Communicating Air Quality with a Sense of Audience: What Do Survey and Textual Analyses Offer Science Dissemination Studies?

Andrew Mitchell, Mark Lemon, Gill Drew

andrew.mitchell@dmu.ac.uk

www.bioairnet.co.uk

Overview:

- The study – an introduction

Overview:

- The study – an introduction

- PGR survey – risks and exposure to BioPM non-specialist responses

Overview:

- The study – an introduction

- PGR survey – risks and exposure to BioPM non-specialist responses

- Text-mining BioAirNet T4 on-line workshop transcripts (March 21, 2023)

Overview:

- The study – an introduction

- PGR survey – risks and exposure to BioPM non-specialist responses

- Text-mining BioAirNet T4 on-line workshop transcripts (March 21, 2023)

- Corpus linguistic analysis of BioAirNet T4 on-line workshop transcripts

Overview:

- The study – an introduction

- PGR survey – risks and exposure to BioPM non-specialist responses

- Text-mining BioAirNet T4 on-line workshop transcripts (March 21, 2023)

- Corpus linguistic analysis of BioAirNet T4 on-line workshop transcripts

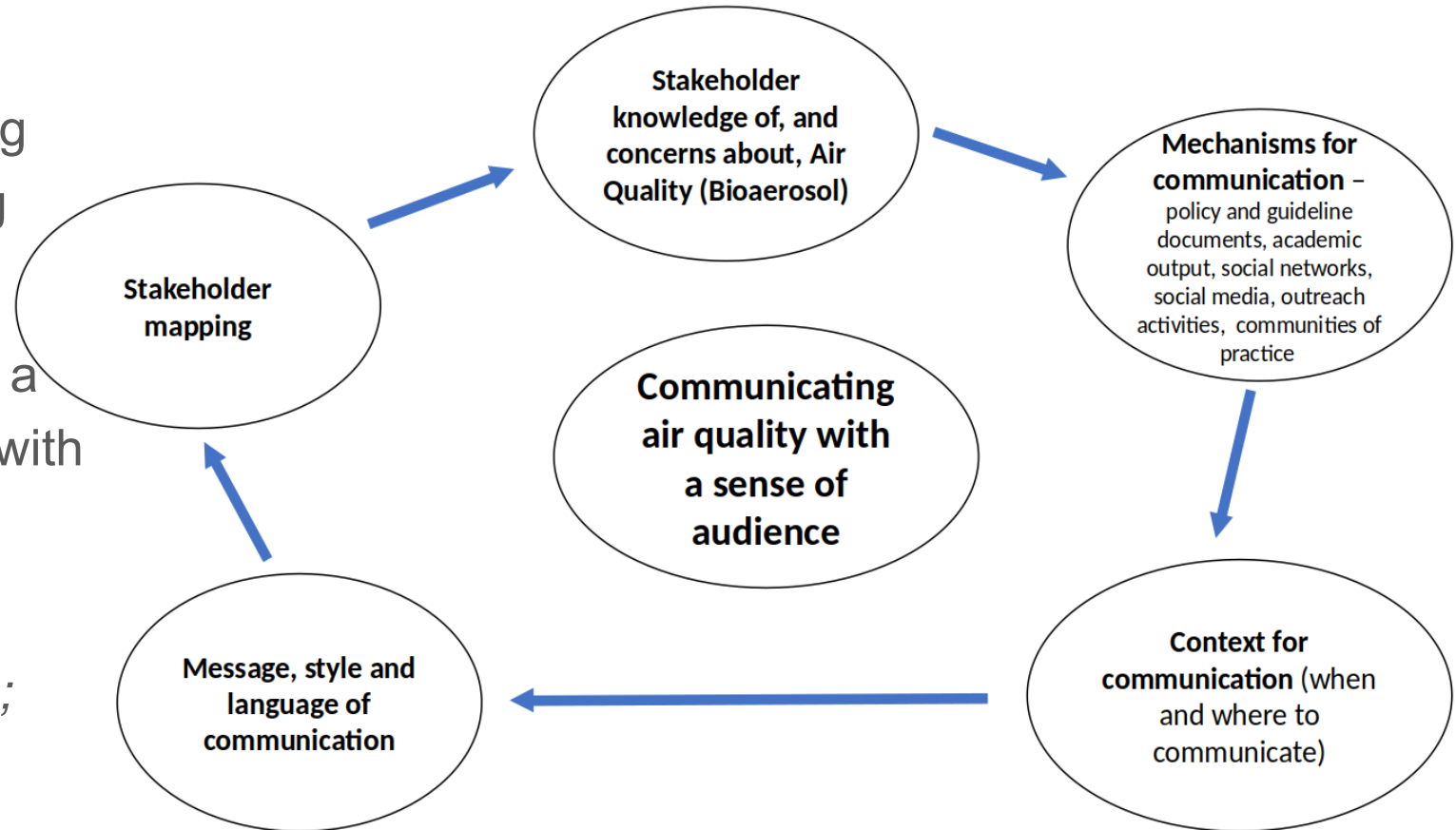- Thematic analysis of BioAirNet T4 on-line workshop transcripts

Overview:

- The study – an introduction

- PGR survey – risks and exposure to BioPM non-specialist responses

- Text-mining BioAirNet T4 on-line workshop transcripts (March 21, 2023)

- Corpus linguistic analysis of BioAirNet T4 on-line workshop transcripts

- Thematic analysis of BioAirNet T4 on-line workshop transcripts

- Conclusions

The study – an introduction

- Multi-modal analysis using an **early-stage** schematic model for describing processes to support communicating with a sense of audience

- The aim is to operationalise and test a five-phase model of communication with a sense of audience

- Empirically **explore three** of the five phases, viz. *Stakeholder Knowledge; Communication Mechanisms*; and *Communication Context.*

The study – an introduction (continued)

- To explore the *Stakeholder Knowledge* phase of the model, we examined a **survey** of over 900 responses to 145 questions. This provides insight into how **non-specialist** respondents' **perceive risks** and **environmental sources** of BioPM as a **proxy** for the general public.

- We also used **text-mining** to explore BioAirNet T4 on-line **workshop transcripts** (March 21, 2023)

- To explore the *Communication Mechanisms* and *Communication Context* phases of the model, we drew on **corpus linguistic** and **thematic analytic** methods and applied these to the workshop transcripts

PGR survey – risks and exposure to BioPM non-specialist responses

- Survey was developed and administered by one of Gill Drew's PhD candidates

- 145 questions – most as Likert scale 0 to 10

- > 900 responses

- This analysis did not use any biographical information entered by respondents

- Focus of analysis was on respondents' **perceptions** of **lethality risk** from BioPM, the **sources** of (potentially) lethal BioPM, and the **risks of exposure**

- Survey also asked after **trust** in and **knowledge of science** behind **air pollution** and **climate change**
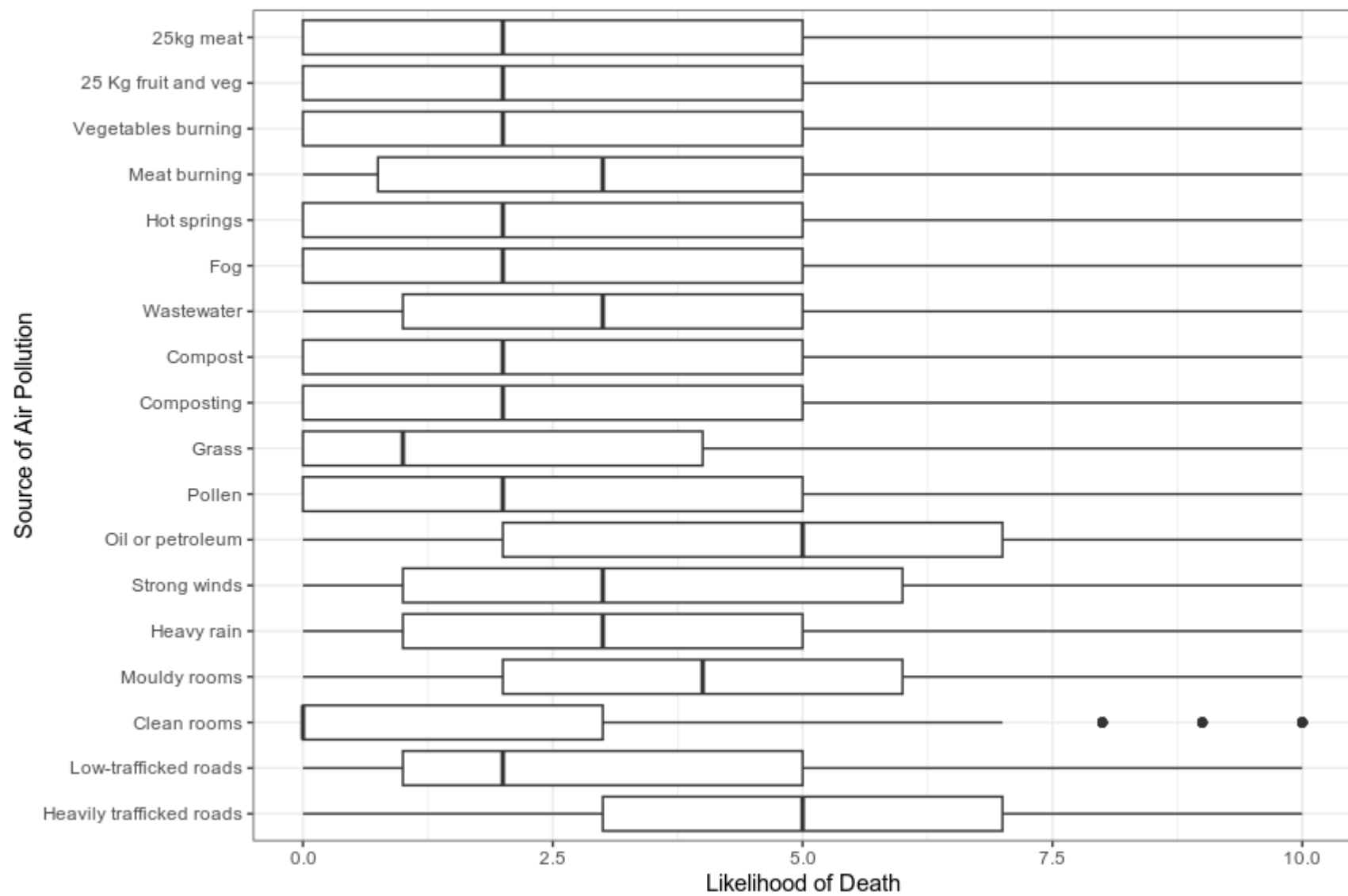
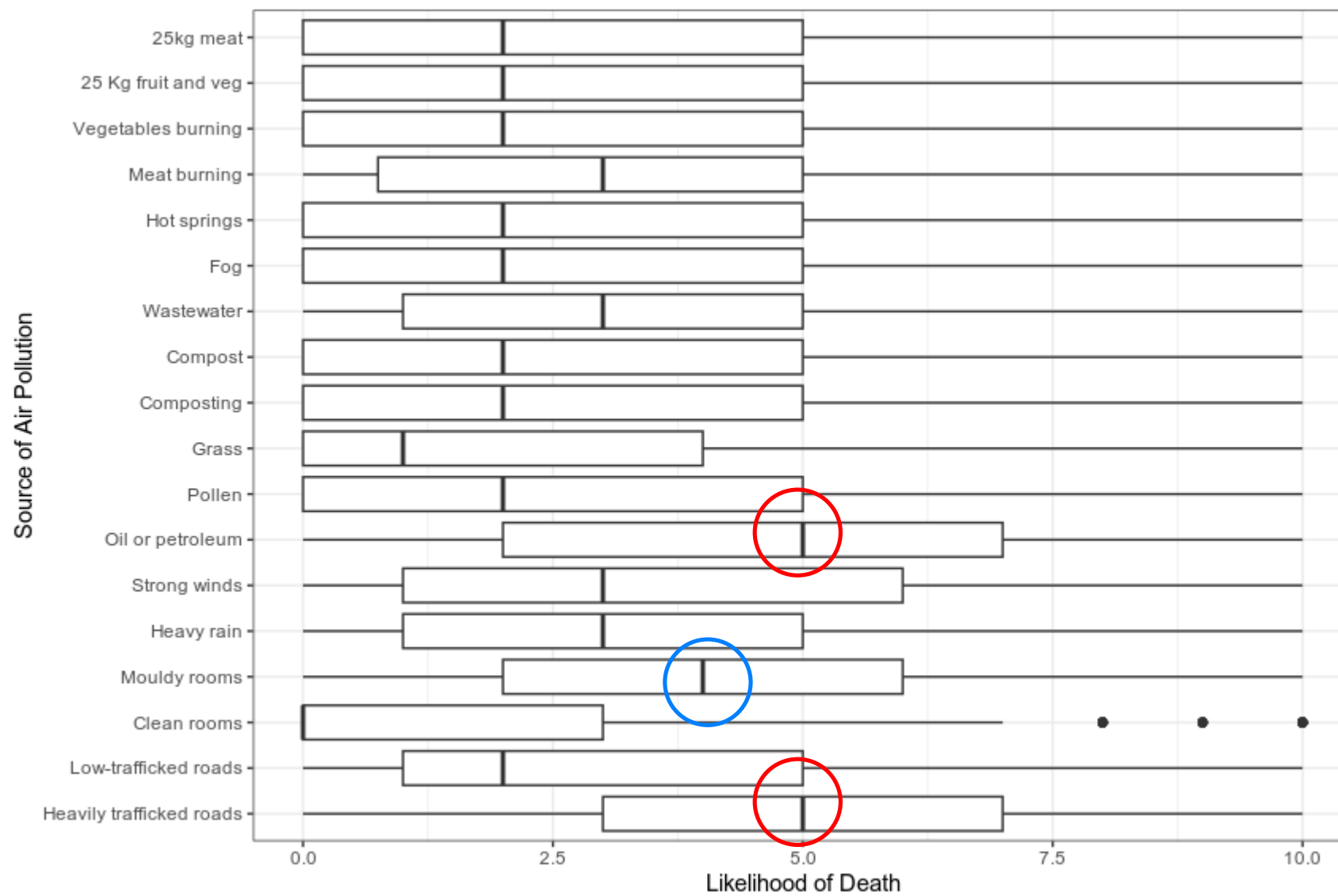Risk of Death by Threat Type
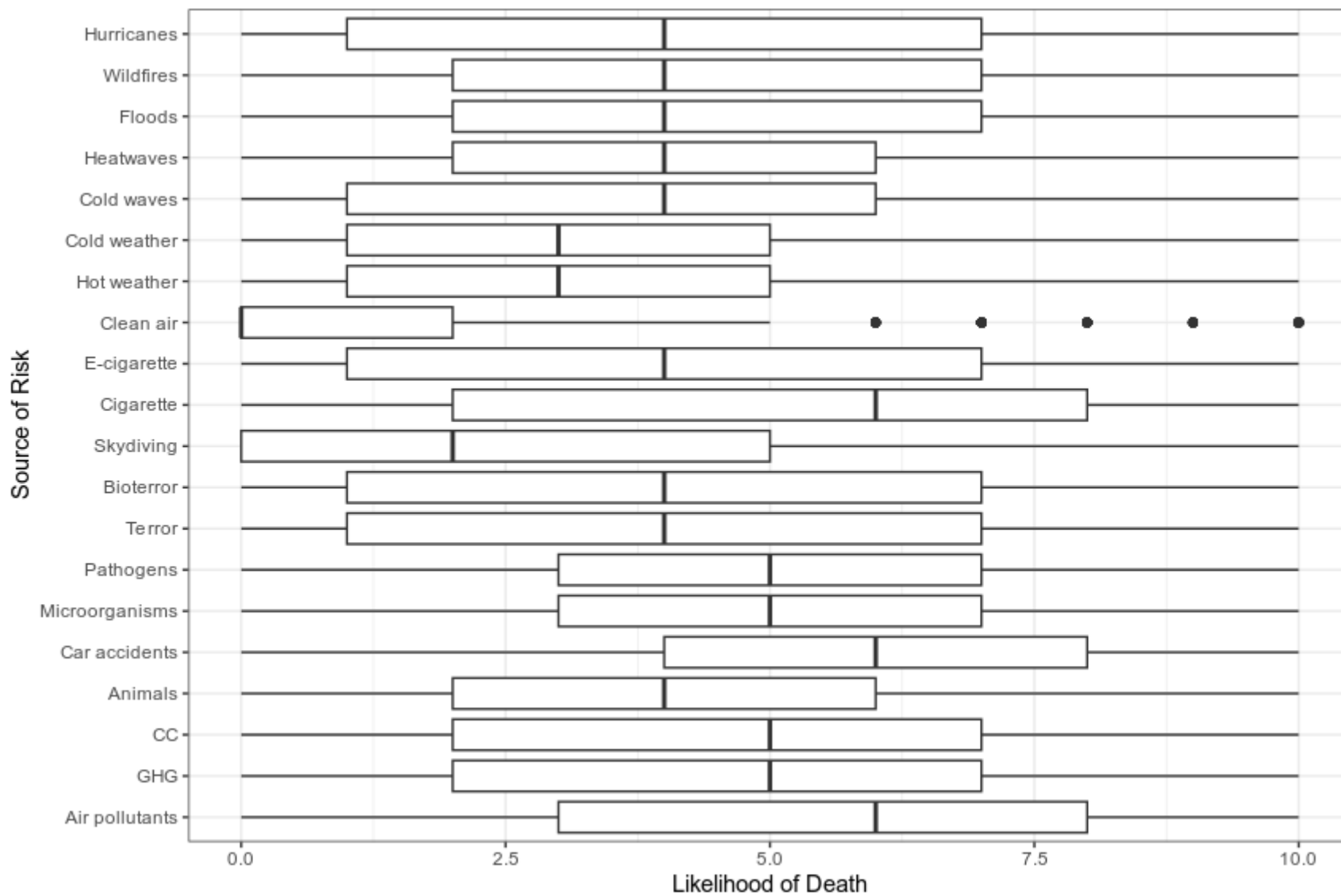
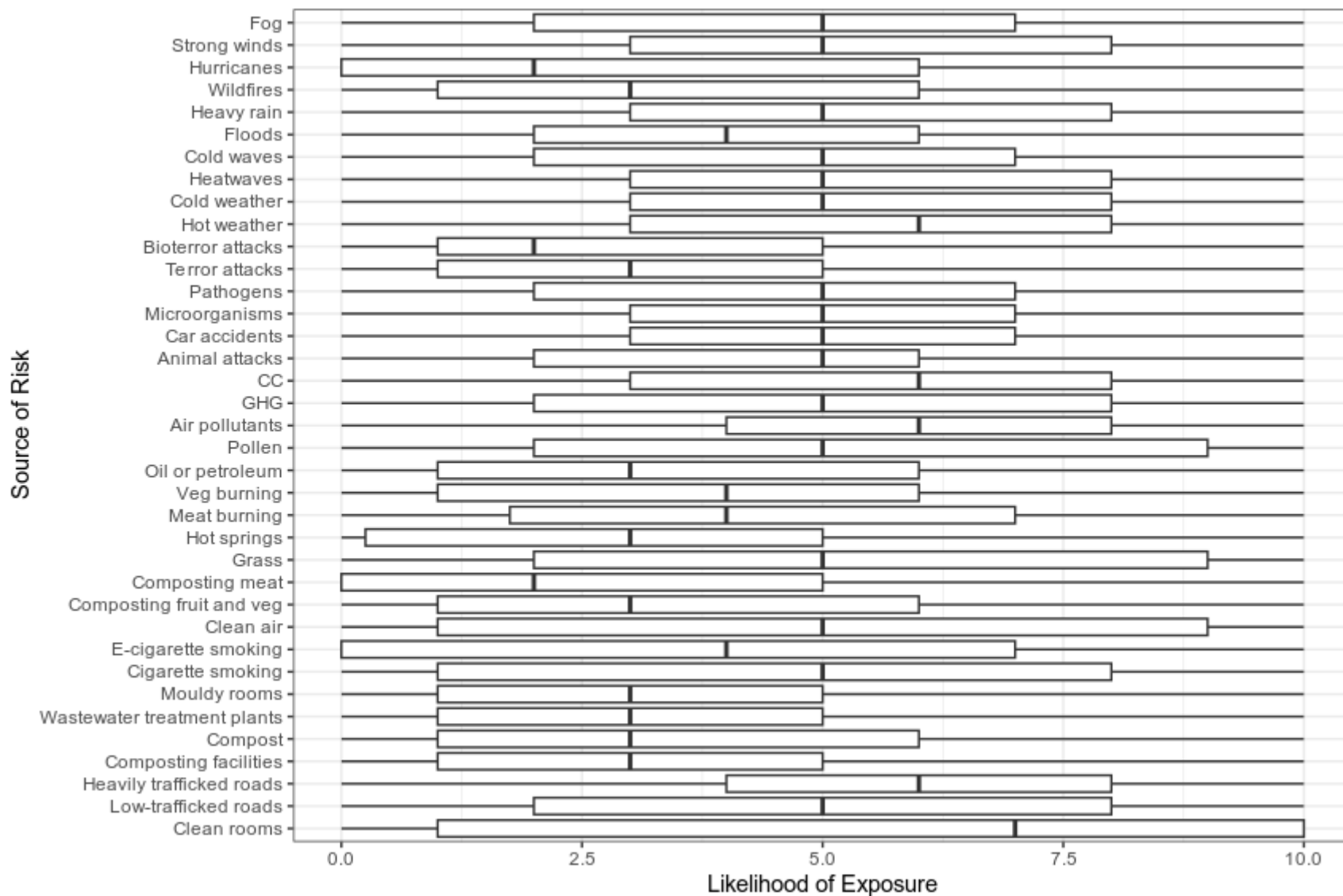**Risk of Death by Threat Type**
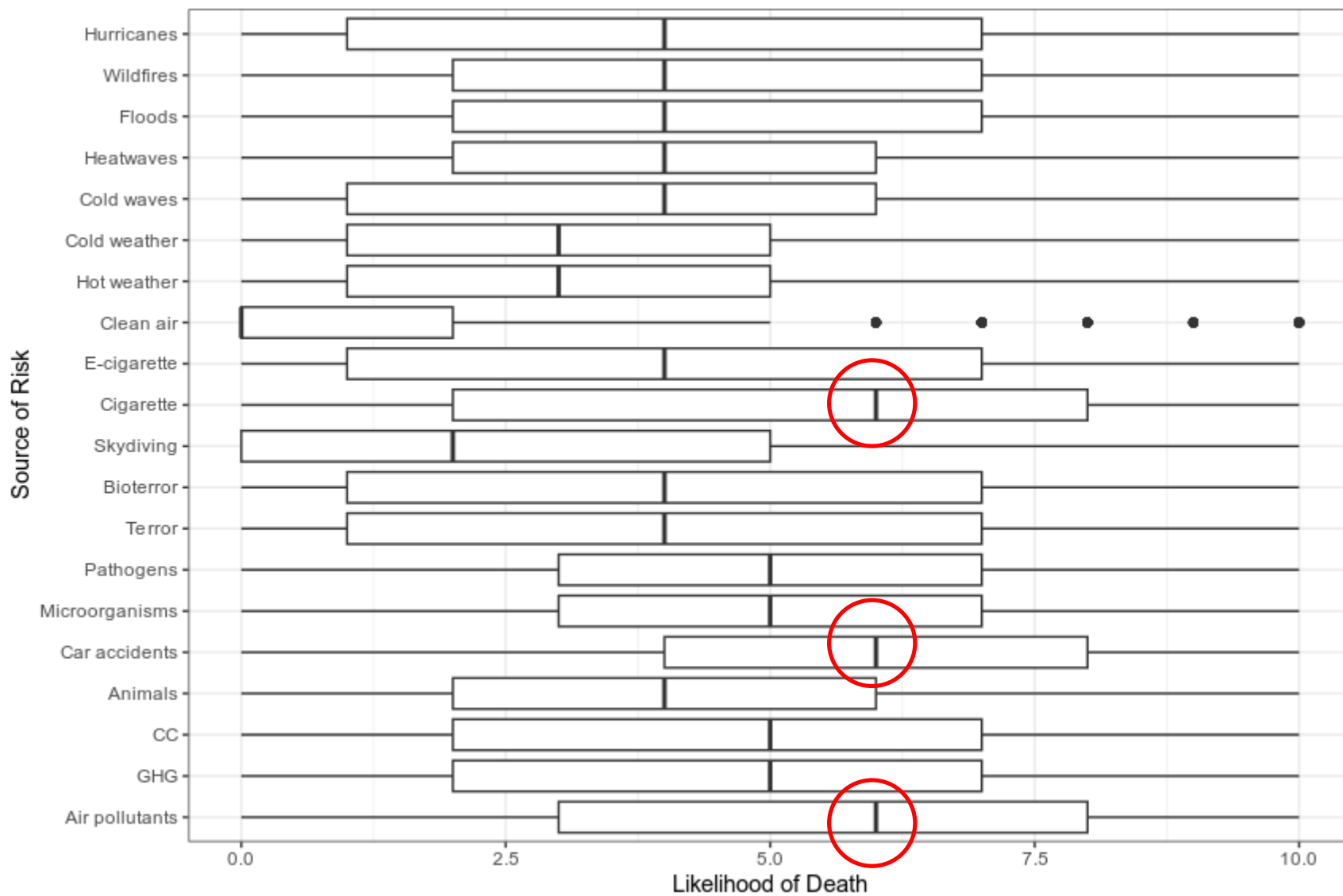
Sources of BioPM

Sources of BioPM

Risk of Exposure to BioPM

Risk of Exposure to BioPM
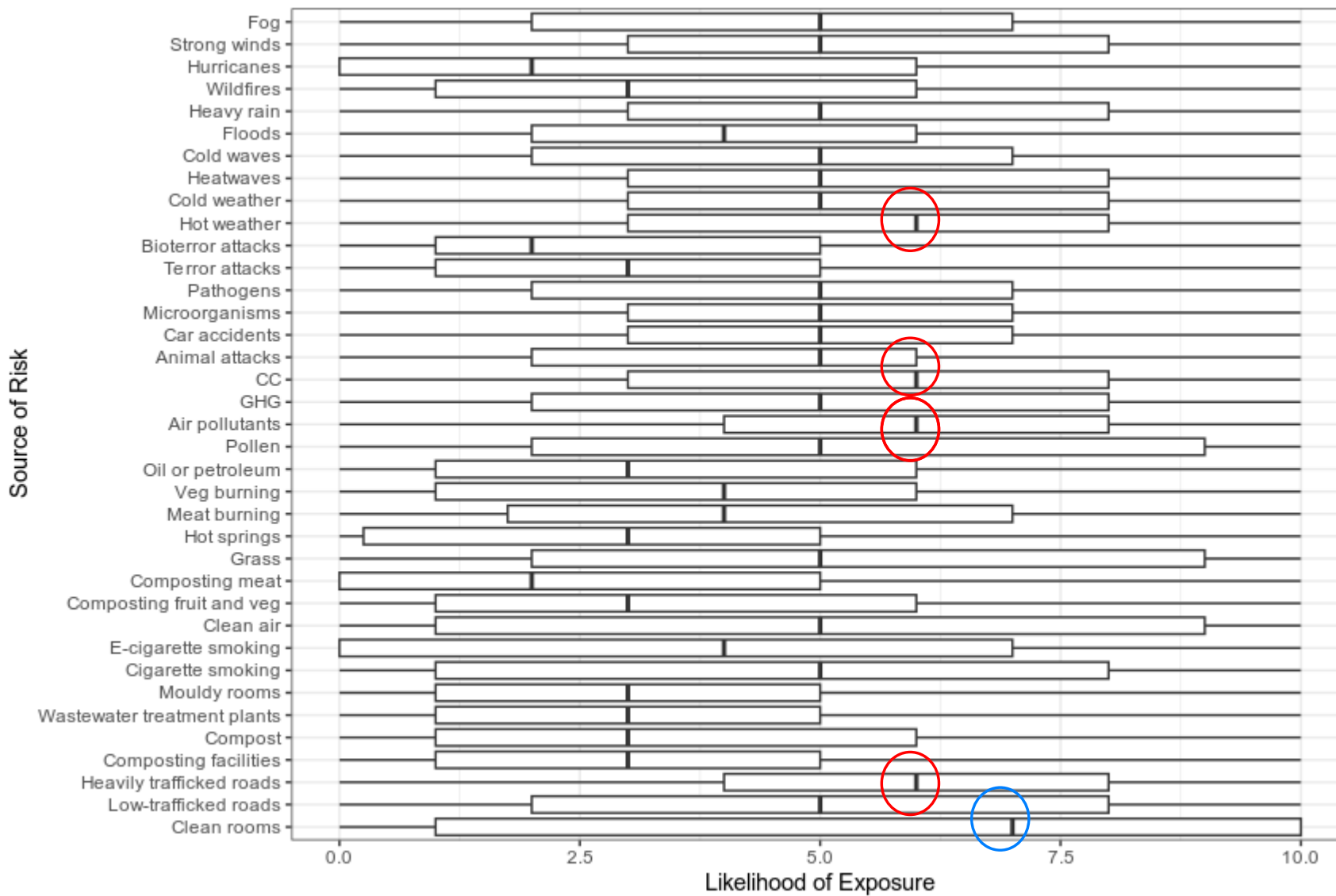
Trust in Science and Risks Associated with Air Pollution and Climate Change
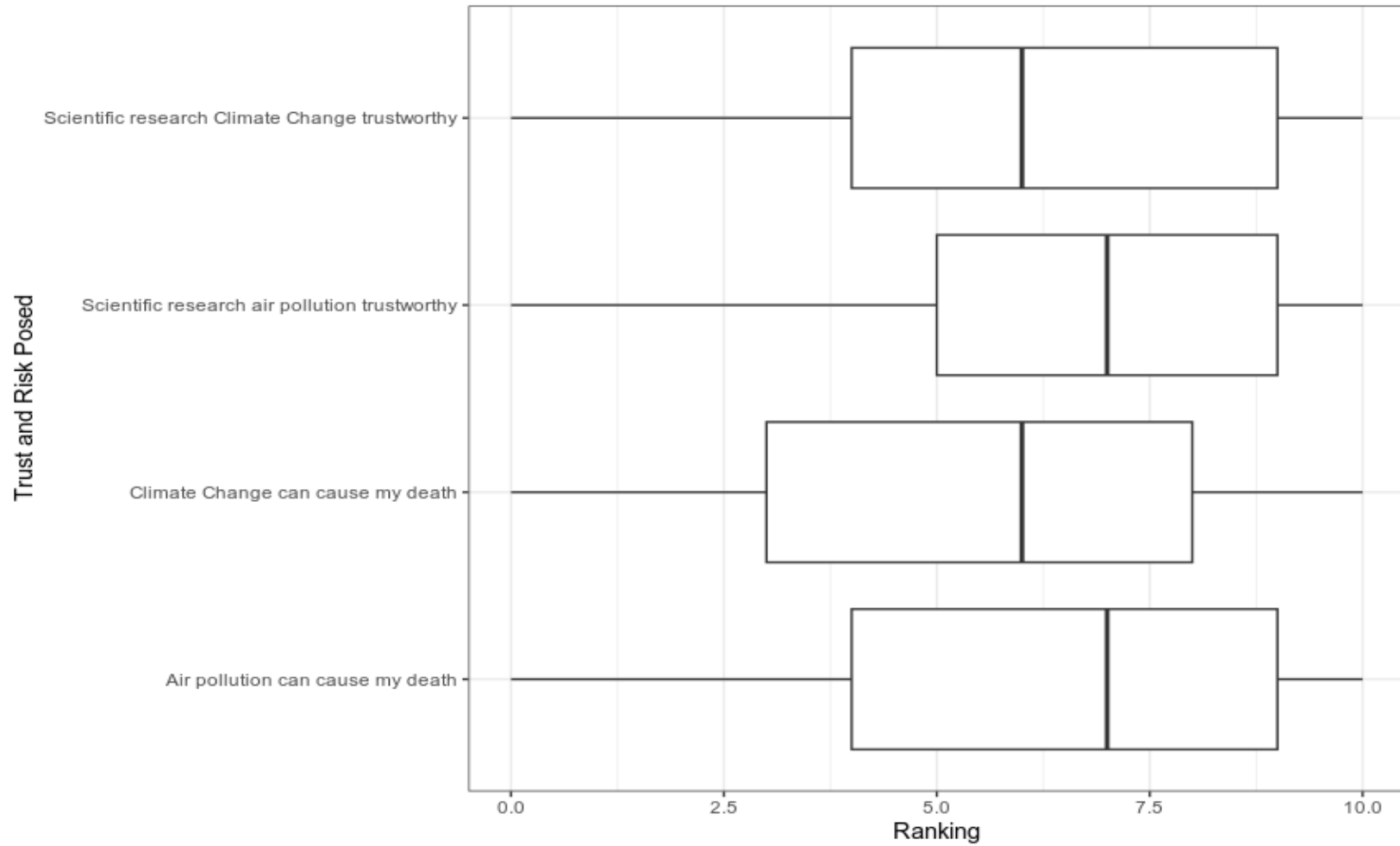
Lethality of Air Pollution, Trust in and Knowledge of Science
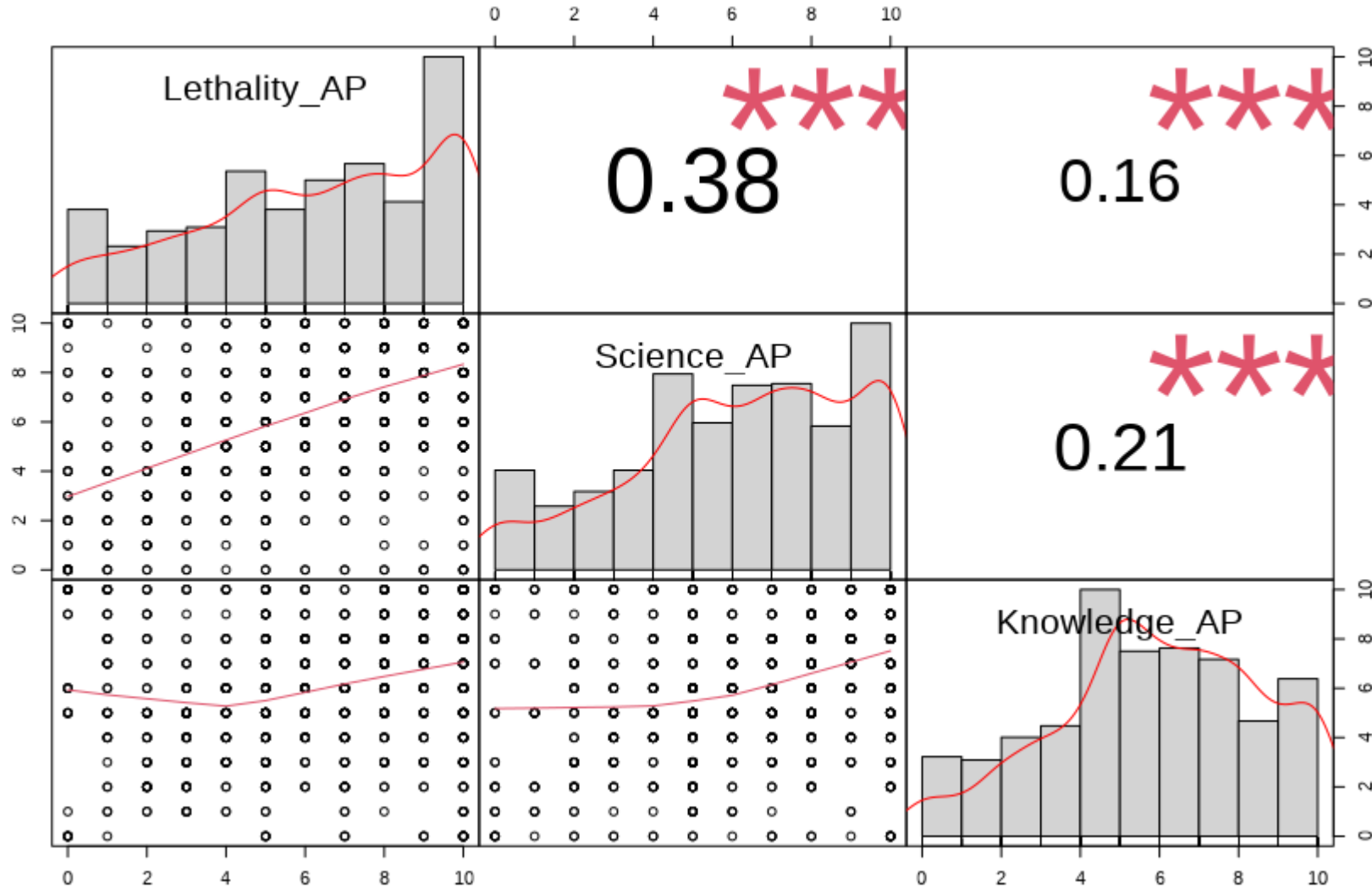
p < 0.001

Lethality of Climate Change, Trust in and Knowledge of Science
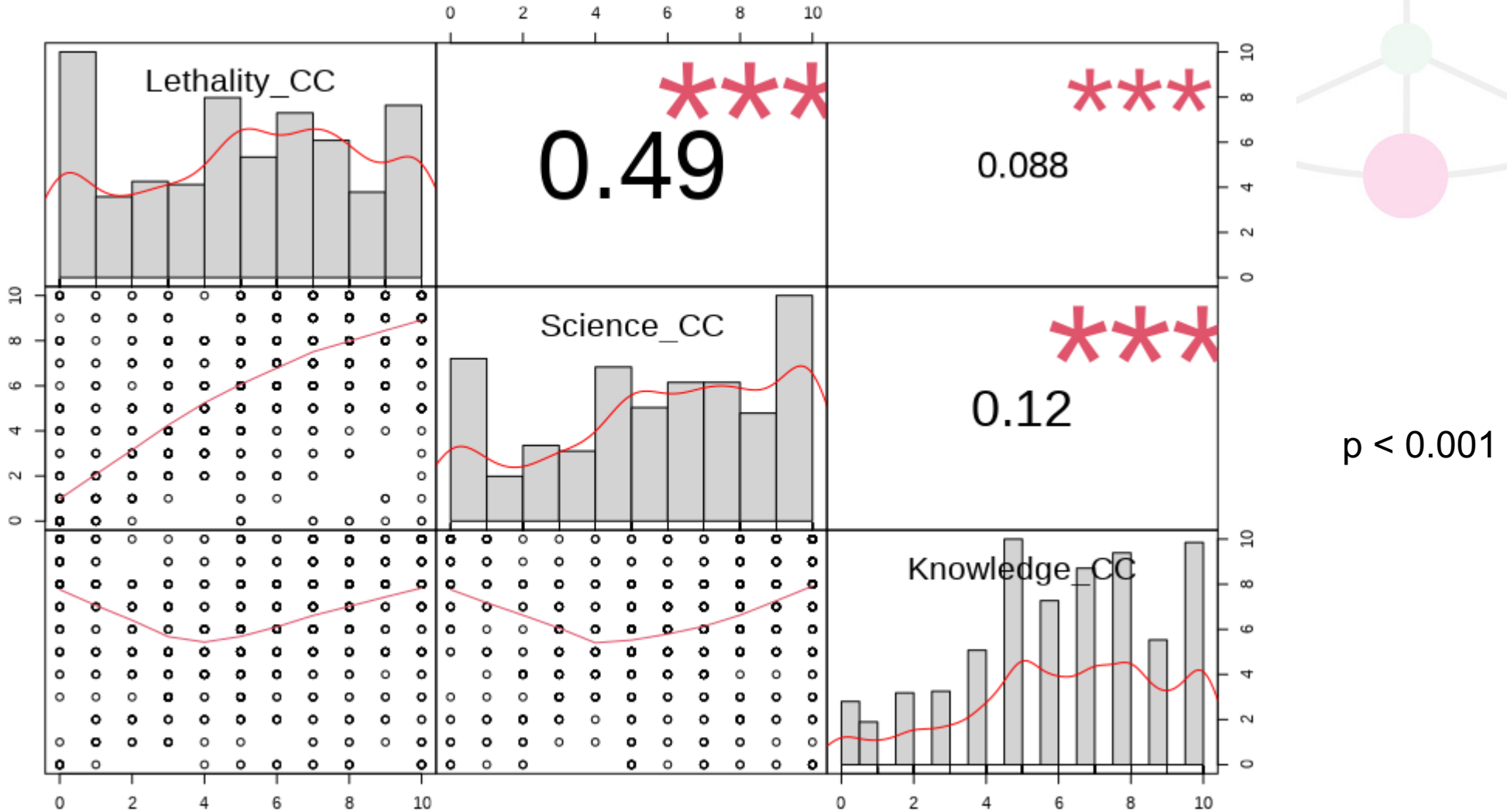
Text-mining BioAirNet T4 on-line workshop transcripts (March 21, 2023)

- Used four analytic methods:

    - TF-IDF,

    - cluster topic modelling,

    - LDA topic modelling, and

    - n-gram analysis

- Text is **unstructured data** and text-mining assumes corpus is a '**bag-of-words**' – semantic patterns are not preserved

- Only statistical patterns are revealed – operation is **syntactically 'blind'**

- Used *R* and packages '*tm*' and '*tidytext*' and **removed stop words** (e.g., 'and', 'the', etc.) which are statistically frequent but have low analytic value

Text-mining #1: TF-IDF

- Term Frequency – Inverse Document Frequency (TF-IDF)

- Calculates the **ratio** between **word frequency** and its **distribution** across the whole corpus

- **Balances** against high-frequency but low significance words

- Closer scores are to **zero**, the **more dispersed** they are and the **less significant** due to distribution

First twenty significant words in workshop transcripts using TF-IDF

Word	TF-IDF	Word	TF-IDF	Word	TF-IDF	Word	TF-IDF
workshop	0.0132	*activities*	0.0088	*aspergillus*	0.0084	*impacts*	0.0073
bioaerosols	0.0118	*raise*	0.0087	*video*	0.0084	*practitioners*	0.0073
thinking	0.0108	*raised*	0.0087	*bioaerosol*	0.0082	*trees*	0.0068
damp	0.0102	*training*	0.0087	*involved*	0.0080	*problem*	0.0068
public	0.0094	*year*	0.0085	*theme*	0.0073	*early*	0.0068

NB: TF-IDF scores closer to zero suggest greater dispersion and lower significance

Text-mining #2: Cluster analysis

- Used **dendrogram** word cluster with **k-means = 4** to reduce information to isolate salient descriptions about the corpus as a whole

- Hierarchical cluster analysis involves a **step-wise algorithm** merging two objects which have the **least dissimilarity** at each iteration

- *y-axis* is a measure of **relative distance** (dissimilarity) between individual data points

Topic Modelling #1 : k-means clustering (k = 4)

Cluster Dendrogram

d
hclust (*, "ward.D")
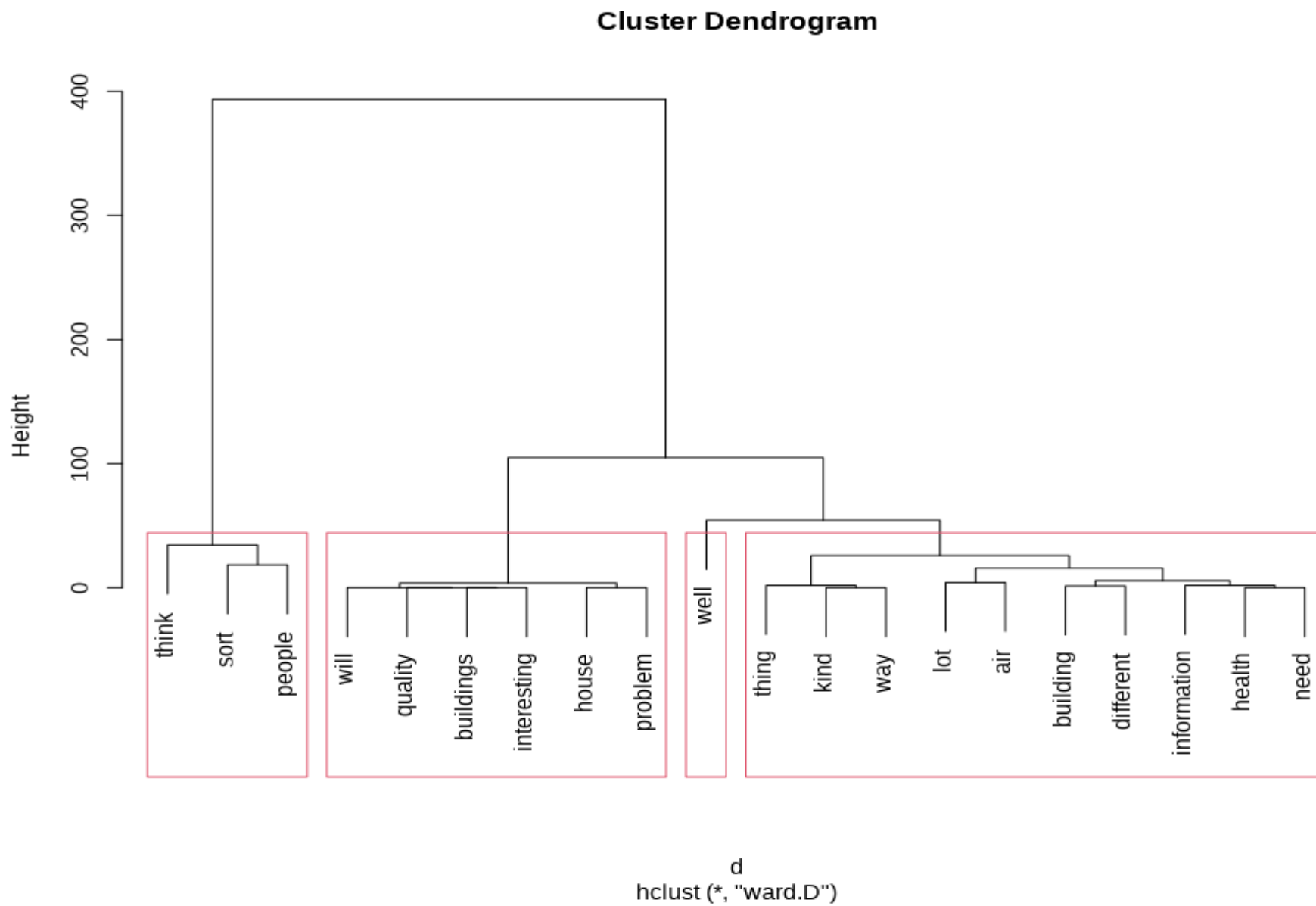
Text-mining #3: Latent Dirichlet Allocation (LDA)

- LDA treats each document as a multivariate '**mixture of topics**' and as a '**mixture of words**'

- It is an **unsupervised** Bayesian **probability-based** approach to revealing latent topics

- Observes **word frequency distribution** among documents to identify k (k = 4) **number of topics** and documents given the **probability for each topic**

- With k = 4, the text is **assigned four probabilities** for being part of a set of terms that suggest a common theme

Topic Modelling #2: LDA (k = 4)
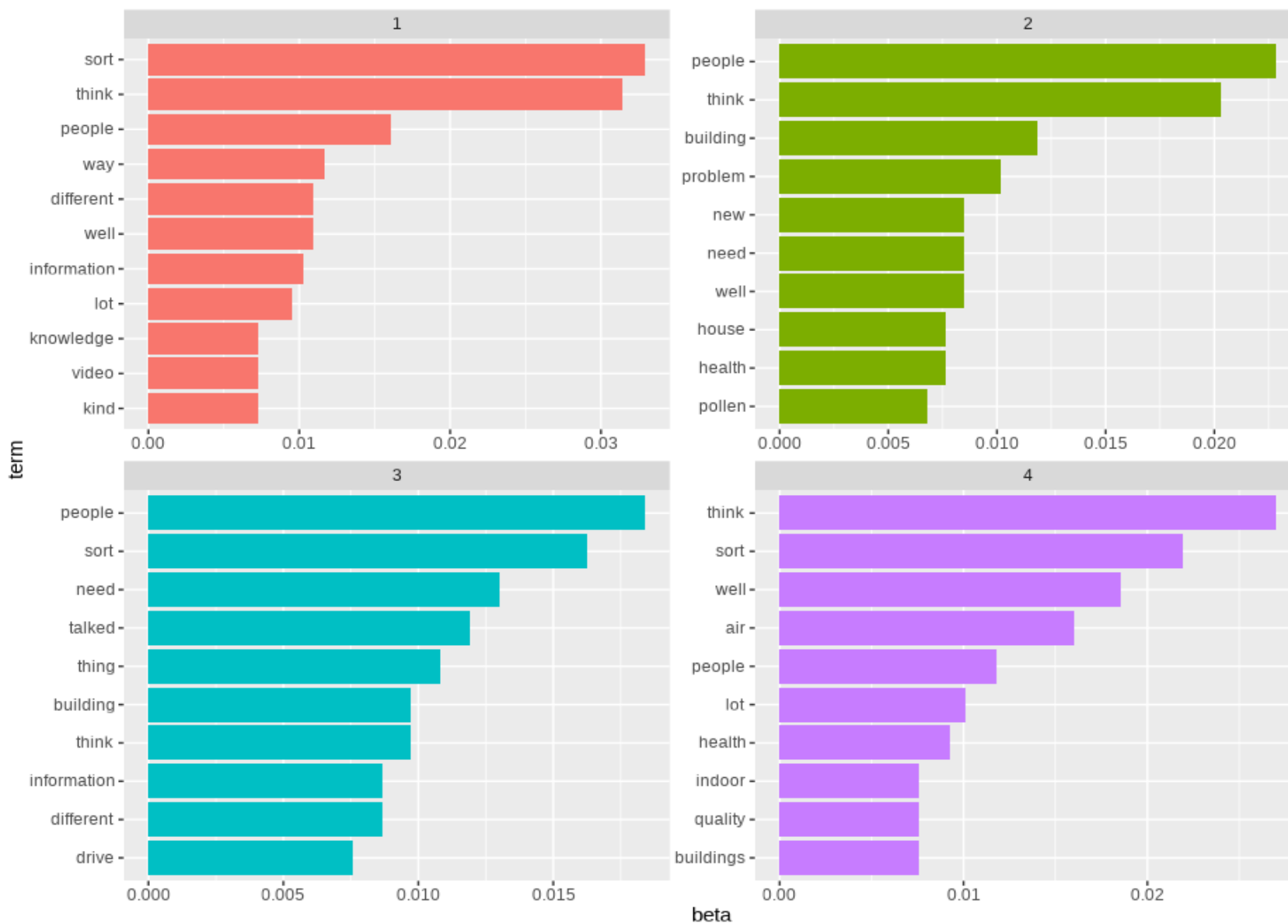
Text-mining #4: n-gram analysis

- *n-grams* are **statistically significant** patterns of **recurring** word forms

- *n-grams* are typically presented as uni-, bi-, or tri-grams, meaning one, two, or three word sequences, respectively

- Calculating n-grams gives insight into **common word pairings**, and in computational linguistics are often used to **anticipate** the probability of a word preceding another given word (e.g., 'air' commonly precedes 'pollution')

Bi-grams (arrow shows directionality of pairing)
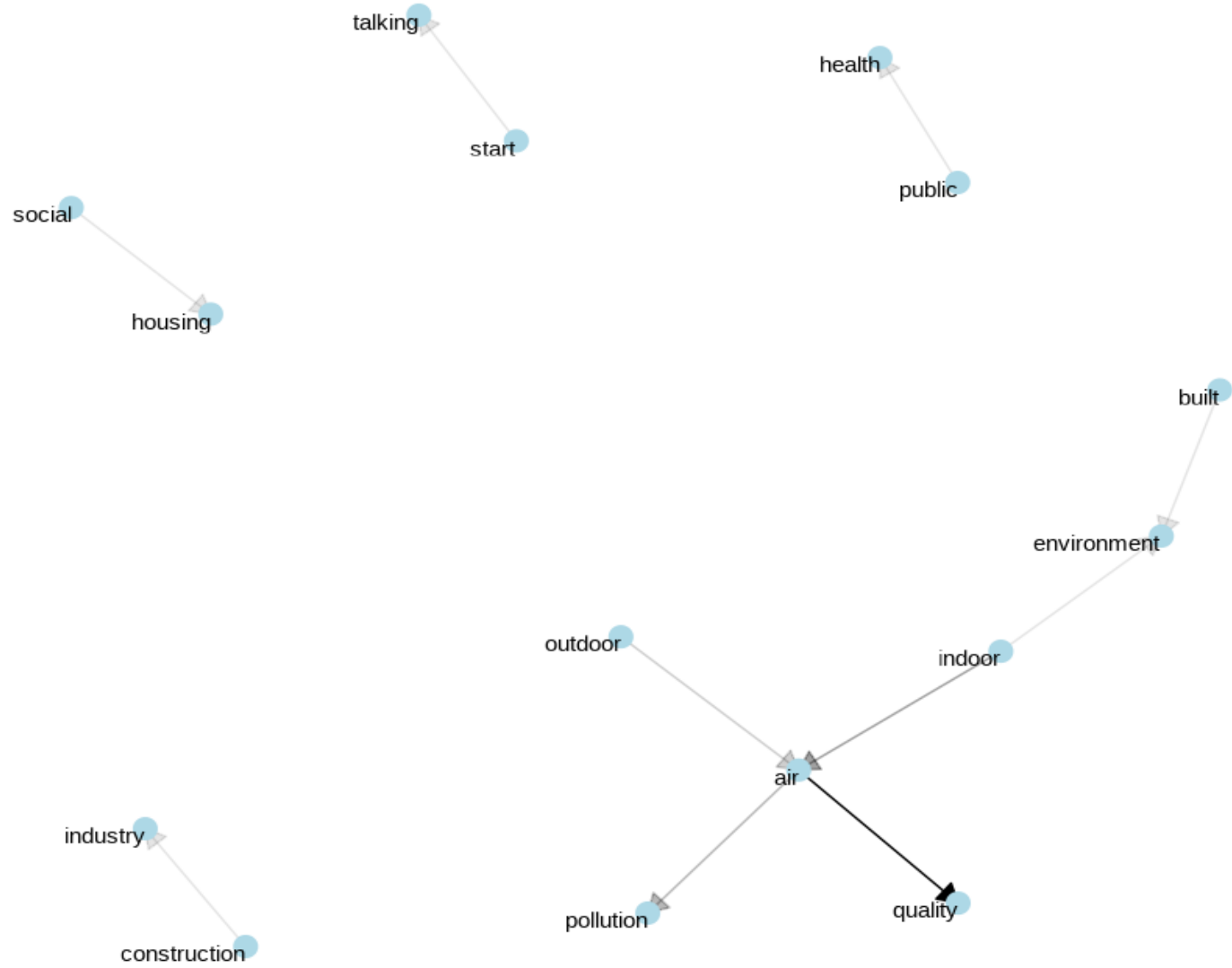
Corpus linguistic analysis of BioAirNet T4 on-line workshop transcripts

- Corpus linguistics is a **mixed methods** approach to studying large volumes of text – a corpus (plural 'corpora', i.e., a 'body')

- Draws on a range of **quantitative techniques** (e.g., collocates – word associations/ pairings; word frequencies, distributions, and 'keyness' analysis)

- However, also requires the analyst to **make judgments** and **interpretations** about results, and this often requires analysing in detail randomly selected **concordance lines** to determine **meanings** and **functions** of key terms in the context of the corpus
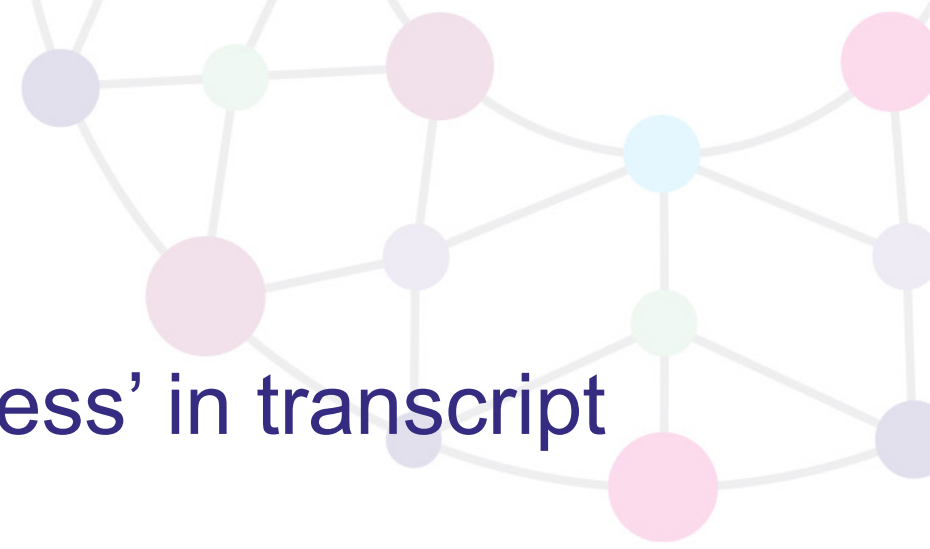
Corpus linguistics #1: Identifying 'keyness' in transcript data

- 'Keyness' refers to the key terms that are comparatively **more significant** to the focal corpus than to a general reference corpus of language usage

- The workshop transcript corpus was compared to the **Spoken British National Corpus** (SBNC) which comprises **11.5-million words** of transcribed conversations recorded by native English speakers across the UK between 2012 and 2016

- Returns a statistically relevant '**profile**' of the corpus or its 'aboutness'
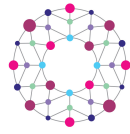
Key Words Significant to Transcript Relative to Spoken BNC

Type	BioAirNet Freq	BioAirNet Disp	SBNC Freq	SBNC Disp	Statistic
health	30	0.38	556	4.32	15.61
information	31	0.42	609	2.77	15.59
air	35	0.69	831	2.72	15.44
buildings	20	0.59	176	5.4	13.95
guidance	17	0.19	20	10.8	13.75
building	33	0.5	1053	3.2	13.04
pollen	16	0.68	21	12.76	12.98
impact	16	0.62	113	5.91	11.95
industry	17	0.59	226	4.58	11.51
indoor	14	0.94	22	9.68	11.47

Type	BioAirNet Freq	BioAirNet Disp	SBNC Freq	SBNC Disp	Statistic
user	14	0.31	24	9.97	11.45
quality	20	0.5	470	2.77	11.23
knowledge	16	0.79	222	3.94	10.92
trees	18	0.72	425	3.55	10.49
construction	13	0.67	52	8.58	10.43
bioaerosols	12	1.07	0	0	10.18
communicate	13	0.59	117	5	9.84
outdoor	12	0.62	55	7.55	9.67
sources	11	0.54	41	8.28	9.06
network	12	0.85	148	6.91	8.92

Corpus linguistics #2: Concordancing

- Concordance analysis is the study of the **key word in context** (KWIC)

- The node/ focal term is located within a **context** of words that **precede** and words that **follow** – the span of these (the 'window') is set by the analyst

- Facilitates **insight** into how words are **used** in context and how they **function** in the sentence and what they **mean** through use.

- **First 5 words** from the key terms relevant to the workshop data are examined – i.e., '**health**,' '**information**,' '**air**,' '**buildings**,' and '**guidance**.'

Concordances for Key Word 'health'

	Left	Node	Right
1	a lot of work around comfortable buildings,	**health**	and well being in the indoor environment,
2	in the impacts of pollen on human	**health.**	When I worked in an allergy clinic
3	and it improving our mental and physical	**health**	but in fact, for 18 million people
4	deeply. We see a lot of the	**health**	impacts here. So consequently we have a
5	if you like, on the sort of	**health**	and wellbeing front when it comes to
6	homeowners around sort of their understanding of	**health**	and wellbeing inside their home. How much
7	a price on a home which promoted	**health**	and well being, etc and actually that
8	more for a building which promoted the	**health**	and well being of themselves and their
9	it's grown a bit I think certainly	**health**	and well being in the built environment
10	study in Modern Methods of Construction, and	**health**	and well being is one of the
11	be the final impact on people's mental	**health**	at the end of this, you know,
12	what we do in relation to human	**health.**	Very little is known about human health.
13	sources and the effects and then the	**health**	on how do we communicate and where,
14	people understand the impact across environments, social,	**health,**	and monetary at that level. And so
15	when we've developed a sort of a	**health**	and well being sensor, or if you
16	of the UK that cause all these	**health**	problems. In general, the birch being one
17	sort of an argument of improving public	**health**	outcome. So, I think. So that's why
18	mean we, when we start talking comfort	**health**	and well being and buildings. If we
19	know, so maybe it's realising what the	**health**	impacts are going to be for the
20	of this evidence we were pushing this	**health**	and well being piece for buildings for

Concordances for Key Word 'information'

	Left	Node	Right
1	it, it's only part of the marketing	**information**	but I don't know if you have
2	grips with I mean there's lots of	**information**	out there in other studies around the
3	is the other argument. What extra	**information**	do we need to be putting in
4	I think there is a lack of	**information**	and clarity of course I think that's
5	pollution, which accompanies that quite nicely. The	**information**	is now there from the medical side.
6	the problems or whether the lack lacking	**information**	is at the moment or lacking pathways
7	to have new mechanisms and translate that	**information**	in a way that makes sense to
8	happy to give them with all the	**information**	that they need all the evidence I've
9	before they said there's a lot more	**information**	around just how to get hold of
10	or something and we can just add	**information**	into different sectors and say oh this
11	how best we can align this available	**information**	and knowledge to the policy and to
12	say providing a shared platform to share	**information**	or discussion board, and we honestly have
13	way that video is sort of an	**information**	medium it's not sort of an entertaining
14	very good way of communicating, very complex	**information**	in a simple way to variety of
15	it's got to be a mixture of	**information,**	I think you've got some clear factual
16	actually even start with just targeting the	**information**	you know it's just boiling it down,
17	in terms of getting that all the	**information**	you've got into guidance so what's the
18	the last seven years gathered so much	**information**	and evidence now, and it's having an
19	about this yes I've got all this	**information.**	But how do I influence the builder,
20	to them and people, but but sharing	**information**	if somebody if we invite people from
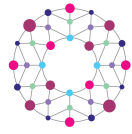
Concordances for Key Word 'air'

	Left	Node	Right
1	the tension between outdoor air pollution indoor	air	pollution, particularly is of concern to us.
2	and have some background in transport and	air	quality, but also work in sustainable building
3	so thermal comfort, acoustics, daylight and indoor	air	quality. So we're quite keen to understand
4	what is out there in the outdoor	air,	with regard to bio PM, we've done
5	Aspergillus is commonly found in in the	air	and it's perfectly sized to be breathed
6	in how the Aspergillus is in the	air	is impacting our patients in much the
7	and people are quite well aware of	air	quality, but somehow this biological part of
8	did you actually when you look at	air	quality do you look at the bioaerosols
9	would advocate for as a business, indoor	air	quality was a key topic in there.
10	you're building in a location with poor	air	quality, We tend to always when talking
11	various things that are in the outdoor	air	but I would say more geared around,
12	from our construction materials into the indoor	air	from the air we know about paints.
13	materials into the indoor air from the	air	we know about paints. What about the
14	If we take thermal comfort, or even	air	quality and start talking TVOCs, VOCs, use
15	the tension between the indoor and outdoor	air	quality. We looked at Massive Datasets we
16	about the minutiae of what's causing indoor	air	problems, concentrate on damp, get rid of
17	publication, and you've now got the NIC	air	pollution including I think it included indoor
18	pollution including I think it included indoor	air	pollution, which accompanies that quite nicely. The
19	building or not sealing the building, letting	air	in when we've don't know the outdoor
20	like so it's looking at temperature, acoustics,	air	quality and light. And we did a

Concordances for Key Word 'buildings'

	Left	Node	Right
1	had sources I think a lot of	**building**	systems, particularly are of concern. But the
2	air quality, but also work in sustainable	**building**	so that cannot link got me on
3	in the development of preinsulated timber frame	**buildings,**	steel frame concrete, whatever it may well
4	do a lot of work around comfortable	**buildings,**	health and well being in the indoor
5	the indoor environment, how do we design	**buildings**	to avoid that. I think looking at
6	interested in the green infrastructure around new	**buildings,**	because we're all very interested in the
7	you know better how the sort of	**buildings**	play a role in terms of exposure,
8	infrastructure or carbon reduction strategy energy efficient	**buildings,**	how that will impact on exposure to
9	sort of standard for the kind of	**buildings**	we would advocate for as a business,
10	got a nicely airtight building, if you're	**building**	in a location with poor air quality,
11	in our guidance around how we build	**buildings**	to to mitigate these further impacts ultimately
12	know, whether it is a client for	**building**	and that's using materials or is trying
13	you know, how people behave inside a	**building,**	it's not just what is built, you
14	and wellbeing front when it comes to	**buildings,**	obviously we've talked about sustainability and energy
15	odd percent would pay more for a	**building**	which promoted the health and well being
16	saying that they're the employees and sick	**building**	syndrome and that type of thing. Whereas
17	building down to a price and they're	**building**	down to a profit margin, and ventilation,
18	to change up the basic regulation of	**building**	construction materials and building, from what I've
19	talking comfort health and well being and	**buildings.**	If we take thermal comfort, or even
20	to bring the quality of what we're	**building,**	and especially working on converting older homes

Concordances for Key Word 'guidance'

	Left	Node	Right
1	we need to be putting in our	**guidance**	around how we build buildings to to
2	into it's actually very difficult to find	**guidance**	on it. Why do you think that
3	to feel it falls under the tenants	**guidance**	to actually maintain it and keep on
4	CIEEM ecology body, the UK in the	**guidance**	that they give to developers on new
5	some involvement. To begin, structuring their their	**guidance**	that they give for planting around schools
6	could potentially speak to, to see what	**guidance**	there already is, if any, or if
7	theme help make that link. I think	**guidance**	would be useful. Well appraised guidance not
8	Think guidance would be useful. Well appraised	**guidance**	not something you pick off Wikipedia, for
9	that what be in the form of	**guidance,**	are the pathways that are big sort
10	that all the information you've got into	**guidance**	so what's the link there that we
11	groups that are going to be giving	**guidance**	to the rest of the industry as
12	trees, but how to get this into	**guidance**	or legislation or tell the people who
13	change is it legislation or is it	**guidance,**	or actually one point was made, is
14	certainly found that in, even in legislation	**guidance**	that's available now, bioaerosols often are not
15	developed quite a lot of tools or	**guidance**	with the funding restrictive shrinking right there's

Thematic analysis of BioAirNet T4 on-line workshop transcripts

- Last of the analytic methods used

- **Counter-balances** statistical approaches with a **closer reading** of transcripts

- Thematic analysis involves the development of **codes** using the words and emerging patterns as the basis – a **bottom-up** approach

- Following several passes over the text, codes are **assembled** into common clusters or **themes** to show higher-level abstractions of **latent meanings** and **discourses**

- Here **five themes** were assembled from the coded text

Thematic analysis #1: 'Communication' theme

[1] "when you start talking about bioaerosols and people are quite well aware of air quality, but somehow this biological part of the air quality always goes on a back burner"
(Participant in the first Breakout group)

[2] "my response was to simplify for our patients, and not worry about the minutiae of what's causing indoor air problems"
(Participant in the second Breakout group)

[3] "an interesting way of trying to communicate what you've actually got to people, people don't understand the problem that they have" (Participant in the second Breakout group)

Thematic analysis #2: 'Gaps' theme (a)

[4] "there's lots of information out there in other studies around the impacts of various things that are in the outdoor air but I would say more geared around, man-made activity or human activity which was causing this rather than the bio side"
(Participant in the first Breakout group)

[5] "in this country, from the base up, the UK is failing to regulate is failing, as, as has come into sharp focus, after the tragic fire in London. There's no accountability."
(Participant in the second Breakout group)

Thematic analysis #2: 'Gaps' theme (b)

[6] "there's a bit of an ownership issue over this problem as well"
(Participant in the second Breakout group)

[7] "it's actually getting to the people that influenced the landscape architects, it's, you know, actually getting to the people that actually say to the builders well this is what you need to do." (Participant in the third Breakout group)

[8] "even in legislation guidance that's available now, bioaerosols often are not really considered"
(Participant in the Plenary)

Thematic analysis #2: 'Inside/ Outside tensions' theme (a)

[9] "especially what's happening at the moment with COVID and people kept inside, and the tension between outdoor air pollution indoor air pollution, particularly is of concern to us."
(Participant in the first Breakout group)

[10] "we've done a lot of work on diesel car fumes, etc. But how do we mitigate that. How do they come into the indoor environment, how do we design buildings to avoid that."
(Participant in the first Breakout group)

Thematic analysis #2: 'Inside/ Outside tensions' theme (b)

[11] "for 18 million people in this country who've got pollen allergies, planting the wrong sorts of plant in a specific area where lots of people, especially around schools and hospitals and things like that can actually be detrimental."
(Participant in the first Breakout group)

[12] "We're going to shut our homes up even more tightly, and we'll have another tidal wave of patients coming to our door."
(Participant in the first Breakout group)

[13] "there's a pollen tree over there, and the pollens coming in your house and causing this person to be ill."
(Participant in the third Breakout group)

Thematic analysis #4: 'Research needed' theme (a)

[14] "so little research being done about how pollens transfer from the outside to the indoor environment. And I think that's an area that would, you know, really benefit from more research"
(Participant in the first Breakout group)

[15] "less research has been done around *Aspergillus*. What research has been done generally is done on visual identification of spores so they can't separate a *Penicillium* from *Aspergillus* so there's even less done than it appears."
(Participant in the first Breakout group)

Thematic analysis #4: 'Research needed' theme (b)

[16] "the emissions from our construction materials into the indoor air from the air we know about paints. What about the other stuff that we manufacture, gypsum block, etc. But equally what's coming out of our manufacturing plants as well."
(Participant in the first Breakout group)

[17] "people make decisions incrementally they choose paint or floor or tiles, you know, nobody's modelling a system, you know the decisions are not being made through a system. So, we have to we have to help people understand the impact across environments, social, health, and monetary at that level."
(Participant in the second Breakout group)

Thematic analysis #1: 'Tools' theme (a)

[18] "we've developed a sort of a health and well-being sensor, or if you like a comfort sensor if you like so it's looking at temperature, acoustics, air quality and light"
(Participant in the second Breakout group)

[19] "which will allow someone to appraise their home properly, whether they want to do the work themselves or whether they want to get a professional"
(Participant in the second Breakout group)

[20] "what we really need is a tool for a normal person just to monitor what's happening in their house to be told this is your problem and then, then what do you do about it."
(Participant in the second Breakout group)

Thematic analysis #1: 'Tools' theme (b)

[21] "if this group could act as almost like a go between or a sort of to use an old fashioned term a telephone directory of people who can who you can contact to discuss particular themes if you have a problem, or do it through a theme." (Participant in the third Breakout group)

[22] "If the goal of this is to communicate existing and emerging knowledge to various stakeholders. I feel like the medium of a short video is always very very effective." (Participant in the third Breakout group)

Conclusions

- The aim was to operationalise and test a five-phase model of communication with a sense of audience

- We mapped findings from the survey and workshop transcripts to three of these phases, viz. *Stakeholder Knowledge*, *Mechanisms of Communication*, and *Context of Communication*

- We applied a multi-modal analysis to the workshop transcripts, deploying text-mining, corpus linguistics and thematic analysis

Conclusions: Some notable findings (a)

- Using survey data we analysed:

  - the severity of risk for potential lethality from BioPMs

  - the perceived sources of threat

  - the perceived likelihood of being exposed to these.

- Respondents identified cigarettes, car accidents and air pollutants as highest **risk of death**

- Respondents rated heavily trafficked roads and oil and petroleum refineries as **sources** of high health risk

- Respondents thought that they were **most likely** to be exposed to hot weather, climate change, air pollutants, heavily trafficked roads, and anomalously to clean rooms

Conclusions: Some notable findings (b)

- Cluster analysis **grouped together** terms such as '*house*', '*building*', '*problems*', and '*quality*', and another cluster grouped words like '*health*', '*buildings*', '*air*', '*need*', and '*information*'. These clusters **suggest concerns** about the **built environment** relative to **air quality** and **wellbeing**, even if the origin and content of the original transcript was unknown. Insights **confirmed** by LDA and common word pairings through n-gram analysis

- Concordance analysis for the first 5 of these key words showed how these words are used and demonstrated that references to '*health*' include public health (as a policy mechanism), but also to well being, to buildings as vectors for health and illness, and also to mental and physical health, and so on.

Conclusions: Some notable findings (c)

- References to '*information*' included the suggestion for a **video** for education. There are also several references to the **people to influence** (builders, different sectors, etc.), and to **where** information is available (e.g., medicine).

- A need was identified to talk about indoor air quality and the **impacts** of chemical and biological particulates **in combination**, in the context of home owners doing **decoration** or **DIY** as well as in **construction**. One potential opportunity may be at the point where paints and other construction **materials are sold** (e.g., at Homebase or B&Q).

- There may also be opportunities to inform home owners about the risks of keeping buildings **sealed.** However, it is also evident that the latter will need to be counter-balanced by the location and risk for **allergic reactions** to outside air-borne BioPM.

Conclusions: Some notable findings (d)

- Participants appear to agree that **guidance** needs to be developed for **industry**.

- However, this requires **careful planning** and development (i.e., it is "not something you pick off Wikipedia").

- Some commentators also note that even in the mechanisms where one would expect such guidance, bioaerosols are **not well represented**.

Conclusions: Some notable findings (e)

- Opportunities exist to leverage **change** in policies and the development of guidance at the point when buildings, especially houses, are being **designed**, as well as when they are built, including the **types of materials** used in the process.

- One potential impediment seems to be the costs involved in changing **construction practices**, and how this may negatively impact the construction companies' profit margins. To overcome this, **government legislation** may be necessary.

- Opportunities to engage **estate agents** in evaluating indoor air quality as one of a property's **selling points**, or for air quality to be **periodically tested** in the same way as the energy rating of a property is rated.

Conclusions: And finally ...

- Our study contributed insights into both how BioPM is **perceived and understood** by non-participant stakeholders, but also the ways in which the 'pluralistic' and 'context-based' engagement of specialists and stakeholders participated in the **co-production** of BioPM-related knowledge.

- We conclude that knowledge is co-produced, and, like meaning, emerges in that liminal **interactive space** between interlocutors. How we communicate knowledge must take this co-productive process into **account**.

Thank you

www.bioairnet.co.uk