], more such research is needed, especially from biodiversity hotspots in less intensively studied parts of the world.

Biodiversity conservation in agricultural production landscapes

In South Africa, the conservation of threatened ecosystems now lies predominantly in the hands of private land managers[19] as approximately 80% (by area) of the most rare and threatened vegetation types occur in privately owned production landscapes. One of the first studies on how socio-economic, agro-physical, demographic and human behavioural factors affect conservation attitudes and behaviours was done in the Renosterveld on the West Coast of South Africa.[20,21] The authors concluded that for land managers who have to make a living from the land, the economic role of that land takes precedence over its aesthetic values. A study of the attitudes and behaviour of landholders toward the conservation of Renosterveld[22] showed that younger land managers were more willing to conserve, were not necessarily better educated, and owned larger farms (>500 ha) with greater extents of remnant Renosterveld (>300 ha) than those less willing to conserve. Attitudes toward Renosterveld were generally negative because it was not economically advantageous to retain it.[22] In the Little Karoo, farmers' attitudes towards existing conservation policies were most likely to be improved via the provision of extension services and the public recognition of their contribution to private conservation areas.[23] Personal interaction had a greater potential than any other method for persuading land managers to change behaviour, while restrictive legislation was found to have no decisive influence on conservation behaviour in the Renosterveld.[24,25]

In order to allocate scarce resources effectively to improve conservation behaviour among farmers, it is useful to identify farmers who present 'conservation opportunity', i.e. those who are willing and able to participate in conservation.[9] In South African grasslands, an interaction between attitudes towards the relative importance of conservation and levels of interest in wattled crane conservation was found to be the best predictor of conservation behaviour such as conservation-friendly habitat management.[26] Research in the Eastern Cape of South Africa mapped feasible opportunities for implementing effective conservation[9] and land restoration[8] actions on private land based on the land manager's knowledge, behaviour, willingness to participate and willingness to collaborate. The findings provide conservation and land management professionals direction on where and how implementation of local-scale conservation should be undertaken.

Aims and focus of this study

The Little Karoo is a semi-arid agricultural region within the Succulent Karoo biodiversity 'hotspot' of South Africa[27], but extensive grazing and browsing by domestic small stock, and especially ostriches, over the last 120 years has resulted in severe degradation of more than 50% of the region[28]. This degradation has led to a loss in productivity and biodiversity over large areas.[29,30]

Free-range ostrich farming has been the leading cause of land degradation in the Little Karoo, and the currently prevalent grazing systems and land management practices are ecologically unsustainable.[31] Recommended stocking rates are considered economically unsustainable by farmers and ecologically unsustainable by conservationists.[32] Whereas the dominant ostrich farming practice in other countries is pen breeding, which has a smaller footprint, ostrich farmers in the Little Karoo have not readily adopted this system. The reasons for this are unclear, and little is known about the environmental attitudes, knowledge and behaviours of ostrich farmers or their willingness to cooperate in farming practices that are more conservation-orientated.[33]

In this study, our aim was to develop an understanding of the attitudes and behaviour of ostrich land managers in the Little Karoo, and to identify factors that may positively influence a land manager's decision to conserve their natural environment. This is an important step in determining conservation opportunity, and towards developing implementation strategies for landscape conservation in the Little Karoo.[9]

The three specific aims were:

1. to document land managers' farming practices, environmental attitudes, levels of conservation knowledge, attitudes towards different types of conservation incentives, and willingness to collaborate with environmental organisations;

2. to explore the correlation and causal relationships between environmental attitude, conservation knowledge, conservation behaviour and willingness to collaborate; and

3. to identify attributes of farmers who display higher than average levels of conservation behaviour, and who are open to collaboration with organisations promoting conservation and sustainable agriculture.

Methods

Study area

The study was conducted in the Oudtshoorn Basin, a 10 163 km² area in the northeast of the Little Karoo which includes Oudtshoorn, the largest town. The Little Karoo is a semi-arid, inter-montane basin where three biodiversity hotspots (Cape Floristic Region, Maputaland-Pondoland-Albany and the Succulent Karoo) intersect.[29] The low-lying parts of the basin are dominated by dwarf shrublands associated with the Succulent Karoo biome.[28] Only 8.6% of undeveloped land remains intact, while 67.9% is moderately degraded and 23.5% severely degraded.[28] Ostrich farming is the largest contributor to the economy of the Little Karoo, and more than 80% of all South African ostrich farms are situated in this region.[31,34] All operating ostrich farms (256 at the time of the study) are registered with the South African Ostrich Business Chamber (SAOBC), the coordinating body for the ostrich industry.

Ostrich farming comprises several different production systems, of which all or a subset may be found on any particular farm in the study area. Breeding practices implemented by ostrich land managers include free-range flock breeding on natural land, intensive group breeding in small camps, and intensive pen breeding. Free-range flock breeding on natural land is the most environmentally destructive ostrich farming practice[31], whereas intensive pen breeding is considered to be a more ecologically and economically sustainable alternative because of its considerably smaller footprint[35].

Several biodiversity initiatives have been launched in the Western Cape following the development of fine-scale maps of Critical Biodiversity Areas. Critical Biodiversity Areas are terrestrial and aquatic features that must be kept in a natural state in order to retain a reasonable proportion of biodiversity pattern in an ecologically functional and resilient landscape and represent the most area-efficient option for meeting all stated biodiversity thresholds. Critical Biodiversity Areas form the basis for conservation initiatives in the Little Karoo.[36] The Ostrich Industry Biodiversity Management Project is a business and biodiversity initiative implemented by the SAOBC and its aim is to investigate and promote alternative ostrich farming practices in the Little Karoo.

Sample selection and data collection

The Oudtshoorn Basin was chosen because it has, since 1865, been the centre of the Little Karoo's ostrich industry.[28] All SAOBC registered ostrich farms within the Oudtshoorn Basin that were affected by the avian influenza outbreak in 2011 were identified. All ostriches on these farms were slaughtered by order of the South African Department of Agriculture, which provided data on ostrich numbers and densities.[33] We assumed that the farms that were affected by avian influenza represented a random sample of ostrich farms in the Oudtshoorn Basin, as viruses exist and flow through complex agri-food systems in a random fashion, making it impossible to predict the direction or speed of their movement through livestock farming regions.[37-39] From the affected ostrich farms, all those that occurred within the Little Karoo Critical Biodiversity Area map's critically endangered vegetation types were selected. Of the 41 affected registered ostrich farms in the Oudtshoorn Basin, 27 farms were situated within a Critical Biodiversity Area. This gave us a sample of 27 farms (10.5% of registered ostrich farms in the Oudtshoorn Basin) operated by 23 land managers, 22 of whom owned their farms.

A structured questionnaire (Supplementary appendix 1) was developed, based on earlier research.[9,22,31] This comprised a combination of quantitative, qualitative, closed- and open-ended questions. The questionnaire was tested in two pilot interviews with land managers not included in the sample. The questionnaire was developed in English and translated into Afrikaans, as most land managers in the Little Karoo are Afrikaans speaking. Face-to-face interviews were conducted with the 23 land managers on their farms between May and November 2012. Land managers were informed that interviews would be confidential, that they would remain anonymous, and that participation was voluntary. All land managers approached agreed to be interviewed.

We identified four dimensions hypothesised to comprise a land manager's conservation character[9,22], namely environmental attitudes (EA), conservation knowledge (CK), conservation behaviour (CB) and land manager willingness to collaborate (WC). The sections of the questionnaire that targeted EA and CB comprised 13 and 9 statements, respectively, with agreement or disagreement being recorded on a 5-point Likert scale. The section targeting CK assessed whether land managers knew about, or were aware of, biodiversity and conservation facts pertinent to the study area and to ostrich farming. Answers were scored using a scale consisting of 3-point Likert items (yes, unsure or no). The section targeting WC contained three subsections: (1) willingness to work with each of 24 land-management organisations (measured using a 5-point scale); (2) the degree of willingness to engage in conservation (agreement on a 5-point scale with seven statements regarding the landowner's role in conservation); and (3) the degree of interest in nine types of incentives for participating in conservation measures (measured using a 5-point scale).

The questionnaire was grouped into eight sections. The first section gathered background information on the land managers, including the number of years the land manager had been on the farm and the size of the farm. The next four sections of the questionnaire targeted the dimensions/constructs EA, CK, CB and WC, as described above. The sixth section comprised a range of closed- and open-ended questions about ostrich farming practices, including their spatial and temporal scale, and why land managers chose specific practices and avoided others. The seventh section comprised a range of closed- and open-ended questions about ostrich stocking rates, including land managers' flock breeding preferences, ostrich stocking rates on natural land, estimates of natural land stocking capacity, the perceived condition of their natural land, and their assessment of the profitability of the stocking rate of 22.8 ha/ostrich recommended by the Provincial Department of Agriculture. The final section of the questionnaire gathered demographic information, including age and level of education.

Data analysis

To refine, or sharpen, the scales that we used to measure the latent variables (LVs) of our study (i.e. the dimensions CB, CK, EA and WC), we relied mainly on classical test theory. Conceptually and mathematically, classical test theory is simpler than item response theory, and it is also better suited to small samples.[40] We assessed the internal consistency and dimensionality of our scales using Cronbach's alpha[41] (summarised by α), McDonald's hierarchical factor analysis[42] (summarised by the omega hierarchical coefficient ωh), and Revelle's item-cluster analysis[43] (summarised by the beta coefficient β). In addition to these classical test theory tools, we used Mokken scale analysis[44] and item-information curves from item response theory to help identify the most internally consistent and unidimensional set of items. The final scales were formed by eliminating items that reduced the measures of internal consistency and/or dimensionality to unacceptably poor values[9,43] (Table 1).

We explored relationships between the LVs and likely predictor variables (Supplementary table 1) using multivariate regression trees (MRT).[45] The method allows for the identification of groups of land managers who hold similar attitudes to conservation and whose conservation-related behaviour is similar, based on easy-to-obtain general characteristics such as the size of the farm and the number of years spent farming. The decision trees (or rules) that result from the analysis allow one to recast the results as a principal component analysis of instrumental variables, or redundancy analysis[46], where the instrumental variable is given by the MRT-determined combinations of the predictor variables[45]. The constrained principal components are linear combinations of the column-centred LVs (technically, response variables). The MRT analysis provides an F-test of the significance of the effect of the MRT-transformed predictors (or rules) on the transformed LVs, i.e. the constrained principal components. We conducted the test by regressing each constrained principal component on a three-level factor whose levels are given by the MRT-determined rules.

We also mined out a set of causal models relating the LVs to each other, using the PC algorithm of Spirtes et al.[47] To test the plausibility of the models, we used Shipley's C-test[48], and followed up by fitting a pair of partial least-squares path models to the best-fitting model to test model effects.

Analyses were done using R[49], supported by the contributed packages bnlearn[50] (for the causal model), semPLS[51] (for the partial least-squares path model), mvpart[45] (for the MRT and associated analyses), and multcomp[52] (for the post-hoc contrast analysis of the constrained principal components).

Table 1: Statistics on the reliability and dimensionality of the final scales to measure the four latent variables

Scale	RV[a]	Reliability[b]					Dimensionality[c]			
		α	λ6	ωt	MS	LC$_R$	H	β	ECV	ωh
Conservation behaviour (CB)	0.913	0.881	1.000	0.990	0.841	0.777	0.593	0.814	0.502	0.704
Conservation knowledge (CK)	0.806	0:849	1.000	0.999	0.655	0.763	0.553	0.631	0.563	0.710
Environmental attitude (EA)	0.838	0.905	0.979	0.991	0.904	0.835	0.653	0.785	0.655	0.839
Willingness to collaborate (WC) [C19]	0.846	0.945	1.000	0.984	0.926	0.900	0.671	0.841	0.632	0.783

[a]RV indexes the extent to which the matrix of scores of the sharpened scale matches or approximates that of the unsharpened scale.

[b]Reliability if the item is dropped. The first three indices derive from classical test theory, the last two from Mokken scale analysis (non-parametric item response theory): α is Cronbach's alpha; λ6 is the sixth of Guttman's coefficients; ωt is McDonald's total omega; MS is Molenaar and Sitjsma's coefficient of reliability; LC$_R$ is coefficient of latent class reliability.

[c]Dimensionality if the item is dropped. H is Loevinger's coefficient of scale scalability (from Mokken scale analysis); β is Revelle's beta (from item-cluster analysis); ECV is Reise's explained common variance; ωh is McDonald's hierarchical omega.

Results

Land managers' attitudes, knowledge and behaviour

All 26 land managers were white men aged between 33 and 77 years, with the majority (61%) aged ≤50 years (Supplementary table 2). Farming experience ranged from less than 10 years to over 50 years, with the majority (74%) of respondents having farmed for between 11 years and 30 years. Farms ranged in size from <50 ha to 12 000 ha, with 48% being between 100 ha and 500 ha in size. Most (91%) land managers indicated that ostrich farming was their primary land-use activity. First-generation land managers comprised 22% of the sample, with the remainder representing a range from 2nd generation to 7th generation land managers; 65% of land managers had learnt how to farm ostriches from their fathers. Although 17% had no tertiary qualifications, 35% had a formal qualification in Agriculture.

Flock breeding was the most widespread breeding practice, implemented by 65% of land managers, primarily because it is less intensive and easier to manage than pen breeding, and was considered the most profitable practice. Only two (9%) respondents reported that they implemented flock breeding primarily for historical reasons, and because they were not interested in other practices. Other reasons included poor success with pen breeding and having access to sufficient natural land available for ostriches to roam. Camp size and the perceived carrying capacity of the land were the primary criteria (both 26%) given for deciding upon ostrich stocking densities.

Regarding ostrich stocking rates, two farmers (9%) thought that ostriches should not be kept on natural land because they cause excessive degradation. Two (9%) thought that low (11–50 ha/ostrich) stocking rates are ecologically appropriate, whereas one (4%) thought that moderate (11 and 20 ha/ostrich) stocking rates are ecologically appropriate. The remaining land managers either expressed no opinion (22%) or thought that a stocking rate of less than 10 ha/ostrich is sustainable (56%). The majority of land managers (87%) held that the recommended stocking rate of 22.8 ha/ostrich is 'very unprofitable'. The farms of 21 of the 23 land managers surveyed have flat areas or valleys where most ostrich farming activities take place. The vegetation of these areas was considered to be in good condition by all 21 land managers who use them.

Attitudes toward the environment (EA) were generally positive (Supplementary table 3). The majority of land managers 'agreed' or 'strongly agreed' (hereafter 'agreed') with several statements indicating pro-environmental attitudes, a sense of responsibility for the environment, and an interest in biodiversity stewardship. Most (96%) agreed with the statement that 'Protection of plants and animals that occur outside of protected reserves should be the responsibility of private landowners'.

Conservation knowledge was generally high (Supplementary table 4). Most respondents (74%) were aware that the vegetation of the Little Karoo is endangered and knew why the lower lying areas of the Little Karoo should be conserved. All but two were aware of the Ostrich Biodiversity Management Project, and of the stocking rates prescribed by the government, but more than half were unaware of the CapeNature Stewardship Programme.

Most land managers reported positive conservation behaviour (Supplementary table 5); 87% had instituted soil conservation measures for reducing soil erosion in the last 5 years, and 70% indicated that they had undertaken nature conservation activities such as surveys, restoration, and monitoring on their farms during that period. More than half (65%) thought it was necessary to have an environmental management plan for their farm, and most (91%) implemented healthy waste-management practices on their farms. Only 34% formally monitored the condition of their land, and 48% regularly attended conservation-related meetings.

When asked whether incentives would motivate them to promote conservation on their land, 95% of respondents agreed (Supplementary table 6). The majority of land managers indicated an interest in incentives (Supplementary table 7). A large proportion (48%) were not interested in public or community recognition for their conservation efforts. Those who responded positively to the incentives question were most interested in subsidy for conservation work (92%), assistance with fencing and land management (87%), assistance with farm environmental plans and maps (83%), law enforcement support (83%), access to scientific information and support (83%), and tax or rate rebates for conservation activities (78%).

Of the 24 organisations suggested in the questionnaire, land managers were most willing to collaborate with the National Department of Agriculture, Fisheries and Forestry (NDAFF), the South African Police Service (SAPS), a Farmers' Association (AgriKK), and neighbouring farmers (Supplementary table 8). The Wildlife and Environmental Society of South Africa (WESSA), Rhodes University, and local and district municipalities received the lowest willingness to collaborate scores. Many land managers had never heard of the conservation organisations WESSA (43%), the Gouritz Biosphere Reserve (22%), or the Cape Leopard Trust (22%).

Correlation and causal relationships between latent and ancillary variables

Relationships among the four latent variables (CB, CK, EA and WC), and between them and four key ancillary variables (Years Farming (YrF), Farm Size (FSz), Farmer Age (FAge), and Level of Education (LoE)), are shown in Supplementary figure 1. Conservation behaviour was positively correlated with EA ($p<0.001$) and CK ($p<0.01$), but only marginally with WC ($p=0.1$). A near-significant positive relationship ($p<0.1$) was found between CK and EA, but no significant relationships were found between CK and WC, or between CB and WC. Farmer age and YrF showed a strong positive correlation ($p<0.001$). There were near-significant ($p<0.1$) positive correlations between FAge and CB and CK (but not EA), and a near-significant negative correlation between YrF and WC ($p<0.1$). Farm size showed a weak positive relationship with CK ($p<0.1$). Level of Education was not significantly correlated with any other variable.

Both causal models (Figure 1) are plausible descriptions of causal relationships between the latent variables, based on Shipley's C-test ($C_{df=6}=2.947$, $p=0.815$ and $C_{df=8}=7.756$, $p=0.458$). The model shown in Figure 1a fits the observed data marginally better than does the model shown in Figure 1b, based on the log-likelihood (-90.525 versus -90.539). The two models differ in how they describe the causal relationship between CB and EA, with CB being a determinant of EA in the model shown in Figure 1a and the opposite being the case in the other model. Relationships that are common to both models are, firstly, that CK has no causal effect on CB, EA or WC. Secondly, and likewise, WC has no effect on CB, CK or EA. Finally, both models indicate that CK is a consequence of CB, rather than it influencing, or being a cause of, CB.
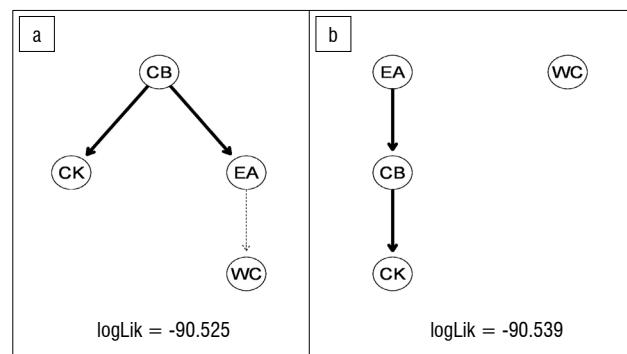

Figure 1: Causal models of the latent variables. The weights of the paths are proportional to the degree of confidence in the path, based on a Monte-Carlo simulation analysis (800 000 permutations). All effects are positive, and both models are plausible descriptions of causal influences based on Shipley's C-test (see the text).
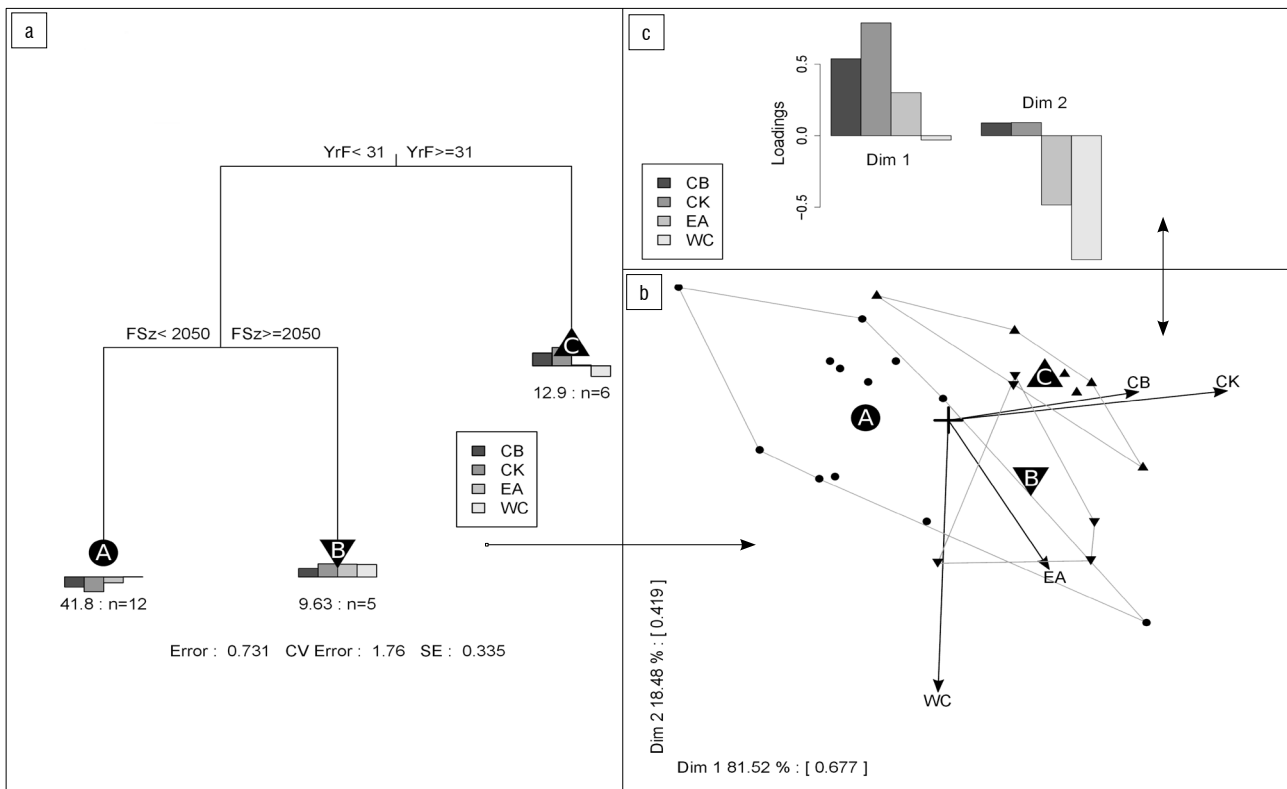
87

Figure 2: (a) Multivariate regression tree (MRT) of the predictors of the latent variables (LVs). The model has an approximate R^2 of 0.2698, meaning it explains approximately 26.98% of the variance of the LVs. Bars show deviations from the grand mean of each scale. (b) Distance (i.e. row-preserving) biplot of the constrained principal component analysis on instrumental variables (PCAIV). The origin of the biplot represents the average position. Convex hulls delimit each group and the big points within each hull show the centroid (or MRT-predicted mean) of each group. An estimate of respondents' individual-level scores on the LVs may be derived by orthogonally projecting the small points onto the vectors of the LVs. The values in square brackets after the variance explained by each dimension are interset correlations, the correlation between respondents' scores on each dimension and the LVs. (c) Variable loadings on dimensions 1 and 2 of the PCAIV shown in (b).

Predicting conservation opportunity from farmer attributes

A MRT showed that there were three groups of land managers based on their scores for the four LVs (Figure 2a). The three groups were defined by the number of years they had been farming (YrF) and by farm size (FSz). Land managers with 30 or fewer years of farming experience were split into two groups. The first group (Group A in Figure 2) of 12 land managers manage smaller farms (<2050 ha). They have below-average scores for EA, AB and CK, and have average scores for WC. The second group (Group B) comprised five land managers whose farms are large (≥2050 ha) and who have above average scores for all four LVs. Scores for CB and CK were highest in land managers ($n=6$) who had farmed for ≥31 years (Group C in Figure 2). Their EA scores were average and their WC scores were the lowest of the three groups. This group represents more experienced and knowledgeable land managers who show little willingness to collaborate with a wide range of institutions. Parametric and non-parametric omnibus tests reject the null hypothesis that there are no differences between the grand means of the three groups (parametric test: ρ=0.0061; non-parametric test: ρ=0.0045), although the effect size was relatively small (26.98% variance explained).

Figure 2b illustrates relationships between the MRT-determined groups and the LVs in greater detail, on a component-by-component basis. The first dimension explained 81.52% of the total variation and is highly significant ($F_{[df=2,20]}=8.48$, $p=0.0022$). It is mainly a contrast between the two groups of respondents on the right-hand side of the origin (Groups B and C), characterised by higher than average scores for CK, CB and EA (in order of decreasing weight or importance), and the group on the left-hand side of the origin (Group A), with lower than average scores for these LVs. Willingness to collaborate (WC) projects close to the origin on the first dimension and therefore makes virtually no contribution to

the separation of groups in that dimension (variable loadings on the two dimensions are shown in Figure 2c). Post-hoc analysis of the dimension shows that Groups B and C differed significantly from Group A but did not differ significantly from each other (Table 2).

Table 2: Post-hoc pairwise comparisons (Tukey contrasts) showing the estimated difference (linear contrast) between the three groups of land managers along the first and second dimensions of the constrained principal component analysis (Figure 2b). The null hypothesis tested is that there is no difference between group-scores; p-values were adjusted for multiple testing using a single-step method based on the t-distribution.

Comparison	Dimension 1			Dimension 2		
	Estimate	*t*-value	*p*-value	Estimate	*t*-value	*p*-value
A − B = 0	-1.940	-3.56	0.0053	-0.538	-1.04	0.5598
C − B = 0	-0.151	-0.23	0.9715	-1.293	-2.06	0.1229
C − A = 0	1.790	3.08	0.0153	-0.755	-1.37	0.3737

The second dimension of the analysis explains 18.48% of the total variation. Figure 2b shows a minor separation between Groups B and C in the second dimension, because of different scores on WC, and to some extent on EA (see the bars for WC and EA of Groups B and C in Figure 2a). The second dimension is not, however, statistically significant ($F_{[df=2,20]}=2.13$, $p=0.15$). Given this fact, and the fact that almost all of the variability in WC lies in the non-significant second dimension, we may conclude that the three groups of respondents do not differ from

each other to any significant degree in their scores for WC. This is borne out by a post-hoc analysis of the dimension (Table 2).

Figure 2b also illustrates the extent to which respondents in each group diverge from the predicted (or mean) score of the group (marked by the large symbol at the centroid of the convex hull enclosing each group). It illustrates the extent to which respondents do not fit the pattern of LV-scores predicted by the MRT model, i.e. by the social identifiers 'years spent farming' and 'farm size'. If we exclude the unusual member of Group A in the bottom right-hand corner of the biplot (quadrant IV), with the highest, or close to highest, scores on all four LVs, and two members of Group B, who marginally overlap with Group C, then all three groups are distinct with respect to their scores on the LVs. Despite the relatively small size of the sample, the MRT-determined characteristics of 'years spent farming' and 'farm size' are thus good predictors of respondents' scores and therefore of their attitudes and behaviour.

Discussion

Land managers' current practices and attitude to change

Despite research efforts there is still much to be learned about why farmers voluntarily adopt, or fail to adopt, improved technologies and sustainable practices.[11] The majority of land managers interviewed maintained a traditional flock breeding system on natural veld, even though this practice has been reported to lead to land degradation.[31] South Africa is one of the few countries (along with Israel and Zimbabwe) where traditional flock breeding of ostriches in large populations is widely practised.[53] More economically and ecologically sustainable alternatives to flock breeding have been recommended by agricultural economists and by the Ostrich Industry Biodiversity Management Project. Intensive pen breeding in particular is supported by the SAOBC, because its impact is limited to much smaller spatial scales than that of extensive ostrich flock breeding.[31,35] This raises the question of why land managers still overwhelmingly implement flock breeding in the Little Karoo, and elsewhere in South Africa. The reasons for flock breeding in our study area, and likely other parts of the Little Karoo, appear to be a combination of a preference for established low-intensity practices, and positive (although seemingly unrealistic) perceptions of the economic and ecological sustainability of current practices.

The stocking rate preferred by most land managers was 10 ha/ostrich. This stocking rate is more than double the recommended agricultural stocking rate of 22.8 ha/ostrich that land managers considered unprofitable. There is a wide gap between these rates considered to be economically and ecologically sustainable, and most studies on sustainable ostrich stocking rates have only addressed a single dimension of sustainability.[32] The lack of agreement on an ecologically and economically sustainable stocking rate is likely to be an important factor hindering the adoption of alternative, more sustainable farming practices in the Little Karoo.

Attitudes, knowledge, behaviour and willingness to collaborate

On the whole, land managers reported positive attitudes towards the natural environment of the Little Karoo, a good knowledge of the global importance of its biodiversity and conservation, and conservation behaviours such as soil conservation, ecological monitoring, and environmental management. This corresponds to the farmers' own assessments as operating in an environmentally sustainable manner, but contrasts with ecologists' views of the actual impacts of extensive ostrich farming.

A key aim of this study was to assess whether positive environmental attitudes and high levels of conservation knowledge corresponded to high levels of conservation behaviour (actual practices) and a willingness to collaborate (the potential to influence and support the improvement of practices). Both CK and EA were moderately and significantly correlated with conservation behaviour, but were more weakly and insignificantly correlated with each other. Inferences that can be drawn from the two causal models are, firstly, that little would be gained from investing in improving CK as a way of improving CB, EA or WC, because CK has no causal effect on them. Secondly, and likewise, there is no point encouraging a willingness to collaborate (WC), because WC was found

to have no effect on CB, CK or EA. Finally, both models indicate that CK is a consequence of CB, rather than CK influencing or being a cause of CB.

The finding that CK does not lead to an increase in conservation-related behaviour (or to a greater willingness to collaborate) is echoed by the findings of other recent studies of environmental behaviour among farmers. Environmental awareness was not sufficient to motivate the adoption of water conservation in Finland[16], and information from the media did not significantly influence farmers' intentions to farm sustainably in Ethiopia[11]. The theory of planned behaviour[5] describes behaviour as being driven by intention (itself a function of attitude, subjective/societal norm, and perceived behaviour control), and by perceived behaviour control directly. Knowledge is not included in this model as a direct determinant of behaviour, although some modifications[1] include it as an indirect driver in the form of awareness of the need for, and consequences of, the behaviour. Several recent studies have applied the theory of planned behaviour and variants thereof to understand the drivers of farmer behaviour, including the adoption of technologies and sustainable practices.[11] A common finding, similar to that in this study, is that farmers often have positive attitudes towards the environment but do not act on them because of a perceived inability to change, with economic constraints being a commonly cited factor.[18]

Willingness to collaborate was found to be distinct from the other scales and to be uncorrelated with EA, CB and CK. This finding suggests that farmers who had positive attitudes and behaviours were not necessarily willing to collaborate with a variety of agencies and institutions to achieve greater ecological sustainability. It is probable that respondents' stated willingness to collaborate with particular organisations reflected their trust/distrust, or pragmatism and discernment with regard to particular organisations, rather than a fundamental (un)willingness to collaborate with organisations to achieve better land management and conservation. As it was implemented in this study, the WC scale measures at least three things: (1) a willingness to collaborate *sensu stricto* (i.e. in the sense of a latent trait), (2) a willingness to collaborate conditional on the nature of the organisation (i.e. additional to the latent trait, but tempered by discrimination/judgement), and (3) a general knowledge of organisations that engage in conservation or in conservation-related activities. Future work to identify land managers who are open to engagement on conservation issues, and who potentially would be willing to adapt their management practices, should be designed to disentangle these different, interrelated, dimensions of willingness to collaborate.

In addition to individual attitudes, norms and values, the most commonly applied theories in environmental psychology recognise two other important determinants of environmental behaviour.[5,54] These determinants are societal or subjective norms (what is expected by society, especially peers, and how strong such pressure is), and perceived behaviour control (an individual's perceived capability to perform a behaviour). Initiatives encouraging the adoption of conservation practices should therefore target entire farming communities, in an effort to change societal norms in tandem with personal ones[4], and to address perceived constraints to behaviour change. Farmers' perceived ability to change behaviour, even when they recognise the value of nature conservation, is very often constrained by economic factors.[18,55] Incentives such as tax rebates and other material support may help to reduce this obstacle to adopting more sustainable farming practices, as indicated by farmers in this study. Lastly, frequent behaviours become more strongly influenced by habit and less determined by intent.[1] Influencing routine behaviours thus requires not only a change in personal and societal attitudes but also a disruption of established habits.

Predicting conservation opportunity based on farmers' attributes

We expected that older land managers with more years of farming experience would be less likely to change their traditional farming practices and therefore would be less willing to innovate. We also expected that individuals who own or manage larger farms would be in a better position to implement conservation actions, given that such farmers were likely to have more financial and other resources, as well as being better placed to benefit from economies of scale if changing

their practices. Farm size and years of farming did, indeed, emerge as the key predictors of an individual's EA, CK, CB and WC profile and could be used in conjunction with other information to identify land managers who present an opportunity for championing more ecologically sound ostrich farming practices. Age and farm size also emerged as factors correlated with landholder attitudes and behaviour toward the conservation of Renosterveld, with younger farmers, and those with bigger farms and more remnant Renosterveld vegetation, being more inclined towards conservation.[22] Cultivated acreage size (together with attitudes and past behaviour) was also found to be a significant factor in influencing an intention to adopt sustainable practices in Italy.[13]

Land managers with 30 or fewer years of experience and larger properties (≥2050 ha) had above-average mean scores for all four LVs. In terms of actual practices, attitudes and willingness to collaborate, this group (Group B in Figure 2) presents the greatest opportunity for intervention to improve ecological sustainability, as well as for championing newer and more sustainable farming methods. Conservation behaviour and CK were highest in respondents who had farmed for ≥31 years (Group C in Figure 2). Land managers in this older group of land managers had average EA scores and, as expected, were the least willing to collaborate. Given their long experience (and hence likely good standing among their peers) and high CB scores, this group of land managers is important to engage. Our findings suggest that engaging older farmers would require a different approach than that used for working with the younger farmers with larger properties (Group B), whose willingness to collaborate was already high. It is important to determine whether the low CK scores in this group of older farmers is because of a fundamental conservatism and unwillingness to collaborate, and/or their perception of the particular set of organisations they were presented with in the questionnaire.

The largest group in our sample comprised land managers with more limited farming experience (<31 years) and smaller farms (<2050 ha; Group A in Figure 2). Given their low scores for all latent variables, this group should not be a priority for an investment of time and effort in an attempt to change attitudes or behaviours. Although their lack of experience could mean that with more input they could become more aware and involved, land managers with small farms may have a more limited opportunity for change as they are unable to compete without the economies of scale available to larger farms.[33] There is also potentially less natural land to conserve on smaller farms, meaning that the return-on-investment of partnerships built to initiate positive changes of attitude and behaviour is likely to be relatively small.

Conclusions

Despite positive attitudes to biodiversity and conservation, the majority of ostrich farmers in the Little Karoo practise flock breeding, which is deemed environmentally unsustainable. Farmers themselves believed that flock breeding at stocking rates far exceeding the recommended carrying capacity was economically viable, and that land used in this way was in good ecological condition. The big difference between the assessments and recommendations of conservationists, ecologists and the majority of farmers points to a need to close this gap.[32] This requires communication and co-operation, and also perhaps a more tempered and realistic assessment of what ecological sustainability in a production (as opposed to a conservation) landscape should look like. The factors that constrain farmers from adopting new methods and finding effective and appropriate ways to make alternative methods more viable for land managers, should be research priorities. Agricultural organisations that are trusted by farmers have a key role to play in bridging the gap between farmer and conservation perceptions. At a time when shrinking budgets reduce the availability of extension services and other face-to-face farmer support, social learning among farmers should be fostered and supported.[56]

Acknowledgements

Authors' contributions

A.W., A.T.K. and S.V. conceptualised and designed the research. A.W. collected the data. M.D. performed data analyses. A.W. and S.V. wrote the manuscript with substantial editorial contributions from A.T.K. and M.D.

References

1. Klöckner CA. A comprehensive model of the psychology of environmental behaviour – A meta-analysis. Glob Environ Chang. 2013;23:1028–1038. https://doi.org/10.1016/j.gloenvcha.2013.05.014

2. Tremblay KR Jr, Dunlap RE. Rural-urban residence and concern with environmental quality: A replication and extension. Rural Sociol. 1978;43:475–491.

3. Williams JA Jr, Moore HA. The rural-urban continuum and environmental concerns. Great Plains Res. 1991;1:195–214.

4. Quinn CE, Burbach ME. A test of personal characteristics that influence farmers' pro-environmental behaviours. Great Plains Res. 2010;20:193–204.

5. Ajzen I. The theory of planned behaviour. Organ Behav Hum Dec. 1991;50:179–211. https://doi.org/10.1016/0749-5978(91)90020-T

6. Kaiser FG, Wölfing S, Fuhler U. Environmental attitude and ecological behaviour. J Environ Psychol. 1999;19:1–19. https://doi.org/10.1006/jevp.1998.0107

7. Milfont TL, Duckitt J. The Environmental Attitudes Inventory: A valid and reliable measure to assess the structure of environmental attitude. J Environ Psychol. 2010;30:80–94. https://doi.org/10.1016/j.jenvp.2009.09.001

8. Curran P, Smedley D, Thompson P, Knight AT. Mapping restoration opportunity for collaborating with land managers in a carbon-credit-funded restoration program in the Makana Municipality, Eastern Cape, South Africa. Restor Ecol. 2012;20:56–64. https://doi.org/10.1111/j.1526-100X.2010.00746.x

9. Knight AT, Cowling RM, Difford M, Campbell BM. Mapping human and social dimensions of conservation opportunity for the scheduling of conservation action on private land. Conserv Biol. 2010;24:1348–1358. https://doi.org/10.1111/j.1523-1739.2010.01494.x

10. Reimer AP, Thompson AW, Prokopy LS. The multi-dimensional nature of environmental attitudes among farmers in Indiana: Implications for conservation adoption. Agric Hum Val. 2012;29:29–40. https://doi.org/10.1007/s10460-011-9308-z

11. Zeweld W, Van Huylenbroeck G, Tesfay G, Speelman S. Smallholder farmers' behavioural intentions towards sustainable agricultural practices. J Environ Manage. 2017;187:71–81. https://doi.org/10.1016/j.jenvman.2016.11.014

12. Martínez-García CG, Dorward P, Rehman T. Factors influencing adoption of improved grassland management by small-scale dairy farmers in central Mexico and the implications for future research on smallholder adoption in developing countries. Livest Sci. 2013;152:228–238. https://doi.org/10.1016/j.livsci.2012.10.007

13. Menozzi D, Fioravenzi M, Donati M. Farmer's motivation to adopt sustainable agricultural practices. Bio-Based Appl Econ. 2015;4:125–147.

14. Borges JAR, Tauer LW, Lansink AGJMO. Using the theory of planned behavior to identify key beliefs underlying Brazilian cattle farmers' intention to use improved natural grassland: A MIMIC modelling approach. Land Use Policy. 2016;55:193–203. https://doi.org/10.1016/j.landusepol.2016.04.004

15. Fielding KS, Terry DJ, Masser BM, Bordia P, Hogg MA. Explaining landholders' decisions about riparian zone management: The role of behavioural, normative, and control beliefs. J Environ Manage. 2005;77:12–21. https://doi.org/10.1016/j.jenvman.2005.03.002

16. Grammatikopoulou I, Pouta E, Myyrä S. Exploring the determinants for adopting water conservation measures. What is the tendency of landowners when the resource is already at risk? J Environ Plann Manage. 2016;59:993–1014. https://doi.org/10.1080/09640568.2015.1050551

17. Power EF, Kelly DL, Stout JC. Impacts of organic and conventional dairy farmer attitude, behaviour and knowledge on farm biodiversity in Ireland. J Nat Cons. 2013;21:272–278. https://doi.org/10.1016/j.jnc.2013.02.002

18. Šorgo A, Špur N, Škornik S. Public attitudes and opinions as dimensions of efficient management with extensive meadows in Natura 2000 area. J Environ Manage. 2016;183:637–646. https://doi.org/10.1016/j.jenvman.2016.09.024

19. Cumming T, Driver A, Pillay P, Martindale G, Purnell K, McCann K, et al. The business case for biodiversity stewardship. A report produced for the Department of Environmental Affairs. Pretoria: South African National Biodiversity Institute; 2015.

20. McDowell C. Factors affecting the conservation of Renosterveld by private landholders [doctoral thesis]. Cape Town: University of Cape Town; 1988.

21. McDowell C, Sparks R. The multivariate modelling and prediction of farmers' conservation behaviour toward natural ecosystems. J Environ Manage. 1989;28:185–210.

22. Winter SJ, Prozesky H, Esler KJ. A case study of landholder attitudes and behavior toward the conservation of Renosterveld, a critically endangered vegetation type in the Cape Floral Kingdom, South Africa. J Environ Manage. 2007;40:46–61. https://doi.org/10.1007/s00267-006-0086-0

23. Pasquini L, Cowling RM, Twyman C, Wainwright J. Devising appropriate policies and instruments in support of private conservation areas: Lessons learned from the Klein Karoo, South Africa. Conserv Biol. 2009;24:470–478. https://doi.org/10.1111/j.1523-1739.2009.01344.x

24. McDowell C. Legal strategies to optimize conservation of natural ecosystems by private landholders: Restrictive legislation. Comp Int Law S Afr. 1986;19:450–460.

25. McDowell C. Legal strategies to optimize conservation of natural ecosystems by private landholders: Economic incentives. Comp Int Law S Afr. 1986;19:461–474.

26. Savy C. Conserving wattled crane (Grus canrunclulatus) on private land in KwaZulu-Natal, South Africa: Attitudes, economics and potential incentives [MSc thesis]. Cape Town: University of Cape Town; 2003.

27. Mittermeier RA, Gil PR, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, et al. Hotspots revisited: Earth's biologically richest and most endangered eco-regions. Chicago, IL: University of Chicago Press; 2005.

28. Thompson MW, Vlok JHJ, Rouget M, Hoffman MT, Balmford A, Cowling RM. Mapping grazing-induced degradation in a semi-arid environment: A rapid and cost-effective approach for assessment and monitoring. Environ Manage. 2008;43:585–596. https://doi.org/10.1007/s00267-008-9228-x

29. Vlok JHJ, Cowling RM, Wolf T. A vegetation map for the Little Karoo. Unpublished maps and report for the SKEP project. 2005.

30. Herling MC, Cupido CF, O'Farrell PJ, Du Plessis L. The financial costs of ecologically non-sustainable farming practices in a semi-arid system. Restor Ecol. 2009;17:827–836. https://doi.org/10.1111/j.1526-100X.2008.00419.x

31. Cupido CF. Assessment of veld utilization practices and veld condition in the Little Karoo [MSc (Agric) thesis]. Stellenbosch: Stellenbosch University; 2005.

32. Wheeler A, Knight AT, Vetter S. Examining the evidence for ecologically sustainable ostrich breeding practices on natural veld in the Little Karoo, South Africa. Afr J Range Forage Sci. 2015;32:233–241. https://doi.org/10.2989/10220119.2014.982708

33. Wheeler A. Factors influencing ecological sustainability in the ostrich industry in the Little Karoo, South Africa [MSc thesis]. Grahamstown: Rhodes University; 2014.

34. Le Maitre DC, Milton SJ, Jarmain C, Colvin CA, Saayman I, Vlok JHJ. Linking ecosystem services and water resources: Landscape-scale hydrology of the Little Karoo. Front Ecol Environ. 2007;5:261–270. https://doi.org/10.1890/1540-9295(2007)5[261:LESAWR]2.0.CO;2

35. Murray MI. Economic case for pen-breeding. Unpublished report for the South African Ostrich Business Chamber Biodiversity Management Project supported by Critical Ecosystems Protection Fund (CEPF). 2008.

36. Pasquini L. Assessing the suitability and feasibility of implementing a biosphere reserve in the Gouritz Initiative Domain. Internal report for the Gouritz Initiative. 2008.

37. Bickerstaff K, Simmons P. The right tool for the job? Modelling, spatial relationships, and styles of scientific practice in the UK foot and mouth crisis. Environ Plann D. 2004;22:393–412. https://doi.org/10.1068/d344t

38. Hinchliffe S. Indeterminacy in-decisions – Science, policy and politics in the BSE (Bovine Spongiform Encephalopathy) crisis. Trans Inst Brit Geog. 2001;26:182–204. https://doi.org/10.1111/1475-5661.00014

39. Law J. Disaster in agriculture: On foot and mouth mobilities. Environ Plann A. 2006;38:227–239. https://doi.org/10.1068/a37273

40. De Champlain AF. A primer on classical test theory and item response theory for assessments in medical education. Med Educ. 2010;44:109–117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

41. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334. https://doi.org/10.1007/BF02310555

42. McDonald RP. Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum Associates; 1999.

43. Revelle W. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behav Res. 1979;14:57–74. https://doi.org/10.1207/s15327906mbr1401_4

44. Mokken RJ. A theory and procedure of scale analysis. Berlin: De Gruyter; 1971. https://doi.org/10.1515/9783110813203

45. De'ath G. Multivariate regression trees: A new technique for modeling species-environment relationships. Ecology. 2002;83:1105–1117.

46. Rao CR. The use and interpretation of principal component analysis in applied research. Sankhya A. 1964;26(4):329–358.

47. Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. 2nd ed. Cambridge, MA: MIT Press; 2000.

48. Shipley B. A new inferential test for path models based on directed acyclic graphs. Struct Equ Modeling. 2000;7(2):206–218. https://doi.org/10.1207/S15328007SEM0702_4

49. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016. Available from: http://www.R-project.org/

50. Scutari M, Denis J-B. Bayesian networks with examples in R. New York: Chapman & Hall/CRC; 2014. https://doi.org/10.1201/b17065

51. Monecke A, Leisch F. semPLS: Structural equation modelling using partial least squares. J Stat Softw. 2012;48(3):32.

52. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biometrical J. 2008;50:346–363. https://doi.org/10.1002/bimj.200810425

53. Deeming DC, Burbier NE. Behaviour in natural and captive environments. In: Deeming DC, editor. The ostrich: Biology, production and health. Wallingford: CABI Publishing; 1999. p. 83–104.

54. Jackson T. Motivating sustainable consumption: A review of evidence on consumer behaviour and behavioural change. Guildford: Centre for Environmental Strategy, University of Surrey; 2005.

55. Vanclay F, Lawrence G. Farmer rationality and the adoption of environmentally sound practices; a critique of the assumptions of traditional agricultural extension. Eur J Agric Educ Extension. 1994;1:59–90. https://doi.org/10.1080/13892249485300061

56. Genius M, Koundouri P, Nauges C, Tzouvelekas V. Information transmission in irrigation technology adoption and diffusion: Social learning, extension services, and spatial effects. Am J Agric Econ. 2013; 96(1):328–344.

AUTHORS:
Douw G. Breed[1] 🆔
Tanja Verster[1] 🆔

AFFILIATION:
[1]Centre for Business Mathematics and Informatics, North-West University, Potchefstroom, South Africa

CORRESPONDENCE TO:
Tanja Verster

EMAIL:
Tanja.Verster@nwu.ac.za

An empirical investigation of alternative semi-supervised segmentation methodologies

Segmentation of data for the purpose of enhancing predictive modelling is a well-established practice in the banking industry. Unsupervised and supervised approaches are the two main types of segmentation and examples of improved performance of *predictive* models exist for both approaches. However, both focus on a single aspect – either target separation or independent variable distribution – and combining them may deliver better results. This combination approach is called semi-supervised segmentation. Our objective was to explore four new semi-supervised segmentation techniques that may offer alternative strengths. We applied these techniques to six data sets from different domains, and compared the model performance achieved. The original semi-supervised segmentation technique was the best for two of the data sets (as measured by the improvement in validation set Gini), but others outperformed for the other four data sets.

Significance:

- We propose four newly developed semi-supervised segmentation techniques that can be used as additional tools for segmenting data before fitting a logistic regression.

- In all comparisons, using semi-supervised segmentation before fitting a logistic regression improved the modelling performance (as measured by the Gini coefficient on the validation data set) compared to using unsegmented logistic regression.

Introduction

The use of segmentation within a predictive modelling context is a well-established practice in credit scoring.[1-3] According to Thomas[3], its goal is to achieve more accurate, robust and transparent predictive models that allow lenders to better serve the segments identified. The origins of segmentation lie in marketing survey analysis, with the first application by Belson[4] when studying the effects of BBC broadcasts in England. (For more information on the history of segmentation refer to Morgan et al.[5]) The only early approach still in broad use today is chi-squared automatic interaction detection[6], which was developed initially by Kass[7].

Predictive modelling refers to the use of statistical methods to construct formulae to estimate a target variable based on various explanatory variables. For this paper the target variables are binary, i.e. there are only two outcomes. The basis for model comparison is 'lift', i.e. the ability of the models to distinguish between the two outcomes compared to a naïve estimate.[17] There are several ways to measure lift, and for this paper the Gini coefficient was chosen.

In this paper, the focus is segmentation when developing predictive models, irrespective of the application – credit risk, marketing, financial risk management, fraud detection, process monitoring, health and medicine, environmental analysis, etc. The results should therefore be of interest to researchers in any scientific or other field in which such models are applied. More specifically, this paper increases the number of available segmentation techniques available by proposing four alternative semi-supervised segmentation techniques.

There are two main types of segmentation – supervised and unsupervised – and the former are favoured in predictive modelling. Supervised techniques are used to identify cases that *act alike*, i.e. where 'independent' predictors display similar predictive patterns relative to a 'dependent' target variable. Separate segments are required to address interactions in which predictive patterns change with the values of other predictors, especially when developing generalised linear models. Interactions are often related to the target's value, and the focus is typically on maximising target separation – or impurity – between segments.[6] The most obvious examples are decision trees derived using recursive partitioning algorithms, which identify homogenous risk groups on the assumption that they will display the greatest interactions. This is not always the case.

By contrast, unsupervised segmentation[8] identifies subjects that *look alike*, i.e. have variables with similar values. It maximises segments' dissimilarities based on a distance function, with no dependent variable (one does not need a target). The most obvious examples are cluster and factor analysis, most commonly used in marketing.

The choice between supervised and unsupervised segmentation depends on the application and requirements of the models developed[9], and many examples of improved model performance exist for both[10]. However, both focus on a single aspect (i.e. act or look alike) and so using them together may deliver better results.

Their combined use considers both target and explanatory variables, and is called semi-supervised segmentation (SSS).[11-13] It has many similarities with semi-supervised clustering[14], supervised clustering[15] and semi-supervised semantic segmentation[16] (more used in image processing). For more detail on the differences and similarities see Breed[11]. In this paper, we explore four newly developed variations of an existing technique to see whether they can provide further benefits.

Six data sets were used from different disciplines, each of which was split into a training and validation set. The five different SSS approaches were applied to each to see which worked best, with models built per segment using logistic regression. A further model was developed on the unsegmented data. Of the five approaches, four are new alternatives and form the main contribution of this paper. They were inspired by an existing technique, semi-

supervised segmentation using k-means clustering and information value (SSSKMIV), which was explored in Breed et al.[13] and described in more detail in a recent PhD thesis[11]. K-means clustering is used to measure the independent variable distribution, and information value for target separation. A 'supervised weight' controls the balance between the two aspects.[13] The algorithm is quite complex and calculation intensive, so alternatives were sought. The four variations are:

Variation 1: We replaced the information value with the chi-squared test statistic and call this technique SSSKMCSQ (semi-supervised segmentation as applied to k-means using chi-squared). The chi-squared calculation has similarities with the Hosmer–Lemeshow statistic, and further information can be found in Hand[6].

Variation 2: We developed a density-based semi-supervised technique using Wong's density-clustering algorithm.[18] We call this the SSSWong technique (semi-supervised segmentation applied to Wong's density clustering methodology).

Variation 3: We developed a semi-supervised technique with segment size equality (SSE). We call this the SSSKMIV$_{SSE}$ technique (semi-supervised segmentation applied to the k-means algorithm using information value as supervised component, with the addition of segment size equality).

Variation 4: These techniques (SSSKMIV, SSSKMCSQ, SSSKMIV$_{SSE}$) have some similarities with the k-means semi-supervised segmentation algorithm, proposed in Peralta et al.[19] which is called LK-Means. This methodology has many similarities to SSS techniques, but also has a number of clear differences.[11] Our fourth variation augments other existing semi-supervised techniques[11] to make its results comparable to the others. It is thus not really new, but an existing supervised technique adapted to be comparable with other SSS techniques.

Semi-supervised techniques

Both unsupervised and supervised segmentation make intuitive sense depending on the application and the requirements of the models developed[9] and many examples exist in which the use of either improved model performance[10]. However, both focus on a single aspect (i.e. either target separation or independent variable distribution) and using them in tandem might deliver better results. Five semi-supervised techniques are described here, four of which are new.

Semi-supervised segmentation: SSKMIV

This approach is explored in Breed et al.[20] and described in more detail in a recent PhD thesis[11] and will be used as the first (original) segmentation method. It is called SSSKMIV, an abbreviation for semi-supervised segmentation using k-means clustering and information values, where k-means is used to assess independent variable distributions, and information values for target separation.

The implementation of this approach is quite complex and calculation intensive.[11,20] Further, the information value formula demands that there be at least one event and non-event each time (to avoid division by zero), and results can be distorted by small numbers. A general rule is that each bin and segment combination must have at least five events and five non-events.

Semi-supervised segmentation: SSKMCSQ (Variation 1)

In this variation the information value is replaced with the chi-squared[6] test statistic for the supervised part, and we call this SSSKMCSQ (semi-supervised segmentation as applied to k-means using chi-squared).

The chi-square statistic is often used as a measure of separation. A good example is chi-squared automatic interaction detection, which is a recursive partitioning algorithm used to construct decision trees.[6] It is used here to compare observed target values for each segment against naïve estimates (i.e. counts per class proportional to those for the population).

Using the chi-squared for the supervised part has two main advantages:

- It is always defined within a segmentation scheme (no division by zero). Our techniques do have the option that a user can set a

minimum number of cases. A popular rule of thumb is to have at least 5% of cases of the sample in each segment.[2]

- It works for both binary and continuous variables – which allows its application to a broader range of problems.

Details of the k-means clustering technique are provided below, followed by a formal definition of chi-squared.

Consider a data set with n observations and m characteristics and let $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots x_{im}\}$ denote a single observation in the data set. The $n \times m$ matrix comprising all characteristics for all observations is denoted by $\mathbf{X}$. Let $\mathbf{X}_p = \{X_{1p}, X_{2p}, \dots, X_{np}\}$ denote a vector of all observations for a specific characteristic p.

On completion of the k-means clustering algorithm all observations $\mathbf{x}_i$, with $i = \{1,2,\dots,n\}$, will have been assigned to one of the segments $S_1, S_2, \dots, S_K$ where each S_j denotes an index set containing the observation indices of all the variables assigned to it. That is, if observation $\mathbf{x}_i$ is assigned to segment S_j, then $i \in S_j$.

Further, let $\mathbf{u}_j = \{u_{j1}, u_{j2}, \dots u_{jm}\}$ denote the mean (centroid) of segment S_j, for example u_{j1} will be the mean of characteristic $\mathbf{X}_1$. The distance from each observation x_i to the segment mean $\mathbf{u}_j$ is given by a distance function $d(\mathbf{x}_i, \mathbf{u}_j)$. If a Euclidian distance measure is used, then $d(\mathbf{x}_i, \mathbf{u}_j) = ||x_i - \mathbf{u}_j||^2$ where $||.||^2$ defines the distance. Note that the double vertical bars indicate distance and hence imply that a square root is used.

The objective of ordinary k-means clustering is to minimise within-segment distances. For notational purposes, we introduce $c \in \mathbf{C}$ as an index of an assignment of all the observations to different segments, with $\mathbf{C}$ the set of all combinations of possible assignments. The notation S_{cj} is now introduced to reference all the observations for a given assignment $c \in \mathbf{C}$ and for a given segment index j. In addition, $\mathbf{u}_{cj}$ is the centroid of segment S_{cj}. The objective function of the ordinary k-means clustering algorithm can now be stated in generic form as

$$\min_{c \in \mathbf{C}} \sum_{j=1}^{K} \sum_{i \in S_j} d(\mathbf{x}_i, \mathbf{u}_{cj}), \tag{1}$$

Note that the notation used for the k-means clustering is the same notation as used in Breed et al.[20]

For the newly proposed SSSKMCSQ technique, a function is required to inform the segmentation process. For the supervised component, we will use the chi-squared value (rather than the information value).

The chi-square statistic is calculated as

$$X_c^2 = \sum_{j=1}^{K} \frac{\sum i \in s_{ch} Y_i - \frac{1}{n} \sum_{h=1}^{n} Y_h |S_{ch}|}{\frac{1}{n} \sum_{h=1}^{n} Y_h |S_{ch}|} \tag{2}$$

where n is the number of observations in the input data set; K is the number of segments over which X_c^2 is calculated; and y is the target variable and can be either binary or continuous. The term $|S|$ is used to represent the number of observations in segment S.

If chi-square is used in semi-supervised segmentation, then the supervised component $\rho(c)$ for each observation $\mathbf{x}_i$ and segment S_{cj} (with i assigned to S_{cj} in each case) can be defined as

$$\rho(c) = x^2 \tag{3}$$

Let $0 \leq w \leq 1$ be a weight that controls how much the clustering function is penalised by the chi-square statistic. The proposed optimisation problem for the SSSKMCSQ technique, taking within-segment distances into account, is the following

$$\min_{c \in \mathbf{C}} \left[w\rho(c) + (1-w) \sum_{j=1}^{K} \sum_{i \in S_j} d(S_{cj}, \mathbf{u}_{cj}) \right]. \tag{4}$$

In this paper, a heuristic approach is followed for the purpose of generating solutions to the optimisation problem in [4]. This includes

determining the optimal weight w for the supervised portion, using an algorithm that consists broadly of 10 steps similar to those of SSSKMIV. For details of the steps, see Breed et al.[20]

Semi-supervised segmentation: SSSWong (Variation 2)

Next, we propose a density-based semi-supervised technique using Wong's density clustering algorithm.[18] We call this the SSSWong technique (semi-supervised segmentation applied to Wong's density clustering methodology).

Predictive models are often developed for relatively large data sets (>1000 observations and 20 or more characteristics), and more common kernel-based density methods (like k-nearest neighbours[21]) are inviable because of their complexity. Wong's methodology combines the speed of k-means with the advantages of density-based clustering. It consists of two stages.[18,21] Note that these two stages are in essence an iterative process.

Stage 1: A preliminary clustering analysis is performed using a k-means algorithm with k much larger than the number of final clusters required.

Stage 2: The k-clusters formed in stage one are analysed and combined based on density-clustering dissimilarities until the required number of clusters are formed, or only a single cluster remains.

Preliminary clusters s_{cr} and s_{ct} are considered adjacent if the midpoint between the centroids u_{cr} and u_{ct} are closer to each other than any other preliminary-cluster mean based on Euclidean distance. Each thus has only one potential cluster with which it can be combined (with ties typically dealt with based on the order of the observations in the data set). The pair combined each time is that with the minimum density-based dissimilarity measure (see Wong[18] for further detail and the derivation):

$$\delta(s_{cr}, s_{ct}) = \begin{cases} \dfrac{\left(\sum_{i \in s_{cr}} d(\mathbf{u}_{ct}, \mathbf{x}_i) + \frac{1}{4}(|s_{cr}| + |s_{ct}| d(\mathbf{u}_{ct}, \mathbf{u}_{cr}))\right)^{\frac{m}{2}}}{(|s_{cr}| + |s_{ct}|)^{1 + \frac{m}{2}}} & \text{if } s_{cr} || s_{ct} \\ \infty & \text{otherwise} \end{cases} \quad (5)$$

where $|s|$ represents the number of observations in segment s and $s_{cr} || s_{ct}$ indicates that s_{cr} is adjacent to s_{ct}.

Wong's clustering was incorporated in the original semi-supervised technique (SSKMIV). We adjusted Wong's second step to incorporate the target variable. Thus, the algorithm optimises both cluster density and target rate differences. Let $c \in \mathbf{C}$ denote an index of an assignment of all the preliminary segments $s_{c1}, s_{c2}, ..., s_{cq}$ to the final segments $S_{c1}, S_{c2}, ..., S_{cK}$ with $K > q$ and with $\mathbf{C}$ the set of all combinations of possible assignments. In this case, q denotes the number of preliminary segments. Note that each will contain at least one observation, but is likely to contain a larger number that reduces computational complexity on large data sets.

The conglomeration of the preliminary segments into the final set of segments is done in a binary fashion, as illustrated by Figure 1.
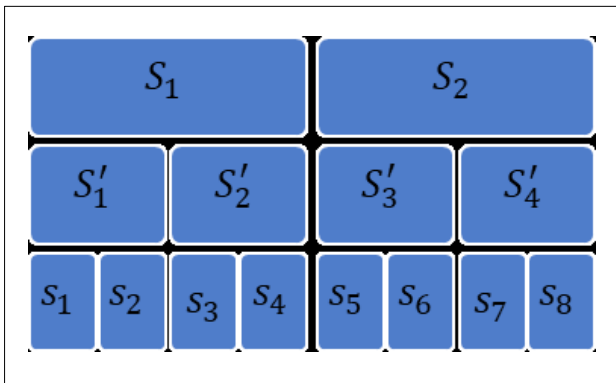

Figure 1: Example for notational purposes: Wong's method.

The final segments for the example are $S1 = \{S'_1, S'_2\} = \{s_1, s_2, s_3, s_4\}$ and $S_2 = \{S'_3, S'_4\} = \{s_5, s_6, s_7, s_8\}$. This previous example covers only one possible combination of assignments. We use the notation S_{cj} to represent any set of segments assigned to it for a given combination $c \in \mathbf{C}$. In order to evaluate the density dissimilarity between two segments or nodes, we make use of the notation $d(S_{cj})$. For example, to calculate the dissimilarity between nodes S'_1 and S'_2, we can calculate $d(S_{c1}) = d(s_1, s_2)$.

The proposed optimisation problem for the SSSWong algorithm is:

$$\min_{c \in \mathbf{C}} \left[w\rho(c) + (1-w) \sum_{j=1}^{K} d(S_{cj}) \right]. \quad (6)$$

Note the values of $\rho(c)$ and $d(S_{cj})$ are standardised for the same reasons as when using SSSKMIV.[20] For a single segmentation analysis using SSSWong, there are five steps:

1. Preliminary segmentation: Similar to Wong's method, the first step creates the preliminary segments that will be iteratively combined using formula [6].

2. Preliminary segment inspection: Preliminary segments are investigated to identify any with no events or non-events (which the information value calculation cannot handle), which are combined using Wong's standard density measure.

3. Determine preliminary segment adjacency: Adjacent segments are identified for each preliminary segment using a k-nearest neighbour type approach.

4. Combine segments until K left: Segments are iteratively combined until the required number of segments remains.

5. Calculate data set statistics: Statistics like information value obtained per segment are calculated and stored for further use.

The details of these steps can be found in Chapter 7 of a recent PhD thesis.[11]

Semi-supervised segmentation: SSKMIV$_{SSE}$ (Variation 3)

For the third variation we developed a semi-supervised technique with segment-size equality (SSE). We call this the SSSKMIV$_{SSE}$ technique (semi-supervised segmentation applied to the k-means algorithm using information value as supervised component, with the addition of segment size equality). Its purpose is to discourage the formation of small segments, or rather encourage segments of similar (or more equal) size.

Only minor adjustments were needed to the SSSKMIV's objective function[11], by introducing v as the SSE weight and δ as the SSE function. We define δ as

$$\delta(c) = \prod_{j=1}^{K} \frac{|S_{cj}|}{n_c} \quad (7)$$

where n_c is the total number of assigned observations for $c \in \mathbf{C}$. The function is at its maximum when all segment sizes ($|S_{c1}|, ..., |S_{cK}|$) are equal. Incorporating v and δ into the SSSKMIV technique results in a new objective function:

$$\min_{c \in \mathbf{C}} \left[w\rho(c) + v\,\delta(c) + (1-w-v) \sum_{j=1}^{K} \sum_{i \in S_{cj}} d(S_{cj}, \mathbf{u}_{cj}) \right] \quad (8)$$

where $w + v \leq 1$.

Semi-supervised segmentation: LK-Means (Variation 4)

The SSSKMIV variation has some similarities with the k-means semi-supervised segmentation algorithm, proposed in Peralta et al.[19] which is called LK-Means. This methodology has many similarities to SSS algorithms, but also some clear differences.[11] For our fourth variation we augmented the LK-Means methodology. It is thus not really a new

SSS technique, but an existing technique adapted to be comparable with others presented in this paper.

All four variations of semi-supervised segmentation methods (as well as the original SSSKMIV) were implemented in SAS software (Version 9.4, SAS Institute Inc., Cary, NC, USA). The detail of the technical specifications (e.g. the optimal number of segments, the weight parameters in SSS, the optimal value of k in the k-mean algorithm, and a heuristic example) can be found in Breed[11].

To facilitate representing the objective function of the LK-Means algorithm mathematically, we expand the S_{cj} notation to S_{cjl}, to reference all the observations for a given assignment $c \in \mathbf{C}$, for a given segment index j and a given label l. Similarly, u_{cjl} represents the mean, or centroid of S_{cjl}. For this algorithm, the assumption is that the labels (or target variable values) take on L discrete values and are not continuous. The objective function of the LK-Means algorithm to be minimised becomes

$$\min_{c \in \mathbf{C}} \left[W \sum_{j=1}^{K} \sum_{l=1}^{L} \sum_{i \in S_{cj}} d(S_{cj}, \mathbf{U}_{cj}) v_{cjl} + (1-w) \sum_{j=1}^{K} \sum_{i \in S_{cj}} d(S_{cj}, \mathbf{U}_{cj}) \right] \quad (9)$$

where v_{cjl} is the ratio of the number of observations assigned to cluster j with label l divided by the number of total observations assigned to cluster j. This ratio represents the 'confidence' of label l in cluster j. The distortion weight, w, is similar to the weight in SSSKMIV and again adjusts the supervised element with values between 0 and 1. More details of these steps can be found in Chapter 7 of the PhD thesis.[11]

How to measure model performance: Data splitting and Gini coefficient

In order to compare model performance, each data set was divided randomly into equally sized development and validation sets. Data splitting is the dividing of a sample into two parts and then developing a hypothesis using one part and testing it on the other.[22] Picard and Berk[23] review it in the context of regression and provide specific guidelines for the validation of regression models, i.e. 25% to 50% of the data is recommended for validation. Faraway[24] illustrates that split-data analysis is preferred to a full-data analysis for predictions with some exceptions.

We used the development set (i.e. training data) to develop the predictive models, whilst the validation set (i.e. hold-out data) was used to assess model performance (hereafter the 'lift'). Lift was measured by calculating Gini coefficients[2], to quantify a model's ability to discriminate between two possible values of a binary target variable[17]. Cases are ranked according to the predictions, and the Gini then provides a measure of correctness. It is one of the most popular measures used in retail credit scoring[1-3,25], and has the added advantage that it is a single number[17]. For this paper, values are calculated for the combined validation data sets. Although we used only Gini in this paper, more measures were used in the original PhD thesis.[11]

Description of data sets

The above segmentation techniques were compared on six different data sets, described below. All explanatory variables were standardised by transforming them into z-scores, i.e. subtracting the mean and dividing by the standard deviation of each based on the full development data set. Weights of evidence or dummy variables would have been preferable, but were not considered because of the added complexity of binning each predictor – especially if done per segment. We cannot say whether or how the transformation methodology might have affected the results.

The data sets are the same as those used in the previous study.[12] A short summary of the data used is given in Table 1. Details on the data sets can be found in Breed[11].

Table 1: Details of the six data sets

Data set	Short description	Size
1. Direct marketing	This data set contains information about a bank's customers, the products they have with the bank, and their utilisation of and behaviour with those products. The target variable is binary and indicates whether the customer responded to a direct marketing campaign for a personal loan or not.	24 explanatory variables and 4720 observations
2. Protein structures[26,27]	This data set contains results of experiments performed by the Protein Structure Prediction Centre[27] on the latest protein structure prediction algorithms.	9 explanatory variables and 45 730 observations
3. Credit applications[2]	This data set contains 10 characteristics of customers who applied for credit. The target variable is binary, indicating whether or not the customer experienced a 90 days' or worse delinquency.	10 explanatory variables and 150 000 observations
4. Wine quality[26,29]	This data set contains physicochemical properties of wines that are extracted through analytical tests that can be easily performed on most wines. The target variable is derived from a score between 0 and 10 which indicates the quality of the wine as scored by tasting experts. The binary target variable that is used for this analysis indicates whether the score is greater than 6, thereby indicating a great quality wine (only 20% of the wines score greater than 6).	11 explanatory variables and 6497 observations
5. Chess king-rook vs king[26,30,31]	The data set is an 'Endgame database', which is a table of stored game-theoretic values for the legal positions of the pieces on a chessboard. This data set was first described by Clarke[32].	18 explanatory variables and 28 056 observations
6. Insurance claims[28]	The data set was used in a competition named 'Claim Prediction Challenge (Allstate)' concluded in 2011. The binary target that was used in this data set indicates whether or not a claim payment was made. The independent variables have been hidden but, according to the website, it contains information about the vehicle to which the insurance applies as well as some particulars about the policy itself.	12 explanatory variables and 14 782 observations[8]

Empirical results

The five semi-supervised segmentation techniques described above were applied to all six data sets, with performance assessed on the validation data. Results for all five are presented in Tables 2–7, respectively, with Table 8 providing a summary of the results. Note that for a comparison of supervised and unsupervised techniques, please refer to other research studies.[11-13] Also, while our focus was to compare different semi-supervised segmentation techniques, we have also included an unsegmented logistic regression in each table as a further baseline.

Table 2 summarises the performance of the modelling techniques when applied to the **direct marketing** data set (as measured by the Gini coefficient calculated on the validation set). SSSKMCSQ achieved the best result, with SSSKMIV second.

Table 3 summarises the results for the **protein tertiary structures** data set, where the ranking order is completely different from that in Table 2. As a start, SSSKMCSQ ranks fourth of five. Best is SSSKMIV, with SSSKMIV$_{SSE}$ second. The Gini coefficients are between 65% and 70%, which are quite high values.

Table 4 shows results for the **credit application** data set, where SSSKMIV again outperforms the other techniques. Note that strong bureau data as well as internal data were available on this credit application data set, hence the relatively high Gini values. The large difference between the unsegmented and segmented results is highly unusual, and may be related to the use of z-scores (i.e. standardisation of variables). It may be that the variables that predict credit risk (delinquency) best, are the least normally distributed.

For the **wine quality** data set, Table 5 shows that one of our new variations takes top position: LK-Means.

Table 6 shows results for the chess king-rook vs. king data set, where LK-Means again dominates. It is interesting that the Gini coefficients achieved are very high, from 75% to almost 88%. It seems that it is easier to obtain efficient ranking in this data set, which relates to a highly structured game.

Table 7 shows the results for the last data set, which is for **insurance claim prediction**. In this case, SSSKMIV$_{SSE}$ works best. The Gini coefficients are very low (Gini ranging between 12% and 16%), which makes one wonder about whether predictive models can provide any value in this domain.

And finally, Table 8 provides a summary of the median and average ranks for all five semi-supervised segmentation techniques.

Table 2: Direct marketing data set: comparison of performance of techniques

Modelling technique	Best Gini obtained	Rank
Unsegmented logistic regression	22.02%	
Semi-supervised segmentation with logistic regression		
SSSKMIV	27.89%	2
SSSKMCSQ	29.14%	1
SSSWong	19.04%	5
SSSKMIVSSE	23.53%	4
LK-Means	24.29%	3

Table 3: Protein tertiary structures data set: comparison of performance of techniques

Modelling technique	Best Gini obtained	Rank
Unsegmented logistic regression	59.39%	
Semi-supervised segmentation with logistic regression		
SSSKMIV	70.37%	1
SSSKMCSQ	67.99%	4
SSSWong	65.35%	5
SSSKMIVSSE	70.14%	2
LK-Means	69.51%	3

Table 4: Credit application data set: comparison of performance of techniques

Modelling technique	Best Gini obtained	Rank
Unsegmented logistic regression	38.45%	
Semi-supervised segmentation with logistic regression		
SSSKMIV	66.25%	1
SSSKMCSQ	65.39%	3
SSSWong	61.34%	5
SSSKMIVSSE	65.44%	2
LK-Means	64.54%	4

Table 5: Wine quality data set: comparison of performance of techniques

Modelling technique	Best Gini obtained	Rank
Unsegmented logistic regression	61.17%	
Semi-supervised segmentation with logistic regression		
SSSKMIV	66.97%	2
SSSKMCSQ	66.91%	3
SSSWong	64.26%	5
SSSKMIVSSE	66.86%	4
LK-Means	68.05%	1

Table 6: Chess king-rook vs. king data set: comparison of performance techniques

Modelling technique	Best Gini obtained	Rank
Unsegmented logistic regression	71.03%	
Semi-supervised segmentation with logistic regression		
SSSKMIV	86.60%	2
SSSKMCSQ	86.49%	3
SSSWong	75.52%	5
SSSKMIVSSE	86.30%	4
LK-Means	87.33%	1

Table 7: Insurance claim prediction data set: comparison of performance of techniques

Modelling technique	Best Gini obtained	Rank
Unsegmented logistic regression	8.29%	
Semi-supervised segmentation with logistic regression		
SSSKMIV	15.19%	2
SSSKMCSQ	14.89%	3
SSSWong	12.80%	5
SSSKMIVSSE	15.24%	1
LK-Means	14.76%	4

Table 8: Median and average rank of the semi-supervised segmentation (with logistic regression) techniques across all data sets

Modelling technique	Median rank	Average rank
SSSKMIV	2	1.67
SSSKMCSQ	3	2.83
SSSWong	5	5.00
SSSKMIV$_{SSE}$	3	2.83
LK-Means	3	2.67

The original SSSKMIV technique performed best with a median rank of 2 and average of 1.67. Three of the four variations (SSSKMCSQ, SSSKMIV$_{SSE}$, LK-Means) achieved a median rank of 3, while LK-Means achieved an average rank of 2.67 (only slightly higher than SSSKMCSQ and SSSKMIV$_{SSE}$). The overall loser was SSSWong, which came in last across the board.

Comments on using Gini as an absolute value

The analysis above illustrates the problem of using Gini as an absolute value.[27] The best was 87.33% for LK-Means on the chess data set, but for the insurance data the best was SSSKMIV$_{SSE}$ with a Gini of 15.24%. Such results are not a reflection of the techniques being used, but the data under consideration.[33] It is unreasonable to have a minimum Gini that is broadly applied.[34] Using Gini coefficients for comparison makes sense only if the data are comparable – in this instance different models applied to the same data.

Concluding remarks

We proposed four newly developed semi-supervised segmentation techniques and provided their mathematical notation. Additionally, we evaluated our four variations against the original semi-supervised technique, SSSKMIV, on six different data sets, with Gini coefficients derived using combined validation data for each segment. The original SSSKMIV technique performed best overall and was the outright winner for two of the data sets, but other variations dominated elsewhere. Best performers were SSSKMIV in the protein and credit data sets, LK-Means in the wine quality and chess data sets, SSSKMIV$_{SSE}$ in the insurance prediction data set and SSSKMCSQ in the direct marketing data set. The SSSWong technique produced the worst overall results, perhaps because some of k-means' weaknesses were already addressed by SSSKMIV[11] and the additional complexity of SSSWong adds no additional benefit.

We conclude that the four alternatives provide additional tools for segmenting data before fitting a logistic regression. Of the four, SSSWong is quickest to perform on a standard PC, but performs worst (as per the results observed). SSSKMCSQ is most versatile (as it can be performed on both binary and continuous variables) and achieves reasonable results. The most optimal variation will, however, be dependent on the characteristics of the data set being analysed.

The benefit of segmentation was also clearly illustrated in the six data sets used in previous work,[12] although the impact of the transformation methodology is not known. In this study, we have also clearly highlighted the danger of using an absolute Gini coefficient to evaluate the performance of any predictive model. The relative Gini value is more appropriate. Future research could include investigating which properties of data sets contribute to the differences in performance between the techniques. Another extension of the research could be to use measures other than Gini and information value; many other measures exist that could be alternatives to these values. Further comparisons could be done using an array of such alternative measures. It would also provide value to investigate transformation methodologies other than the z-score when doing such research.

Acknowledgements

We gratefully acknowledge the valuable input given by the anonymous referees. This work is based on research supported in part by the

Authors' contributions

D.G.B. was responsible for conceptualisation; methodology; data collection; data analysis; validation; data curation; writing revisions. T.V. was responsible for conceptualisation; sample analysis; data analysis; writing the initial draft; revisions; student supervision; project leadership and project management.

References

1. Anderson R. The credit scoring toolkit: Theory and practice for retail credit risk management and decision automation. New York: Oxford University Press; 2007.

2. Siddiqi N. Credit risk scorecards. Hoboken, NJ: John Wiley & Sons; 2006.

3. Thomas LC. Consumer credit models. New York: Oxford University Press; 2009. http://dx.doi.org/10.1093/acprof:oso/9780199232130.001.1

4. Belson W. Effects of television on the interests and initiative of adult viewers in greater London. Br J Psychol. 1959;50(2):145–158. https://doi.org/10.1111/j.2044-8295.1959.tb00692.x

5. Morgan JN, Solenberger PW, Nagara PR. History and future of binary segmentation programs. University of Michigan: Survey Research Center, Institute for Social Research; 2015. Available from: http://www.isr.umich.edu/src/smp/search/search_paper.html

6. Hand DJ. Construction and assessment of classification rules. West Sussex: John Wiley & Sons; 1997.

7. Kass GV. An exploratory technique for investigating large quantities of categorical data. J R Stat Soc Ser C Appl Stat. 1980;29(2):119–127. https://doi.org/10.2307/2986296

8. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Berlin: Springer; 2001. https://doi.org/10.1007/978-0-387-21606-5_14

9. Cross G. Understanding your customer: Segmentation techniques for gaining customer insight and predicting risk in the telecom industry. Proceedings of the SAS Global Forum 2008; 2008 March 16–19. Cary, NC: SAS Institute Inc.; 2008. Paper 154-2008. Available from: http://www2.sas.com/proceedings/forum2008/154-2008.pdf

10. Fico. Using segmented models for better decisions [document on the Internet]. c2014 [cited 2015 Jan 05]. Available from: http://www.fico.com/en/node/8140?file=9737

11. Breed DG. Semi-supervised segmentation within a predictive modelling context [PhD thesis]. Potchefstroom: North-West University; 2017.

12. Breed DG, Verster T. The benefits of segmentation: Evidence from a South African bank and other studies. S Afr J Sci. 2017;113(9/10), Art. #2016-0345, 7 pages. http://dx.doi.org/10.17159/sajs.2017/20160345

13. Breed DG, De La Rey T, Terblanche SE. The use of different clustering algorithms and distortion functions in semi supervised segmentation. In: Proceedings of the 42nd Operations Research Society of South Africa Annual Conference; 2013 September 15–18; Stellenbosch, South Africa. Available from: http://www.orssa.org.za/wiki/uploads/Conf/ORSSA2013_Proceedings.pdf

14. Grira N, Crucianu M, Boujemaa N. Unsupervised and semi-supervised clustering: A brief survey. In: A review of machine learning techniques for processing multimedia content: Report of the MUSCLE European Network of Excellence [document on the Internet]. c2005 [cited 2018 Jan 08]. Available from: http://cedric.cnam.fr/~crucianm/src/BriefSurveyClustering.pdf

15. Eick CF, Zeidat N, Zhao Z. Supervised clustering algorithms and benefits. In: Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI04); 2004 November 14–17; Boca Raton, FL, USA. Washington DC: IEEE; 2004. p. 774–776. http://dx.doi.org/10.1109/ICTAI.2004.111

16. Hung WC, Tsai YH, Liou YT, Lin YY, Yang MH. Adversarial learning for semi-supervised semantic segmentation. CoRR (Computer Vision and Pattern Recognition) [preprint]. arXiv 2018; #1802.07934. Available from: http://arxiv.org/abs/1802.07934

17. Tevet D. Exploring model lift: Is your model worth implementing? Actuarial Rev. 2013;40(2):10–13.

18. Wong MA. A hybrid clustering method for identifying high-density clusters. J Am Stat Assoc. 1982;77(380):841–847. http://dx.doi.org/10.2307/2287316

19. Peralta B, Espinace P, Soto A. Enhancing K-Means using class labels. Intell Data Anal. 2013;17(6):1023–1039. http://dx.doi.org/10.3233/IDA-130618

20. Breed DG, Verster T, Terblanche SE. Developing a semi-supervised segmentation algorithm as applied to k-means using information value. ORION. 2017;33(2):85–103. http://dx.doi.org/10.5784/33-2-568

21. Wong MA, Lane T. A kth nearest neighbour clustering procedure. In: Eddy WF, editor. Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface. New York: Springer; 1981. https://doi.org/10.1007/978-1-4613-9464-8_46

22. Barnard GA. Discussion of cross-validatory choice and assessment of statistical predictions. In: Stone M. Cross-validatory choice of statistical predictions. J R Stat Soc Ser B Stat Methodol. 1974;36:111–135.

23. Picard P, Berk K. Data splitting. Am Stat. 1990;44(2):140–147. http://dx.doi.org/10.1080/00031305.1990.10475704

24. Faraway JJ. Does data splitting improve prediction. Stat Comput. 2016;26:49–60. http://dx.doi.org/10.1007/s11222-014-9522-9

25. Baesens B, Roesch D, Scheule H. Credit risk analytics: measurement techniques, applications, and examples in SAS. Hoboken, NJ: Wiley; 2016.

26. Lichman M. Physicochemical properties of protein tertiary structure [data set on the Internet]. In: UCI Machine Learning Repository. c2013 [cited 2016 May 06]. Available from: http://archive.ics.uci.edu/ml

27. Protein Structure Prediction Center [homepage on the Internet]. c2015 [cited 2016 Jun 04]. Available from: http://predictioncenter.org/

28. Kaggle [homepage on the Internet]. c2016 [cited 2016 Sep 23]. Available from: http://www.kaggle.com

29. Cortez P, Cerdeira A, Almeida F, Matos T, Reis J. Modeling wine preferences by data mining from physicochemical properties. Decis Support Syst. 2009;47(4):547–553. http://dx.doi.org/10.1016/j.dss.2009.05.016

30. Casti JL, Casti JL. Five golden rules: Great theories of 20th-century mathematics and why they matter. New York: John Wiley & Sons; 1996.

31. Russell SJ, Norvig P. Artificial Intelligence: A modern approach. 2nd ed. New Delhi: Pearson Education; 2003.

32. Clarke M. A quantitative study of king and pawn against king. In: Clarke MRB, editor. Advances in computer chess. Edinburgh: Edinburgh University Press; 1977. p. 108–118.

33. Hamerle A, Rauhmeier R, Rösch D. Uses and misuses of measures for credit rating accuracy. Technical Paper from the University of Regensburg. c2003. [cited 2017 Jan 17]. Available from: http://mx.nthu.edu.tw/~jtyang/Teaching/Risk_management/Papers/Testing/Uses%20and%20Misuses%20of%20Measures%20for%20Credit%20Rating%20Accuracy.pdf

34. Engelman B, Rauhmeier R. The Basel II risk parameters: Estimation, validation, and stress testing. 2nd ed. Berlin: Springer; 2011. https://doi.org/10.1007/978-3-642-16114-8

AUTHORS:
Maryke Schoonen[1] (iD)
Albertus S. Seyffert[2] (iD)
Francois H. van der Westhuizen[1] (iD)
Izelle Smuts[3] (iD)

AFFILIATIONS:
[1]Human Metabolomics, North-West University, Potchefstroom, South Africa
[2]Centre for Space Research, North-West University, Potchefstroom, South Africa
[3]Department of Paediatrics and Child Health, Steve Biko Academic Hospital, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Maryke Schoonen

EMAIL:
mschoonen28@gmail.com

A bioinformatics pipeline for rare genetic diseases in South African patients

The research fields of bioinformatics and computational biology are growing rapidly in South Africa. Bioinformatics pipelines play an integral part in handling sequencing data, which are used to investigate the aetiology of common and rare diseases. Bioinformatics platforms for common disease aetiology are well supported and continuously being developed in South Africa. However, the same is not the case for rare diseases aetiology research. Investigations into the latter rely on international cloud-based tools for data analyses and ultimately confirmation of a genetic disease. However, these tools are not necessarily optimised for ethnically diverse population groups. We present an in-house developed bioinformatics pipeline to enable researchers to annotate and filter variants in either exome or amplicon next-generation sequencing data. This pipeline was developed using next-generation sequencing data of a predominantly African cohort of patients diagnosed with rare disease.

Significance:
- We demonstrate the feasibility of in-country development of ethnicity-sensitive, automated bioinformatics pipelines using free software in a South African context.
- We provide a roadmap for development of similarly ethnicity-sensitive bioinformatics pipelines.

Introduction

The research fields of bioinformatics and computational biology are both growing rapidly in South Africa, with an ever-increasing number of both small and large laboratories having access to next-generation sequencing (NGS) technologies. This increase in sequencing capability continues to stimulate the development of a variety of platforms, databases and initiatives, such as H3Africa (https://www.h3africa.org), the South African Human Genome Programme (http://sahgp.sanbi.ac.za), and the South African Bioinformatics Institute (https://www.sanbi.ac.za), to support sequencing data analysis. However, to date, the majority of these developments have been focused on common disease related research, because diseases like HIV, fibromyalgia, tuberculosis and malaria are major health challenges in southern Africa.[1]

Southern African researchers doing rare-disease related research (diseases affecting 6–8% of the global population)[2] still rely on flexible, international cloud-based tools, such as Bystro[3], BrowseVCF[4], and RD-Connect[5], as their analysis needs have not yet been fully met locally. For European populations, these tools allow researchers to leverage well-established genotype–phenotype correlations to guide investigations into rare disease aetiology. However, these correlations do not always hold for African populations, thus significantly reducing the power of and degree to which these online tools can be relied on in the South African research context. A niche therefore exists for a tool that is both sensitive to population heterogeneity and general enough in nature to enable effective domestic rare disease research.

African population groups are heterogeneous with large genetic variety and limited information on genotype–phenotype correlations.[6] Even though data are processed using the same reference genome, some aspects, such as allele frequency for disease-causing variants and genotype–phenotype correlation, could differ between ethnically diverse population groups[7] and must be taken into account.

In this paper, we present a bioinformatics pipeline developed in-house to address these limitations. This pipeline processes NGS data of ethnically diverse population groups without any strong prior assumptions regarding genotype–phenotype correlations. This offline pipeline, suitable for the analysis of both exome and amplicon sequencing data, is written in Bash (or the Bourne-Again SHell) and uses only open-source software. To allow other researchers to benefit from this work, the pipeline has been made available on GitHub under the GNU GPLv3 software licence.[8] The Ensembl Variant Effect Predictor (VEP)[9] offline script is used for variant annotation, and the Genome Mining (GEMINI)[10] command line database management tool for variant filtering. The pipeline is easily adjustable with regard to what annotations are made, and how they are filtered, which is especially useful when working with NGS data from ethnically diverse patients. The bioinformatics pipeline for variant annotation and filtering of amplicon and exome sequencing data presented here has been successfully used in research forming part of the project 'Investigating the aetiology of South African paediatric patients diagnosed with mitochondrial disorders'.[11-14]

Bioinformatics pipeline

Workflows leveraging NGS technology mainly consist of two parts: primary and secondary analysis. Figure 1 shows what this workflow looks like when our newly developed pipeline is incorporated. During primary analysis, patient samples are prepared and sequenced on a specialised platform such as Ion Torrent (used in our research case) or Illumina. These platforms further perform the requisite signalcalling, basecalling, reference sequence alignment and variant calling. The output from this primary analysis, for a single sample, is a Variant Call Format (VCF) file listing all the variants for that sample (*variants_patientX.vcf.gz* in Figure 1).[15] Secondary analysis, often with the identification of disease-causing variants in mind, entails variant annotation and filtering. Variant annotation is typically done using purpose-built tools, such as ANNOVAR[16] and VEP runner (used in our research case) that

annotate variants with relevant metadata such as variant type, variant allele frequency and predicted impact. Variant filtering, in which variants of interest are identified, can be done using tools like BrowseVCF, RD-Connect and GEMINI (used in our research case). These tools filter variants based on the metadata associated with each, which allows a researcher to find the variants most relevant to their investigation.
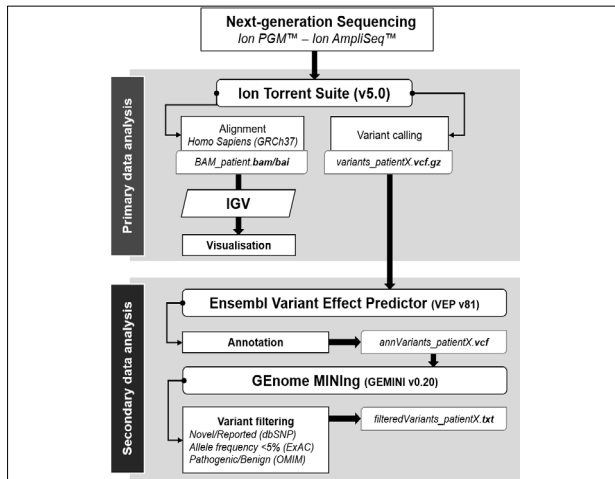
PGM, Personal Genome Machine; BAM, binary alignment/map; IGV, Integrative Genomics Viewer; VCF, Variant Call Format; dbSNP, database for single nucleotide polymorphisms; ExAC, Exome Aggregation Consortium; OMIM, Online Mendelian Inheritance in Man.

Figure 1: Next-generation sequencing bioinformatics pipeline followed for identifying disease-causing variants from Ion Torrent sequencing data. The first component of this illustration is primary data analysis, which is a semi-automated process done using the Ion Torrent Software (Torrent Suite). The second component illustrates software used for secondary data analysis.

The pipeline presented in this paper is focused exclusively on secondary analysis because the problematic potential population biases, which stem from variant annotations made based on European-population-centric research results, influence only the secondary analysis results. The pipeline utilises the offline VEP script's flexibility to annotate variants sufficiently comprehensively so that population-sensitive variant filtering can be achieved using GEMINI's high-specificity querying capabilities. Our pipeline is intended to form part of a larger NGS data analysis workflow, and is only responsible for variant annotation and filtering.

Secondary data analysis

Bash is the default command shell on most modern Unix-like systems (and Unix-like tools for Windows), and incorporates useful features from the Korn shell and the C shell.[17,18] The ability for Bash users to write powerful scripts to automate analysis has made it a popular choice for research implementations, typically in the form of command line tools and analysis pipelines. Many of these command line tools are available as free software, giving researchers who use the Bash shell access to a plethora of bioinformatics tools with which to do their research.

The main advantage of using Bash is that it allows powerful automation of established tools in a natural way while minimising both the number of introduced software dependencies and the need for more advanced analysis infrastructure. These are important advantages considering the rare-disease focus of this pipeline. For researchers who focus exclusively on rare diseases, whose expertise and research interests often lie outside bioinformatics, a pipeline based on Bash allows flexible analyses that remain easily repeatable over long periods of time, while necessitating minimal skills development and infrastructure investment.

Here, we present the bioinformatics pipeline developed using these tools during our research. Scripts were run on Slackware v14.12 with GNU Bash, version 4.4.12.

As primary analysis, using the Ion Torrent system, delivers sequencing results in batches, it seemed prudent to write the secondary analysis scripts to operate in a batch fashion. This approach, coupled with the built-in automation that comes with scripting, ensures consistency across the samples in each batch and across different batches.

In the secondary analysis pipeline, the scripts implement two steps: variant annotation and data mining. Variant annotation is done using the VEP script, which can be downloaded from Ensembl's website.[9] Data mining is done using the command line tool GEMINI, which uses the SQLite relational database management system to enable effective sequencing data mining.[10] Each of these steps has a Bash script dedicated to it: **vep_single.sh** and **gemini_single.sh**, respectively (the '.sh' extension indicates that these text files are Bash scripts). These scripts are embedded into the **vep_batch.sh** and **gemini_batch.sh** scripts, respectively, which run them for each .vcf.gz file in a given directory. The **hetero_annotate.sh** script binds these two batch scripts into a full pipeline, calling each in sequence and managing their inputs and outputs. See Appendices 1–6 in the supplementary material for the full contents of these scripts. A brief description of what each script does is given below.

The **vep_single.sh** script (Supplementary appendix 1) takes an Ion Torrent-generated input .vcf.gz file as its first argument, and runs the VEP for it. It takes an output file as its second argument (a .vcf file), to which the annotated output of the VEP is written. Alongside these two arguments, a number of additional arguments are passed to the VEP script, with the most notable being the --fields [list] argument, which allows for the specification of the required annotation fields included in the annotated output .vcf file. This argument (as well as the specification of the VEP arguments) is handled in an extensible way by the **vep_single.sh** script, and the list of desired fields can be built up over multiple lines, or in multiple groups. A list of all possible fields and a list of all possible additional arguments can be found in the VEP's documentation.[9,19] The VEP also generates a 'statistic run report' (.html file) when run, containing general statistics that give information on, among other things, the number of variants processed, number of overlapping genes, and number of novel/reported variants.

The gemini_single.sh (Supplementary appendix 2) script takes a VEP-annotated input .vcf file as its first argument and loads it into a SQLite database (a .db file). This file is then mined for relevant variants using user-defined SQLite database query specifications. Each line in the queries_spec.txt (Supplementary appendix 3) auxiliary text file contains one such query specification, and consists of two comma-separated fields: the query's name and the relevant SQLite query snippet. Queries can easily be added to or removed from this file, or a different such file can be specified in the gemini_single.sh script. An example of such a query, which filters based on variants' allele frequency in African populations, is: *rareAFR,aaf_1kg_afr <= 0.01*.

What information these queries should return for each variant is controlled by the *cols* variable, which is used to build the full SQLite query. Some examples of columns that can be included in a query are *is_coding* (which is true if the variant is in a coding region), *rs_ids* (which lists the rsIDs associated with each variant), and *aaf_1kg_afr* (which stores the allele frequency of the variant in the AFR population as reported in the 1000 Genomes Project). For a list of all possible columns, see GEMINI's documentation.[20] The *cols* variable is handled similarly to the --fields [list] argument to the VEP, and is similarly extensible. The output of each query is a list of variants, with information from the database columns specified in the *cols* variable, stored in a .txt file that has the query's name as filename. These files constitute the main output of the pipeline.

Once all the queries have been performed, a *meta_info.txt* file is generated that summarises the lengths of the lists contained in the query output files. This file, along with the generated query outputs, is saved in the folder specified as the second argument to the **gemini_single.sh** script.

The **vep_batch.sh** and **gemini_batch.sh** scripts (Supplementary appendices 4 and 5, respectively) each run their counterpart script, as discussed above, for all the .vcf.gz files in a given directory. Both

of these scripts take the directory containing the relevant input files as their first argument, with the second argument being the directory to which output should be written.

Finally, the **hetero_annotate.sh** script (Supplementary appendix 6) administers the operation of the two batch scripts described above. The first argument of this script is the directory containing the Ion Torrent output *.vcf.gz* files (the first argument for the **vep_batch.sh** and **vep_single.sh** scripts), and the second argument is the directory to which the GEMINI output should be written (the second argument to the **gemini_batch.sh** and **gemini_single.sh** scripts).

From here the user can further prioritise and filter the variants according to the criteria of their choice. For our African data set, variants were first filtered based on the novelty of these variants. Second, variants were filtered based on African allele frequencies, as reported in Exome aggregation consortium (ExAC[21]). In addition, for mostly non-African population groups, known disease-causing variants can be identified from the data set using databases such as the Online Mendelian Inheritance in Man (OMIM[22]) and ClinVar[23].

Conclusion

Robust bioinformatics pipelines are key components for diagnosis and research of rare genetic diseases. Here we describe an offline, flexible and open-source bioinformatics pipeline that annotates variants using VEP and filtering of important disease-causing variants using GEMINI from NGS data. It was developed as a first prudent step towards data processing and offers unique advantages for the detection of multiple genetic alterations, including in patients of African descent for whom little information is available. The pipeline can be used on exome as well as amplicon NGS data and was designed using NGS data of a predominantly African cohort with rare disease. Identifying underlying genetic causes for rare disease is particularly challenging in the South African population, with limited bioinformatics support for researchers and non-bioinformaticians. With the increased burden to diagnose rare genetic diseases using NGS and genetic screening, and the limited support and resources in developing countries such as South Africa, it is equally important to develop and provide access to bioinformatics pipelines, such as this one. With continued development, pipelines in South Africa could be further refined and made more user-friendly, making them useful for both researchers and clinicians. These refinements could include, for instance, cloud-based interfaces like those used in developed countries. With such refinements in place, clinicians would more easily be able to investigate genotype–phenotype correlation in rare diseases for African population groups.

Acknowledgements

Authors' contributions

M.S. and A.S.S. were responsible for: conceptualisation, methodology and writing – the initial draft. F.H.v.d.W. was responsible for: writing – revisions, student supervision and project leadership. I.S. was responsible for: writing – revisions, student supervision and funding acquisition.

References

1. Mulder NJ, Christoffels A, De Oliveira T, Gamieldien J, Hazelhurst S, Joubert F, et al. The development of computational biology in South Africa: Successes achieved and lessons learnt. PLoS Comput Biol. 2016;12(2), e1004395, 15 pages. https://doi.org/10.1371/journal.pcbi.1004395

2. Dawkins HJ, Draghia-Akli R, Lasko P, Lau LP, Jonker AH, Cutillo CM, et al. Progress in rare diseases research 2010–2016: An IRDiRC perspective. Clin Transl Sci. 2018;11(1):11–20. https://doi.org/10.1111/cts.12501

3. Kotlar AV, Trevino CE, Zwick ME, Cutler DJ, Wingo TS. Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. Genome Biol. 2018;19(1), Art. #14, 11 pages. https://doi.org/10.1186/s13059-018-1387-3

4. Salatino S, Ramraj V. BrowseVCF: A web-based application and workflow to quickly prioritize disease-causative variants in VCF files. Brief Bioinform. 2016;18(5):774–779. https://doi.org/10.1093/bib/bbw054

5. Thompson R, Johnston L, Taruscio D, Monaco L, Béroud C, Gut IG, et al. RD-Connect: An integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med. 2014;29(3):780–787. https://doi.org/10.1007/s11606-014-2908-8

6. Van der Westhuizen FH, Sinxadi PZ, Dandara C, Smuts I, Riordan G, Meldau S, et al. Understanding the implications of mitochondrial DNA variation in the health of black southern African populations: The 2014 Workshop. Hum Mutat. 2015;36(5):569–571. https://doi.org/10.1002/humu.22789

7. Shearer AE, Eppsteiner RW, Booth KT, Ephraim SS, Gurrola II J, Simpson A, et al. Utilizing ethnic-specific differences in minor allele frequency to recategorize reported pathogenic deafness variants. Am J Hum Genet. 2014;95(4):445–453. https://doi.org/10.1016/j.ajhg.2014.09.001

8. Seyffert AS. A set of BASH scripts for annotating and filtering next-generation sequencing data using Ensembl's offline VEP runner GEMINI [software]. c2018 [cited 2018 Oct 01]. Available from: https://github.com/aseyffert/hetero_annotate

9. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol. 2016;17(1), Art. #122, 14 pages. https://doi.org/10.1101/042374

10. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: Integrative exploration of genetic variation and genome annotations. PLoS Comput Biol. 2013;9(7), e1003153, 8 pages. https://doi.org/10.1371/journal.pcbi.1003153

11. Van der Westhuizen FH, Smuts I, Honey E, Louw R, Schoonen M, Jonck L-M, et al. A novel mutation in ETFDH manifesting as severe neonatal-onset multiple acyl-CoA dehydrogenase deficiency. J Neurol Sci. 2017;384:121–125. https://doi.org/10.1016/j.jns.2017.11.012

12. Louw R, Smuts I, Wilsenach K-L, Jonck L-M, Schoonen M, Van der Westhuizen FH. The dilemma of diagnosing coenzyme Q10 deficiency in muscle. Mol Genet Metab. 2018;125(1–2):38–48. https://doi.org/10.1016/j.ymgme.2018.02.015

13. Van der Walt EM, Smuts I, Taylor RW, Elson JL, Turnbull DM, Louw R, et al. Characterization of mtDNA variation in a cohort of South African paediatric patients with mitochondrial disease. Eur J Hum Genet. 2012;20(6):650–656. https://doi.org/10.1038/ejhg.2011.262

14. Smuts I, Louw R, Du Toit H, Klopper B, Mienie LJ, Van der Westhuizen FH. An overview of a cohort of South African patients with mitochondrial disorders. J Inherit Metab Dis. 2010;33(3):95–104. http://dx.doi.org/10.1007/s10545-009-9031-8

15. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330

16. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16), e164, 7 pages. https://doi.org/10.1093/nar/gkq603

17. Ramey C, editor. Bash, the Bourne−Again Shell. In: Proceedings of The Romanian Open Systems Conference & Exhibition (ROSE 1994), The Romanian UNIX User's Group (GURU); 1994 November 3–5; Bucharest, Romania.

18. Stevens WR, Rago SA. Advanced programming in the UNIX environment. Ann Arbor, MI: Addison-Wesley; 2013.

19. The Ensembl Genome Database Project, NCBI, NIH. Variant Effect Predictor [software]. c2018 [cited 2018 Oct 01]. Available from: https://www.ensembl.org/info/docs/tools/vep/index.html

20. Genome Mining. GEMINI: A flexible framework for exploring genome variation [software]. c2017 [cited 2018 Oct 01]. Available from: https://gemini.readthedocs.io/en/latest/

21. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: Displaying reference data information from over 60 000 exomes. Nucleic Acids Res. 2016;45(D1):D840–D845. https://doi.org/10.1093/nar/gkw971

22. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2014;43(D1):D789–D798. https://doi.org/10.1093/nar/gku1205

23. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2013;42(D1):D980–D985. https://doi.org/10.1093/nar/gkt1113