Evaluating the Mental Models Approach to Developing a Risk Communication: A Scoping Review of the Evidence

Journal:	*Risk Analysis*
Manuscript ID	RA-00474-2015.R1
Wiley - Manuscript type:	Original Research Article
Key Words:	mental models, risk communication, review

Evaluating the Mental Models Approach to Developing a Risk Communication: A Scoping Review of the Evidence

ABSTRACT:

Risk communication is fundamental in ensuring people are equipped with the knowledge needed to navigate varied risks. One generally well-regarded framework for the development of such communications is the Mental Models Approach to Risk Communication (MMARC). Developed during the 1990s, the MMARC has been applied to a range of health, technological and environmental risks. However, as yet, we know of no attempt to collate and review articles that evaluated communications developed using the MMARC. The current paper took a first step at addressing this gap by conducting a scoping review, which aimed to begin to explore the fidelity with which the approach has been applied, explore whether there appeared to be sufficient studies to warrant a future systematic review, and identify future research questions. Although the initial search found over 100 articles explicitly applying the MMARC, only 12 of these developed a risk-related communication which was tested against a control (and thus included in the current review). All studies reported a positive effect of the MMARC vs. control communication for at least some of the outcome measures (knowledge being the most prevalent). However, there was wide variation between studies including type of control, outcomes assessed, and only 5 studies reported adopting a randomised design. The review highlights both the need for greater fidelity in the way future studies operationalise the MMARC approach, and suggests that a full-scale systematic review of the MMARC literature appears justified, especially given the possibility of a large grey literature in this area.

KEYWORDS:

mental models, risk communication, scoping review

1. INTRODUCTION

The world contains many hazards. Some are relatively infrequent, such as major earthquakes and nuclear energy disasters, whilst others are encountered daily, for example car driving or food preparation.[1–4] For many hazards, people are often unable to personally assess the risk. Reasons for this include a) the hazard being undetectable by the senses (e.g. microscopic foodborne pathogens invisible to the naked eye), b) it being too complex for non-specialists to fully comprehend (e.g. financial risks), and c) the fact that people lead busy lives with many competing pressures on their time and attention.[5–7] Regardless of the reason, people often rely on those with a greater knowledge (sometimes referred to as 'experts') to provide them with salient hazard information (although see below for a discussion on two-way communication).[8] This process of information exchange about a hazard is referred to as risk communication.[9–11]

Notwithstanding the fact that members of the public have a wide range of knowledge on risk topics, it is generally acknowledge that 'experts' tend to have a more detailed understanding of the technical aspects of a risk, for example relating to the hazard's effects and the pathways through which it operates. However, additional factors affect how the public perceives a risk, for example its controllability and their familiarity with the hazard, with which experts might be less familiar.[12,13] Given uncertainty over the public's knowledge of a hazard, and these additional factors affecting their perception, a

particular challenge for risk communicators is determining appropriate message content.[10,14,15] On the one hand, there is no point in telling people what they already know, but on the other, information indispensable to understanding a hazard should be communicated if not already known.

Historically, the process of risk communication was often based around a one-way approach, for example where experts formulated the content of the risk communication detached from public input (or vice versa).[8,14] Now, it is recognised that a two-way approach (involving both traditional 'expert' and 'lay' perspectives) is more appropriate because it involves experts working with the public throughout the risk communication process.[10,14,15] Dialogue between these groups helps ensure the risk communication takes into account the audience's knowledge and concerns, in theory more effectively communicating appropriate information.[14,16,17]

One well-established two-way framework for developing risk communication is the Mental Models Approach (MMARC).[10,18–20] The MMARC is based on the idea that people's views of a concept are based on a complex web of information, drawn from personal experience and external sources.[21] Mental models, as these webs are known, are not always based on accurate information.[10] For example Hagemann & Scholderer define a mental model as '*a mix of factual knowledge, erroneous assumptions, value judgements, and uncertainty*'.[22] The MMARC aims to improve the accuracy of people's mental models relating to a risk and thus inform their decision-making.[23] The MMARC assumes that a scientifically-accurate decision model ('expert model') will be technically

more accurate than the public's (i.e. 'decision-maker') mental model of a risk.[10,24,25] The challenge of determining risk communication content is therefore tackled by comparing an expert model with the public's mental models, which forms the first two parts of the five-step process (see Figure 1).[10] The third step is a confirmatory survey of the target audience to determine the prevalence of knowledge gaps and misconceptions, and thus prioritisation of the communication's content. The final two steps involve the development and iterative evaluation of a risk communication explicitly developed from information gleaned during the earlier steps.

The MMARC was developed during the 1990s, and is being increasingly applied to a diversity of topics (see Figure 2). These topics include health, (e.g. HIV, contraception), technological (e.g. genetically-modified foods, nuclear energy) and environmental (e.g. flooding, wildfires) related hazards.[22,26–31] In some cases not all steps of the MMARC are applied. One example is Hagemann & Scholderer, who applied the first two steps of the MMARC and identified mental models of how experts and the public perceived risks and benefits associated with a novel potato, genetically-engineered to contain lower toxin levels than conventional potatoes. However, the authors did not subsequently design and test a new communication which attempted to build on these earlier steps.[22]

The MMARC is well-regarded but, despite this praise, to our knowledge there has been no attempt to collate studies that have evaluated communications developed using the MMARC.[9,32,33] This is despite the fact that Bostrom *et al*. [1994, p.789] stated that although

the MMARC is '*advanced on logical and theoretical grounds...direct empirical evaluation is needed to assess its products*'.[20] The current paper aims to fill this gap by reviewing studies that have evaluated a communication developed using the MMARC. Importantly, we focussed specifically on those studies that included the final step of the MMARC where a communication developed using information from early stages was compared to a control communication not developed using the MMARC approach, rather than those where only the earlier steps were applied (Figure 1). This is to begin to ascertain the efficacy of communications developed using the MMARC process. Guidance for assessing MMARC communications already exists, with common elements including consideration of the target audience, communication format, and outcome.[10,18,34]. These elements were thus incorporated into the data extraction stage of this study.

Specifically, we conducted a scoping review to begin mapping out some of the key research underpinning this area. A scoping review differs from a full-scale systematic review in several ways and has slightly different aims. For instance, a full-scale systematic review explicitly attempts to: a) synthesise all research in the field including that in the grey literature, b) reduce bias in the review process by adopting highly specialised literature searches often including following up with known authors in the field, c) conduct inter-rater reliability for data extraction and quality assessment, and d) when feasible, conduct meta-analyses of the findings. Given the thorough nature of the process such reviews are extremely resource intensive and can take several years to complete. Consequently, it is recommended to conduct, and publish the findings of, an

initial scoping review, in order to investigate whether investing the resources needed for a full-scale systematic review is justified.[35] In the words of Peters et al. (2015), '*Scoping reviews are commonly used for 'reconnaissance' – to clarify working definitions and conceptual boundaries of a topic or field… [they] are particularly useful when a body of literature has not yet been comprehensively reviewed … While scoping reviews may be conducted to determine the value and probable scope of a full systematic review, they may also be undertaken as exercises in themselves to summarize and disseminate research findings, to identify research gaps, and to make recommendations for future research*'.[35, p.141] Following this advice, the aim here was not to try and attempt to conduct a full-scale systematic review, but rather to map the field out in a way which should both provide some important insights in and of itself but also help identify exactly where a systematic review might focus on in future.

2. METHODS

This study followed the steps outlined in the scoping review literature which consist of a structured search to identify relevant studies, followed by data charting and summarising results, as detailed below.[36–38] Of note, and reflecting the scoping nature of the current review, as opposed to the procedures in a full systematic review, all three stages were conducted by the first author in discussion with the other authors, rather than multiple authors attempting to replicate the search strategy independently.

2.1 Search

The search involved a combination of forward citation (to identify references that have cited a particular source), standard searching (using a database to identify references that meet particular key words), and hand searching (e.g. identifying references from the bibliography of a paper). Forward citation was used for four key sources that outlined the MMARC and was particularly suitable as part of the search strategy because the emergence of the MMARC is well-defined in these particular sources.[10,18–20] However, one of these sources for which forward citation was required is the book by Morgan *et al*. (2002), and some academic search engines do not include sources such as books (see Table I).[10] It was therefore decided to also use the search engines Web of Science and Google Scholar (which allowed forward citation on most of the key sources). It is recognised that there are differing opinions about the use of Google Scholar in systematic searches, for example due to reproducibility.[39] However, it was considered a sound choice for this initial scoping study, was not used in isolation, and is also supported by recent studies comparing search engines.[40,41] The search strategy is set out in Table II.

2.2 Identify relevant studies

Citations were exported directly from Web of Science to EndNote. For Google Scholar, Zotero was used to export citations via Mozilla Firefox. All references were imported to a single EndNote X7 file. After duplicate references were removed, the inclusion criteria listed in Table III were applied.

Following the title and abstract screen, full text copies of articles were obtained and the same inclusion criteria applied. During the title / abstract sift, if it was clear that a paper did not report the evaluation of a communication (i.e. it was not relevant to this study's aims), but it was unclear whether the study was applying the earlier interview and survey stages of the MMARC, the full text was checked to ensure the correct exclusion criterion was applied (see Table III).

2.3 Data charting

Data were extracted from studies meeting the inclusion criteria and summarised under the following headings which include the PICO structure for evaluating research (Population, Intervention, Comparison, Outcome):[42]

1. Study description: authors, year, country, topic, whether the study included at least one author of the original study (i.e. Morgan, Fischhoff, Bostrom and Atman), and the knowledge measures (i.e. whether direct or indirect).[10]
2. Participants (target audience): from which population the participants who evaluated the communication were drawn.
3. Intervention: the MMARC communication, and its format.
4. Comparison: whether a control communication was used in each study, and if so its format.

5. Outcome: the effectiveness of the MMARC, and what was measured in participants following exposure to the communication (referred to here as the outcome measure), e.g. knowledge, behaviour.
6. Process: methods / experimental design: how the communication was developed and evaluated (e.g. survey, focus group), and whether randomisation was used.

2.4 Summarising the results

In keeping with scoping review recommendations the extracted data was evaluated to identify trends, gaps, and how the findings related to the broader risk communication field. For example, this included how success was measured, whether particular aspects of the communications affected their impact, and key points for future risk communication evaluations. It did not involve detailed analysis of suitability of statistical procedures or appropriateness of sample sizes in the relevant papers. Where study details were unclear, further information was requested from the authors and incorporated (where provided) in the current review.

3. RESULTS

All searches were conducted during February 2014. A total of 2,504 references were identified. Figure 3 details the screening procedure used to sift the references identified by the searches. Table IV details the 12 studies that met all inclusion criteria and are therefore included in this review.

3.1 Search process

Although we found over 100 papers which reported empirical findings using the MMARC approach, 90 of these did not include the final steps of the approach, i.e. the design and evaluation of a novel communication drawing on the findings of the earlier steps. In fact, we found only 12 studies that actually designed and tested a risk communication based on the approach. Nonetheless, many of these 90 studies concluded with advice, which may have since been used by other organisations involved in risk communication.[27,43–45] Also, there may be an assumption when embarking on the MMARC that sufficient differences exist between the expert model and the public's (collective) mental model to justify the production of a communication, which may not always be the case. Although these 90 papers are not discussed further here, it is important to acknowledge their contribution to the MMARC literature in providing a wealth of examples in which the earlier MMARC stages were applied (i.e. steps 1-3 in Figure 1) and which may want to be included in a future systematic review into the MMARC approach more generally. Furthermore, a future systematic review may also want to consider a more extensive search process where the authors of studies that did not include the evaluative step could be contacted, to request potentially unpublished data exploring an evaluation which could be incorporated into a meta-analysis.

A further two studies that met the inclusion criteria were not included in the analysis (Morgan *et al*., 1992; Cone *et al*., 2013), because the findings in the former are reported in greater detail by Bostrom *et al*. (1994).[20,23,46] Additionally, Cone *et al*. 2013 reported that a communication was developed and evaluated using the MMARC, but referred to details within another paper submitted for publication and not available for review at the time of these searches (Cone & Winters, 2014).[47] No additional papers were found after reviewing reference lists (see Table II) of the 12 papers that met the inclusion criteria, supporting our contention that the current search approach was a satisfactory first step.

3.2 Study description

The 12 studies included in this review were conducted in six countries over the past two decades, and focussed on a range of topics. The USA accounted for over half (seven) of all studies, with one study conducted in each of the other five countries (Canada, Mexico, Germany, Switzerland, UK). All US studies included at least one of the original study authors (i.e. Morgan, Fischhoff, Bostrom and Atman). The 12 studies were spread from 1993 (Maharik & Fischhoff) to 2012 (Vogt & Schaefer).[48,49] Topics on which the studies focussed represented a range of risks, including environmental e.g. radon,[18,20] health e.g. xenotransplantation and disease,[50,51] and technological e.g. mobile phone masts and electrical fields.[52,53]

3.3 Participants (target audience)

Our review suggests that only four studies (33%) specifically focused on the key target audience, both in terms of region and age, at the evaluation stage.[34,49,51,54] For example, Niewöhner *et al*. (2004) evaluated a risk communication about occupational chemical health risks, specifically rosin-based solder flux in the electronics industry, and perchlorethylene in the dry cleaning industry, with participants recruited from these particular industries. In another example, Downs *et al*. developed a MMARC communication which aimed to inform young women about contraceptives and sexually-transmitted disease (STD) risk.[51] Participants were drawn from this particular demographic group. In both cases, participants directly represented the target audience at risk, for whom the communication was intended to improve the mental models. In contrast, participants in other studies did not represent the target audience of people whose collective mental models were elicited and recognised by the authors as incomplete for the topic under investigation.

The majority of studies used convenience sampling to recruit participants to take part in the evaluation of the MMARC communication (see Table IV). In some cases the limitations of this approach were recognised, i.e. that the views of the sample with whom the communication was tested may not represent those views of the communication's intended audience e.g. De Bruin *et al*. (p.1408) and Longstaff (p.36).[50,55] The challenges of recruiting participants from the target population were described by Niewöhner *et al*.[34] The authors reported a low response rate of 7-12% for return of completed postal surveys during the analysis of the communication related to

management of chemical risks amongst small companies (the target audience). It was suggested that tighter profit margins amongst many smaller organisations resulted in a lower expenditure on health and safety, and a corresponding reluctance amongst managers of such organisations to participate in safety surveys. The authors also described how such industries were poorly understood by the research community, which presents a challenge in identifying suitable companies in the first place.

3.4 Intervention (the MMARC communication)

The MMARC communications in the 12 studies varied by format, length and scope. The format of the MMARC communications was almost exclusively written, except for Downs *et al*.[51] Of these 11 studies, a brochure / booklet was used in six studies.[18,20,48,49,52,56] In one study, a comic book was used to communicate carbon monoxide poisoning risks to residents of a Mexican town.[54] Another two studies used a combination of text and graphic, which in one was written on several sheets of paper, and in the other the exact format (e.g. whether a leaflet or brochure) was not stated.[50,53] One study used an information sheet,[34] and another a combination of flow chart / case study / FAQs (frequently asked questions).[55]

Communications' length varied between studies from one-page brochures[34,56] to over 50 pages.[48,49] There was variation in whether participants were asked to view the full communication, which appeared to be connected to length. That is, for longer communications participants were not required to read them in their entirety, but rather

could focus their attention on specific parts of greatest relevance.[48,51] For example, participants viewing the STD risk video were not required to watch the full one hour length and were directed towards relevant parts.[51]

Finally, the breadth of the information in each communication was generally wide, including detail about multiple aspects of a risk. One example was Cousin *et al*. where the booklet included general information about radiation from mobile phone systems, the current state of research and scientific uncertainties.[52] The one exception to this pattern was Read & Morgan, where the communication focussed on correcting a specific component of the audience's mental models of electric fields (range dependency, i.e. the relationship between magnetic field strength and electricity source distance).[53] In this case, because one specific misunderstanding had been identified in the public's collective mental model the communication was correspondingly tightly focussed on this one specific area.

3.5 Comparison (the control communication)

Over half (eight of twelve) studies included here used at least one control.[18,20,34,48,49,51–53] In all of these, except Niewöhner *et al*. and Read & Morgan, multiple controls were used.[34,53] The highest number of controls used was seven, where in all cases the MMARC was identified as structurally superior to the control brochures (communication structure was the outcome measure).[18] The most frequently used control (in seven of the studies) was an alternative communication; in contrast Read & Morgan did not

provide instruction for their control group.[53] Controls were matched to the MMARC communication in at least one of three different ways. Firstly, the control could be topic-matched i.e. if the MMARC communication focussed on radon, a control communication that also focussed on radon was selected. All seven studies with an alternative communication used a topic-matched control. In six studies this was an existing brochure, the exception being Cousin *et al*. where a control text about a topic unrelated to mobile phones (a Swiss abbey) was used.[52] Secondly, in five studies the controls were also format-matched (i.e. if the MMARC was a brochure, so too was the control communication).[18,20,34,48,49] Thirdly, one study used a content-matched control, where a book was developed to provide the same informational content as the MMARC video.[51] In all studies, the MMARC communication performed equal to or better than the control communication(s) for the outcome measures assessed.

3.6 Outcome

Amongst the 12 studies reviewed here, 14 different outcome measures were used to assess the communications developed via the MMARC. Multiple measures were used by 75% of the studies. All 12 studies reported a positive effect of the MMARC communication for at least some (it not all) of the outcome measures against which it was tested (Table IV). There did not appear to be pattern in the outcome of the seven studies that included one of the original study authors compared to those that did not, with predominantly positive effects of the MMARC communication seen in both groups.

Knowledge was the most frequent outcome measure and used in all but one study, i.e. Atman *et al*. who assessed the communication's design and structure.[18] In the remaining 11 studies, most reported a significant improvement in participant knowledge following exposure to the MMARC communication compared to the control or baseline. Two studies concluded that whilst participants in both the MMARC and control conditions scored significantly higher knowledge, the difference between conditions was not significant.[49,51] Knowledge assessments were in most cases direct i.e. topic specific, apart from one study (Table *V*).[34] The measures were often developed by the authors based on prior interview findings (as would be expected using the MMARC). One example was Galada *et al*. who asked participants about the sources and effects of carbon monoxide.[54] Across all 11 studies questions were presented in a variety of formats, including multiple choice, and dichotomous assessments such as true/false.[52,56] Additionally, scales were used allowing participants to express uncertainty, such as Bostrom *et al*. who used a five-point scale (true, maybe true, don't know, maybe false, false), and Vogt & Schaefer used a four-point scale where the neutral option was removed.[20,49]

The most frequent secondary outcome measure was participants' attitude towards the communication topic (e.g. nuclear energy use in space in Maharik and Fischhoff), assessed in three studies.[48,49,54] In all cases there was a positive relationship between the MMARC communication and participant attitude, although in Maharik & Fischhoff and Vogt & Schaefer there was no significant difference in the extent to which the MMARC and control communications appeared to influence participant attitude.[48,49]

Another secondary outcome, behaviour, was assessed in one study which investigated sexually transmitted disease.[51] Here, self-reported sexual behaviour of females was assessed, and complimented by clinical testing at baseline and at six months after exposure to the MMARC communication. Although the MMARC communication was not associated with a significant knowledge increase above the control communications (a book and brochure), for participant behaviour the difference *was* significant, with clinical testing showing participants who viewed the control were twice as likely to be diagnosed with an STD compared to those viewing the MMARC communication.

3.7 Process, methods and experimental design

The process used to create the MMARC communication varied between studies. As shown in Figure 1, developing a communication in line with the MMARC involves five distinct steps. In some of the 12 studies reviewed here, it is not clear to what extent each of these steps was followed, and thus to what extent each of the 12 studies evaluated a communication that is the product of the full MMARC. For example, six studies reported that they developed the communication through comparing expert and public mental models, which *is* in line with Morgan *et al*.'s process.[18,20,34,49,51,54] However, in other studies a slightly different approach was taken (albeit still within the broad scope of the MMARC), in that the communication reflected uncertainties in public mental models. One example was Longstaff, where the confirmatory survey (which

usually precedes development of the communication, see Figure 1) was combined with testing the communication.[55][1]

In general, quantitative methods were used to evaluate the MMARC communication (Table IV). The exceptions to this were Atman *et al*., Bostrom *et al*. and Niewöhner *et al*.[18,20,34] The first two of these three studies used mixed methods (i.e. quantitative and qualitative), and Niewöhner *et al*. used solely qualitative methods.[34] The qualitative methods included analysing communication design, where it was concluded that the MMARC communication was better structured to provide the information salient to participants' mental models.[18] A think-aloud protocol was another qualitative method, whereby participants verbalised their thoughts as they read each communication.[20] Similar 'think aloud' methods were used by Niewöhner *et al*., alongside user evaluation sessions, in which groups of two to four participants discussed positive and negative aspects of each communication.[34]

Most of the 12 studies used between-participant designs to compare the effects of MMARC and control communications, except Atman *et al*. who compared communication design and layout within participants.[18] A longitudinal component was present in two studies.[49,51] Downs *et al*. provided participants with the opportunity to review the communication at intervals after the initial session and so 'top-up' their exposure.[51] This revealed differences in participants' self-reported behaviour,

[1] In some papers where the initial steps of the MMARC process were mentioned only briefly, a reference was provided for the full details e.g. Cousin *et al*. (2011) and Vogt & Schaefer (2012) built on findings published in previous studies.[29,49,52,72]

specifically that participants who viewed the MMARC communication were significantly less likely to report sexual behaviour during the first three months than participants in the control group. However, a significant difference between the groups was not observed for the latter three months, or for some other self-reported behaviours (e.g. condom use). Vogt and Schaefer also reported a longitudinal study, where two brochures about contraceptives were compared.[49] Researchers measured participant knowledge, attitude and intentions at three points: before the intervention (baseline), directly after reading the brochure, and three months later.

One aspect of internal validity, namely randomisation, was also compared between studies. Less than half (five) the studies stated that participants were randomly allocated to the MMARC and control communication conditions.[20,49–52] In short, this one common elements of experimental design was the exception rather than the rule for the studies reviewed here.

4. DISCUSSION

The MMARC is a widely used and recommended framework for developing risk communications.[10,33] The need for such communications arises as the public are often unable to personally assess the numerous hazards they face, instead often relying on ‘expert’ information to safely navigate these risks. One particular challenge in the development of such communications is understanding what the public wants and needs to know. The MMARC aims to remedy this issue by providing a framework to

develop and evaluate targeted communications. The aim of the current paper was to conduct an initial scoping review of the relevant literature in order to map out the field, identify similarities and differences in the ways in which extant studies have applied the MMARC, and explore whether a full systematic review might be feasible and desirable in future. Our focus was on those studies which had used the approach to actually develop and test a risk communication, rather than those who had only used the earlier steps in the model and not proceeded to these final stages (Figure 1). Perhaps the most striking finding of our scoping review was just how few of the studies that claimed to be using the MMARC actually went on to develop and evaluate a risk communication. Moreover, the fidelity with which the MMARC was applied in these 12 studies was extremely varied. Nevertheless, the generally positive results from these few studies suggests that a more detailed and systematic investigation into the approach may be warranted. Below we discuss some of these findings and why we think further synthesis work is justified, as well as identifying those issues we think need particular attention going forward.

4.1 Measuring success of the MMARC

Overall, results indicated that risk communications developed using the MMARC were successful in significantly improving participants' knowledge. Therefore, our initial tentative conclusion, based on this limited set of studies, is that the MMARC might indeed be a useful framework within which to develop a risk communication. However, there are two important caveats, relating to *how* success was determined, and to *what*

the communication was compared. Firstly, determining success using knowledge provides only a partial indication of a communication's impact. That is, it demonstrates information transfer. However, the communication may have additional effects not captured via knowledge assessment, but with implications for risk avoidance e.g. participants' attitudes and / or behaviour towards a risk are modified. The findings of, for example, Read & Morgan are therefore limited, in that only knowledge was measured and not, for example, whether participants' attitudes towards magnetic fields had changed.[53]

Further, although measuring knowledge change is a fundamental part of measuring the success of any MMARC communication, the type of knowledge assessed is also important. For instance, two types of knowledge have been discussed in the literature: direct (subject-specific) and indirect (broader scientific), and each has different relationships with people's action and attitudes.[57–59] Whilst assessing direct knowledge (as was the case in most of the studies reviewed here; Table *V*) may suggest the potential effects of the communication on participant, directly evaluating secondary outcomes can provide a more comprehensive understanding of a communication's impact (i.e. indirect knowledge).

When secondary outcomes are considered, results from the current study still point to the MMARC as being a potentially useful approach to developing successful risk communications.[49,51,52,54] However, the relationship between knowledge and secondary outcomes in the 12 studies was variable, with some showing a significant change in

participant attitudes (Galada *et al*.) while in others no significant effect on attitudes was found (Maharik & Fischhoff).[48,54] However, whereas knowledge was assessed relatively consistently across all studies except one (Atman *et al*.), the choice of secondary outcome(s) was topic dependent and thus highly variable across studies.[18] For example, where individual health risks were the focus of the communication, behaviour (or behavioural intention) was assessed.[49,51] This allowed Downs *et al*. to determine the communication's effect on both participants' self-reported behaviour, which they combined with clinical validation to provide a strong assessment of the communication's impact.[51] On the other hand, where Maharik & Fischhoff trialled a communication relating to nuclear energy use in space, they used the potentially more appropriate secondary measure of participant attitude towards the technology (given that no direct behavioural relationship was possible).[48] Whilst measuring behaviour is suitable for assessing MMARC communications about individual health risks (e.g. smoking, alcohol consumption), where the individual can more easily adjust their behaviour and influence risk exposure, for societal technological risks (e.g. nuclear energy), attitudes may be a more apt measure.

In a sense, it only matters if a communication influences, for instance, attitudes, if influencing attitudes was an explicit goal of the communication. Nonetheless, for the good of the field in general, it might be worthwhile studies consistently measuring a range of constructs, to a) explore the underlying psychological processes behind changes in the main outcome variables (if they are different from, say, attitudes), b) because a communication may have unanticipated impacts, which the assessment of

other secondary outcomes (e.g. attitudes) would help identify, and c) to allow later synthesis across multiple studies containing operationalisations of the same kinds of construct. With such an idea in mind what would a (semi-) standardised set of outcome variables look like? One possibility would be to build on the three evaluative criteria identified by Fischhoff: materiality, proximity and comprehensibility (relating respectively to whether a communication contains information relevant to its audience, can be accessed, and finally understood).[25] These criteria might form a useful checklist to aid researchers in deciding the most appropriate outcomes to assess in future evaluations e.g. in deciding which secondary outcomes should be included alongside knowledge.

The second caveat relating to support for the MMARC is linked to *what* the MMARC communication was compared to. When deciding whether to embark on the time- and resource-intensive MMARC, this decision should be made with an idea of alternative risk communication development frameworks in mind.[60] A quicker, perhaps nearly as effective approach might also be a viable option for researchers. However, one limitation of studies in the current review is that the framework underlying the control communications' development was rarely stated.[20,52][2] This makes it challenging to assess whether the MMARC is more suitable than an alternative framework if the alternative is not clearly described. Our findings thus agree with other recent reviews of risk-benefit communications, which also identified a gap in reporting the underlying

[2] A framework was hinted at in Vogt and Schaefer who stated that the alternative intervention 'followed standard recommendations for [Evidence-Based Patient Interventions]', and provided a reference but no further details.[49]

framework.[61,62][3] An additional challenge related to the control communication is the unavoidable bias in measuring knowledge change when the researchers develop both the MMARC communication, and the questions by which it is assessed. If questions assess information contained solely in the MMARC communication it would not be surprising if there would be a greater chance of identifying a difference compared to the control(s) than if the communications are content-matched. This was the case in Downs *et al*., where there was no significant difference in knowledge between participants exposed to the control compared to those exposed to the intervention.[51] The implication of the method determining its own criteria for success is that comparisons between studies are more difficult, because the assessments are specific to each communication. One option for improving comparability between studies could be to standardise response scales (discussed further in Section 4.4). The lack of content-matched controls (apart from Downs *et al*., as mentioned above) also presents a challenge to being able to draw firm conclusions about which elements of each communication were associated with any change in outcome measures.[51] This is a challenge any future systematic review would need to address.

Moreover, in terms of future research, there may be solutions to these dual challenges of alternative risk communication framework, and fair comparison with the control communication. Firstly, future MMARC evaluations could endeavour to use a control where an alternative risk communication framework was identified. This may not always be possible, in which case there is also a broader need for developing, evaluating and

[3] These reviews were topic-focussed, relating to environmental and food risk communication respectively, unlike the current paper which attempted to review a specific framework, i.e. the MMARC.

reporting alternative risk communication frameworks, to provide the evidence required to inform framework selection. Reporting which framework has been followed is crucial to facilitate a fair comparison of the MMARC with alternative frameworks. Without this information, even a larger sample of studies will not provide greater clarity as to the relative efficacy of the MMARC. Unbiased assessment of the control communication is an inherent challenge in evaluating risk communications. In addition to the use of a content-matched communication as in Downs *et al*., another recommendation for future studies is to measure subjective knowledge i.e. how well-informed people feel, rather than their objective (factual) knowledge.[51,63,64] Although not the case in the 12 studies reviewed here (11 of which measured objective knowledge), subjective knowledge has been included as an outcome measure in other risk communication assessments.[63,64] Applying this measure in future assessments could provide a more complete understanding of a communication's impact.

4.2 What aspects of the communication affected its impact?

As the 12 MMARC communications in this review were associated with near ubiquitous improvements in participant knowledge and secondary outcomes, it is challenging to identify which specific communication facets (e.g. length, use of graphic or text) were responsible for success.[4] It may of course be the *combination* of factors that resulted in improvements in participant mental models. Although reviewing the fine detail of the

[4] The successes may also represent an example of the file drawer effect, with evaluations not achieving statistical significance less likely to be published.[73] Again a future systematic review could take steps to uncover this grey and unpublished literature.

communications developed in each of the 12 studies was outside the scope of this paper, we would recommend that any future systematic review sought to undertake such an analysis. Nevertheless, the studies reported numerous features of the communications, which can inform the development and evaluation of future MMARC communications, including length, format (i.e. use of graphic, presentation), and channel (i.e. whether paper or video).

Whilst the communications' length varied between studies (from 1 to over 50 pages for written communications), it is not clear whether longer or shorter was superior. The next factor was format / presentation, which did appear important (e.g. Atman *et al*.; Longstaff; De Bruin *et al*.), although the exact effect was complex.[18,50,55] Specifically, graphic presentations, or text-graphic combinations were found superior to text alone by De Bruin *et al*., in contrast to another recent study comparing format where graphic communications were insufficient to rectify misunderstandings about Carbon Capture and Storage (Seigo *et al*.).[50,65] Fitzpatrick-Lewis *et al*., in their review of risk communications, suggested mixed formats are superior to single for effective information transfer. Lastly, the 12 studies largely focussed on written communications (e.g. brochures in Fleishman *et al*.), with video used only once (Downs *et al*.).[51,56] However, participants exposed to the video format displayed significantly different behaviour to those exposed to the (written) controls, suggesting the format may have influenced the evaluation outcome.[51] There were no web-based communications which, given the increase in society's use of social media etc., means there is currently a gap in evidence assessing this channel's effectiveness in disseminating MMARC

communications.[66–68] Future, more up-to-date systematic searches of the published and unpublished literature may begin to reveal such studies.

The upshot of these findings is that an effective MMARC communication contains not just relevant information, but presents it clearly. Future communications should consider length, combinations of text and graphics, and use of channels such as web-based communications. The precise combination will however depend on the target population's needs and preferences. Understanding the most appropriate format(s) for a given hazard can begin early in the MMARC process, for example Morss et al, included a diverse range of stakeholders during development of the expert model, resulting in valuable perspectives as to how flood risk warnings could be better communicated.[69] Such an approach also demonstrates the need for two-way communication at an early stage and a need to move away from the simple expert-lay dichotomy. In addition to audience knowledge requirements (including experts' requirements), the researchers also identified needs such as ensuring the messages strike the right balance between sufficient warning and unnecessary precaution.

Similarly to communication format, the target audience also plays a key role in determining the communication's content. MMARC communications are, by their nature, focussed towards a particular population's inaccurate or under-developed mental models. However, the 12 studies varied in whether they trialled the communication with participants from the target population. There was no clear pattern between the outcome of the MMARC assessment and whether participants represented the target

population. In spite of this, the use of participants from the target population appears essential to precisely determine the MMARC communication's impact. The importance of this comes back to one of the basic tenets of the MMARC as described by Morgan *et al*.: '*[communications] need testing with people drawn from the target audience*' (p.103).[10] If a communication is released to a population with whom it was not tested, there would be considerable uncertainty as to its impact i.e. whether it increases / decreases / has no effect on knowledge, or some other unanticipated reaction. For this reason, future evaluations should recruit participants representative of the target population.

Recruiting participants from the target audience can nevertheless pose a challenge for researchers. Such challenges were described by Niewöhner *et al*. in recruiting people from small- and medium-sized businesses despite offering a financial incentive, and using both telephone and postal requests for participation.[34] These methods for encouraging participation represented best practice amongst the 12 studies, and should be considered in future MMARC evaluations. Additionally snowballing, where participants recommend a colleague, may help increase participation in similar circumstances. Other techniques future MMARC studies might consider when recruiting representative samples include market research companies, where specific demographic groups may be targeted. However, whilst market research is a useful tool, as with other approaches it can be challenging to obtain adequate sample response rates.[70]

4.3 Strengthening future evaluations

Two ways in which future MMARC evaluations could improve relate to their methods and study design. Firstly, quantitative methods dominated the 12 studies reviewed here, with qualitative methods used in only a quarter of the evaluations. Our findings support those of Scammell, who reviewed use of qualitative methods in environmental health research and concluded such methods were underutilised.[71] The benefit of qualitative methods is that potentially unanticipated effects of the MMARC communication can be identified and corrected during the trial phase, and before wider dissemination. This should provide the communicators and other stakeholders with greater confidence that their message will have the intended effect. Niewöhner *et al*. described how they learnt as much about how the target audience perceived risks from their (qualitative) evaluation than during earlier unstructured interviews.[34] Quantitative methods are nevertheless fundamental in evaluating MMARC communications, and a *dual* approach, incorporating qualitative methods, would improve future evaluations. Studies applying a qualitative element should inform future MMARC assessments, for example including a 'read-aloud' component.[20,34] This involved participants reading the communication and concurrently verbalising their thoughts, thus providing the researchers with an approximate explorative assessment of the communication's impact, and the opportunity to capture unexpected reactions.

Future MMARC evaluations could also improve study design rigour for the elements considered in this review. The two studies focussing on health risks were also the

strongest in terms of study design (Downs *et al*.; Vogt & Schaefer).[49,51] That is, both were longitudinal randomised controlled trials (RCT), and thus represented best practice. The implications of not randomising the communication's distribution is the potential introduction of bias through incidental, but systematic, variation between the groups. This potentially means another factor (i.e. not the intervention) could be responsible for any relationship identified by the study, and thus caution must be applied when considering the outcomes of such studies in the current review. Downs *et al*. described that no theoretical risk communication approach has demonstrated superiority due to a lack of RCTs, and concluded that comparing the MMARC communication with communications developed using other theoretical approaches is necessary to determine whether one framework is better than another.[51] Our findings support their conclusion. Future studies should utilise RCT study design to ensure robustness of findings and improve internal validity. This would also aid any future systematic meta-analysis of results.

Longitudinal design also provided unique insights about the MMARC communication's impact (Downs *et al*.; Vogt & Schaefer).[49,51] Specifically, the longitudinal analysis revealed whether the communications had a lasting effect on participants, and the degree to which each communication differed in its long-term impact. This is important because risk communicators will often be interested in creating a lasting rather than fleeting impact on their audience. Lasting impacts ensure the population remains informed, and additionally are more efficient if the first communication was sufficiently effective that repeated communication efforts could be avoided. Ideally, future studies

should follow a similar structure to these where outcome measures are measured at several intervals following exposure to the communication, and the results validated (e.g. Downs *et al*. who used clinical testing as this was relevant to their study).[51] If resources do not permit such a detailed follow-up, measurement of participant willingness to change behaviour, alongside knowledge can provide a second best alternative, as in Galada *et al*.[54]

A final point relevant to reporting methods and study design, is a call for future studies to describe more clearly which MMARC components were employed in the communication's development. Guidance in the definitive MMARC book by Morgan *et al*. suggests researchers can adapt the process to their needs.[10] Some of the studies reviewed here, e.g. Fleishman *et al*.; Maharik & Fischhoff were vague about their process beyond stating that the MMARC was followed.[48,56] Others, such as Longstaff, indicated where they deviated from the MMARC process, for example combining the confirmatory survey with trialling the MMARC communication.[55] Some aspects of the process may be more important than others in creating an effective communication. If more studies reported the precise steps they followed, it may be possible to determine which aspects were of most importance, thus allowing future justified refinement of the MMARC without sacrificing its efficacy.

4.4 Towards a revised framework for assessing impact of a communication

A framework for the assessment of MMARC communications would help standardise the evaluation procedure. Currently, a variety of methods and outcome measures were used to assess the MMARC. This would represent a distinct improvement on Niewöhner *et al*. (2004)'s observation that risk communication evaluation is often neglected and lacks robustness. However, it still presents a challenge in comparing studies and in identifying the most important components that contributed to a communication's success. Clearer guidance on evaluating MMARC communications would provide a valuable starting point.[10,20,34] For instance, further work should consolidate methods, suggest reporting criteria, and define common scales to evaluate outcome measures such as knowledge, attitudes, and intentions. By their nature, risk communications are topic specific, with assessments necessarily driven by idiosyncratic factors. However, several scales could be developed depending on discipline e.g. to measure attitude towards technological, health or environmental risks. A common framework would facilitate comparison of the MMARC, and thus identify and address common challenges, and advancing risk communication efforts. Additionally, such a framework may also be of use in evaluating alternative risk communication frameworks.

4.5 Limitations

The current method of a scoping review is the recommended way of mapping out a research field when a full scale systematic review has yet to be conducted.[35–38] Such an approach helps to scope out commonalities and issues within a field, identify future research needs, and is also used to help researchers decide whether or not to invest

the time and resources needed for a full scale systematic review. Importantly, we feel that the findings of the current review suggest that a full scale systematic review are justified because the majority of the studies did show encouraging outcomes (in terms of knowledge increase) but that there was enough heterogeneity across studies to warrant further investigation into identifying best practice. Nevertheless, in keeping with all scoping reviews, the current review has several important limitations that any future systematic review should seek to address.

First, although the search strategy was designed to be comprehensive and involved several databases and used various techniques (e.g. standard searching, forward citations), it is possible some papers or studies were missed. A subsequent systematic review may wish to consider additional databases in the searches, include grey literature e.g. through reaching out to the relevant research community (e.g. Society for Risk Analysis mailing lists), and consult study authors to find out whether, for instance, the 90 papers that only reported the early stages of the MMARC did in fact go on to test a communication developed using these findings but did not publish the results. Such a future review might also be able to begin a meta-analysis of data incorporating studies conducted, for government agencies for example, but not published in the scientific literature.

A second limitation relates to reliability. Specifically, the results were screened, and data extraction was conducted by the first author in consultation with the other authors. Ideally a future systematic review would involve multiple researchers independently

searching the literature for relevant studies, evaluating found studies against agreed inclusion criteria, and extracting the data from included papers. The reliability of these processes could then be assessed, for instance by calculating and reporting Cohen's kappa to assess inter-rater agreement.

Thirdly, the current review's focus was restricted to considering only those studies that have completed the final evaluative step of the MMARC. This is only one way of assessing the MMARC. The excluded studies that applied earlier steps of the MMARC may nonetheless hold valuable insights into the effectiveness of the framework even at the early stages. Clearly this would be a much larger undertaking than a focus on only those studies that actually designed and tested a communication developed using the MMARC, but might nonetheless be informative in terms of discovering its effectiveness, compared to alternatives, at the early stages of the process of eliciting 'expert' and 'lay' knowledge.

Finally, we also recognise that a relatively large number of the papers covered in this review were conducted by the original authors of the approach, which may have introduced bias (i.e. that the authors may have developed a method that they were able to implement with great fidelity but was not easily translatable to other settings). Despite the fact that our review found no evidence to suggest studies by these authors produced more positive findings than those by other authors, we also recognise that a future, more systematic review, would take into account the potential biasing effects of

authorship. Further, it also highlights the need for multiple groups to adopt and try and replicate the basic approach to ensure it is generalizable across contexts and cultures.

5. CONCLUSION

The current scoping review set out to investigate whether the MMARC appears, in general, to provide an effective framework for the development of risk communications. On balance, and given the inherent limitations in the current review protocols, we feel there is enough support for the approach in the currently reviewed studies to warrant the investment of time and resources needed for a full scale systematic review of the approaches' effectiveness in developing risk communications, or at the very least in eliciting lay and expert mental models about a variety of risks. The MMARC is, by its very nature, a very in-depth, multi stage process which requires considerable commitment from a research team. Knowing whether or not this effort is likely to be rewarded will no doubt be a critical factor in whether more teams use it to systematically develop communications. Although the findings from the current scoping review provide several indications that it might be an approach worth pursuing, far more studies are needed, especially one that adopt standardised evaluation protocols, before any firm conclusions can be made. Further, before recommended a wave of new research using the approach, we would also suggest that our findings give sufficient justification for a full scale systematic review of the field to now be conducted which would be able to shed an even clearer light on the full range of MMARC research that has been conducted to date, and also highlight best practice among this extant literature base.

REFERENCES

1. Adak GK, Meakins SM, Yip H, Lopman BA, O'Brien SJ. Disease risks from foods, England and Wales, 1996-2000. Emerg Infect Dis. 2005;11(3):365–72.
2. Burgherr P, Hirschberg S. A comparative analysis of accident risks in fossil, hydro, and nuclear energy chains. Human and Ecological Risk Assessment. 2008;14(5):947–73.
3. Milton JC, Shankar VN, Mannering FL. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis & Prevention. 2008;40(1):260–6.
4. Vere-Jones D. Forecasting earthquakes and earthquake risk. International Journal of Forecasting. 1995;11(4):503–38.
5. Alessandri TM, Ford DN, Lander DM, Leggio KB, Taylor M. Managing risk and uncertainty in complex capital projects. Quarterly Review of Economics and Finance. 2004;44(5 SPEC.ISS.):751–67.
6. Bellou M, Kokkinos P, Vantarakis A. Shellfish-Borne Viral Outbreaks: A Systematic Review. Food and Environmental Virology. 2013;5(1):13–23.
7. Maule AJ, Hockey GRJ, Bdzola L. Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. Acta Psychologica. 2000 Jun;104(3):283–301.
8. Stern PC, Fineberg H V., editors. Understanding Risk. Informing Decisions in a Democratic Society. Washington D.C.: National Academy Press; 1996.
9. Arvai JL. Rethinking of risk communication: lessons from the decision sciences. Tree Genetics & Genomes. 2007;3(2):173–85.
10. Morgan MG, Fischhoff B, Bostrom A, Atman CJ. Risk communication: A mental models approach. Cambridge University Press; 2002.
11. Bier VM. On the state of the art: risk communication to the public. Reliability Engineering & System Safety. 2001;71(2):139–50.
12. Slovic P. Perception of risk. Science. 1987;236(4799):280–5.
13. Ueland Ø, Gunnlaugsdottir H, Holm F, Kalogeras N, Leino O, Luteijn JM, et al. State of the art in benefit risk analysis: Consumer perception. Food and Chemical Toxicology. 2012;50(1):67–76.
14. Fischhoff B. Risk Perception and Communication Unplugged: Twenty Years of Process. Risk Analysis. 1995;15(2):137–45.
15. Bostrom A, Löfstedt RE. Communicating risk: Wireless and hardwired. Risk Analysis. 2003;23(2):241–8.
16. Breakwell GM. Risk communication: factors affecting impact. British Medical Bulletin. 2000 Jan;56(1):110–20.
17. Leiss W. Effective risk communication practice. Toxicology Letters. 2004;149(1–3):399–404.
18. Atman CJ, Bostrom A, Fischhoff B, Morgan MG. Designing Risk Communications - Completing and Correcting Mental Models of Hazardous Processes, Part I. Risk Analysis.

1994 Oct;14(5):779–88.

19. Bostrom A, Fischhoff B, Morgan MG. Characterizing mental models of hazardous processes: A methodology and an application to radon. Journal of Social Issues. 1992;48(4):85–100.

20. Bostrom A, Atman CJ, Fischhoff B, Morgan MG. Evaluating Risk Communications - Completing and Correcting Mental Models of Hazardous Processes, Part II. Risk Analysis. 1994;14(5):789–98.

21. Jones N, Ross H, Lynam T, Perez P, Leitch A. Mental models: an interdisciplinary synthesis of theory and methods. Ecology and Society. 2011;16(1):46.

22. Hagemann KS, Scholderer J. Hot Potato: Expert-Consumer Differences in the Perception of a Second-Generation Novel Food. Risk Analysis. 2009;29(7):1041–55.

23. Morgan G, Fischhoff B, Bostrom A, Lave L, Atman C. Communicating Risk to the Public. First, Learn what people know and believe. Environmental Science & Technology. 1992;26(11):2048–56.

24. Bruine de Bruin W, Bostrom A. Assessing what to address in science communication. Proc Natl Acad Sci U S A. 2013;110 Suppl:14062–8.

25. Fischhoff B. The sciences of science communication. Proceedings of the National Academy of Sciences of the United States of America. 2013;110 Suppl:14033–9.

26. Chakrapani V, Newman PA, Singhal N, Nelson R, Shunmugam M. “If It”s Not Working, Why Would They Be Testing It?’: mental models of HIV vaccine trials and preventive misconception among men who have sex with men in India. BMC Public Health. 2013;13(1):731.

27. Hagemann KS, Scholderer J. Consumer versus expert hazard identification: A mental models study of mutation-bred rice. Journal of Risk Research. 2007;10(4):449–64.

28. Skarlatidou A, Cheng T, Haklay M. What Do Lay People Want to Know About the Disposal of Nuclear Waste? A Mental Model Approach to the Design and Development of an Online Risk Communication. Risk Analysis. 2012;32(9):1496–511.

29. Vogt C, Schaefer M. Seeing things differently: Expert and consumer mental models evaluating combined oral contraceptives. Psychology & Health. 2012;27(12):1405–25.

30. Wagner K. Mental Models of Flash Floods and Landslides. Risk Analysis. 2007;27(3):671–82.

31. Zaksek M, Arvai JL. Toward improved communication about wildland fire: mental models research to identify information needs for natural resource management. Risk Analysis. 2004;24(6):1503–14.

32. McComas KA. Defining Moments in Risk Communication Research: 1996 to 2005. Journal of Health Communication. 2006;11(1):75–91.

33. Ropeik D. The Perception Gap: Recognizing and managing the risks that arise when we get risk wrong. Food and Chemical Toxicology. 2012;50(5):1222–5.

34. Niewöhner J, Cox P, Gerrard S, Pidgeon N. Evaluating the efficacy of a mental models approach for improving occupational chemical risk protection. Risk Analysis. 2004;24(2):349–61.

35. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. International Journal of Evidence-Based Healthcare. 2015;13(3):141–6.

36. Anaby D, Hand C, Bradley L, DiRezze B, Forhan M, DiGiacomo A, et al. The effect of the environment on participation of children and youth with disabilities: a scoping review. Disability and Rehabilitation. 2013;35(19):1589–98.

37. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology. 2005;8(1):19–32.

38. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implementation Science. 2010;5(1):69.

39. Giustini D, Boulos MNK. Google Scholar is not enough to be used alone for systematic reviews. Online Journal of Public Health Informatics. 2013;5(2):214.

40. Bajpai AK, Davuluri S, Haridas H, Kasliwal G, Deepti H, Sreelakshmi KS, et al. In search of the right literature search engine(s). Nature Precedings.

41. Beckmann M, Wehrden H. Where you search is what you get: literature mining: Google Scholar versus Web of Science using a data set from a literature search in vegetation science. Journal of Vegetation Science. 2012;23(6):1197–9.

42. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Medical Informatics and Decision Making. 2007;7(1):16.

43. Dunk MJ, Owen A, McMath SM, Arikans J. Remediation of polluted surface water outfalls–customer communication and changes in behaviour. Water and Environment Journal. 2012;26(2):191–9.

44. Tagashira N, Senda Y. What information should be provided in communications on biomass power generation? Applied Energy. 2011;88(7):2519–29.

45. Thorne SE, Linda V, Cory MB, Sarah L. Developing consumer-focused risk communication strategies related to food terrorism. International Journal of Food Safety, Nutrition and Public Health. 2011;4(1):45–62.

46. Cone J, Rowe S, Borberg J, Stancioff E, Doore B, Grant K. Reframing Engagement Methods for Climate Change Adaptation. Coastal Management. 2013;41(4):345–60.

47. Cone J, Winters KW. Planning and evaluating science videos to aid in understanding science and decisions.

48. Maharik M, Fischhoff B. Risk knowledge and risk attitudes regarding nuclear energy sources in space. Risk Analysis. 1993;13(3):345–53.

49. Vogt C, Schaefer M. Knowledge matters - Impact of two types of information brochure on contraceptive knowledge, attitudes and intentions. European Journal of Contraception and Reproductive Health Care. 2012;17(2):135–43.

50. De Bruin WB, Guevenc U, Fischhoff B, Armstrong CM, Caruso D. Communicating About Xenotransplantation: Models and Scenarios. Risk Analysis. 2009 Aug;29(8):1105–15.

51. Downs JS, Murray PJ, de Bruin W, Penrose J, Palmgren C, Fischhoff B. Interactive video behavioral intervention to reduce adolescent females' STD risk: A randomized controlled trial. Social Science & Medicine. 2004;59(8):1561–72.

52. Cousin M-E, Dohle S, Siegrist M. The impact of specific information provision on base station siting preferences. Journal of Risk Research. 2011;14(6):703–15.

53. Read D, Morgan MG. The efficacy of different methods for informing the public about the range dependency of magnetic fields from high voltage power lines. Risk Analysis. 1998;18(5):603–10.

54. Galada HC, Gurian PL, Corella-Barud V, Perez FG, Velazquez-Angulo G, Flores S, et al. Applying the mental models framework to carbon monoxide risk in northern Mexico. Pan American Journal of Public Health. 2009 Mar;25(3):242–53.

55. Longstaff HA. Frankenfish, Monsatan, and Killer Canola Making Good Decisions in the Midst of Environmental Controversy. Journal of Environmental Informatics. 2005;6(1):33–45.

56. Fleishman LA, De Bruin WB, Morgan MG. Informed Public Preferences for Electricity Portfolios with CCS and Other Low-Carbon Technologies. Risk Analysis. 2010;30(9):1399–410.

57. Guy S, Kashima Y, Walker I, O'Neill S. Investigating the effects of knowledge and ideology on climate change beliefs. European Journal of Social Psychology. 2014;44(5):421–9.

58. Shi J, Visschers VHM, Siegrist M, Arvai J. Knowledge as a driver of public perceptions about climate change reassessed. Nature Climate Change. 2016;(April).

59. Tobler C, Visschers VHM, Siegrist M. Consumers' knowledge about climate change. Climatic Change. 2012;114(2):189–209.

60. Cao Y, McGill WL. LinkIT: A Ludic Elicitation Game for Eliciting Risk Perceptions. Risk Analysis. 2013;33(6):1066–82.

61. Fitzpatrick-Lewis D, Yost J, Ciliska D, Krishnaratne S. Communication about environmental health risks: A systematic review. Environmental Health. 2010;9(67):1–15.

62. Frewer LJ, Fischer ARH, Brennan M, Banati D, Lion R, Meertens RM, et al. Risk/Benefit Communication about Food - A Systematic Review of the Literature. Critical Reviews in Food Science and Nutrition. 2015;

63. Zhang M, Liu GL. The effects of consumers' subjective and objective knowledge on perceptions and attitude towards genetically modified foods: objective knowledge as a determinant. International Journal of Food Science and Technology. 2015;50(5):1198–205.

64. Pieniak Z, Aertsens J, Verbeke W. Subjective and objective knowledge as determinants of organic vegetables consumption. Food Quality and Preference. 2010;21(6):581–8.

65. Seigo SL, Dohle S, Diamond L, Siegrist M. The effect of figures in CCS communication. International Journal of Greenhouse Gas Control. 2013;16:83–90.

66. Krimsky S. Risk communication in the internet age: The rise of disorganized skepticism. Environmental Hazards. 2007;7(2):157–64.

67. Rutsaert P, Regan Á, Pieniak Z, McConnon Á, Moss A, Wall P, et al. The use of social media in food risk and benefit communication. Trends in Food Science & Technology. 2013;30(1):84–91.

68. Veil SR, Buehner T, Palenchar MJ. A Work-In-Process Literature Review: Incorporating

Social Media in Risk and Crisis Communication. Journal of contingencies and crisis management. 2011;19(2):110–22.

69. Morss RE, Demuth JL, Bostrom A, Lazo JK, Lazrus H. Flash Flood Risks and Warning Decisions: A Mental Models Study of Forecasters, Public Officials, and Media Broadcasters in Boulder, Colorado. Risk Analysis. 2015;

70. Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, et al. Research synthesis: AAPOR report on online panels. Public Opinion Quarterly. 2010;74(4):1–71.

71. Scammell MK. Qualitative environmental health research: an analysis of the literature, 1991 to 2008. Environmental Health Perspectives. 2010;118(8):1146–54.

72. Cousin M-E, Siegrist M. Risk perception of mobile communication: a mental models approach. Journal of Risk Research. 2010;13(5):599–620.

73. Rosenthal R. The file drawer problem and tolerance for null results. Psychological Bulletin. 1979;86:638–41.

TABLES

Table I. Search engine selection: the number of references that cited each of the four key sources. Searches run February 2014. *These are a 'calculated set of PubMed citations closely related to the selected article(s) retrieved using a word weight algorithm'.

Key sources	Search engine and number of times each source was cited		
	Google Scholar	Web of Science	PubMed
Morgan *et al*. 2002[10]	812	Item not found	Item not found
Atman *et al*. 1994[18]	144	62	103*
Bostrom *et al*. 1994[20]	106	55	93*
Bostrom *et al*. 1992[19]	275	129	Item not found

Table II. Search strategy used to identify references for consideration

Search type	Search date	Source(s)	Publication dates	Search terms
Standard	4 February 2014	Web of Science, Google Scholar	All	"mental models approach" "risk communication"
Forward citation for four sources in Table I	4 February 2014	As above	All	Not applicable
Hand searching	19 February 2014	Morgan *et al*. 2002	All (up to source publication date of 2002)	Not applicable
Hand searching	26-27 February 2014	Reference lists of the 12 included papers	All	Not applicable

Table III. Inclusion criteria applied to search results.

Inclusion criterion	Detail
Written in English	References not written in English were excluded.
Complete reference	References should be sufficiently complete to allow source identification. Incomplete references were excluded.
Reference type	Must be peer-reviewed journal article. Book chapters, theses, conference proceedings, reports etc. were excluded.
Topic	Studies must report the evaluation of a communication developed in line with the MMARC, and if not were excluded.

Table IV. Summary of MMARC evaluation studies included in this review. *How the MMARC communication was developed: A=after comparing expert and public mental models; B=with expert input; C=in line with the MMARC. 1=quantitative methods; 2=qualitative methods. + positive; - negative; +/- mixed; = no difference.

Study description			Participants (target audience)	Intervention(s) (MMARC communication)	Comparison(s) (control)	Outcome(s)		Process		
Authors; year; country	Original study author?	Risk topic				Measures	Effect of MMARC	MMARC?*	Methods	Randomisation?
Atman *et al*. 1994 USA[18]	Y	Radon	n/a	Brochures	7 brochures	Communication coverage[1]	=	A	Desk-based comparison of each communication	n/a
						Communication structure[2]	+			
Bostrom *et al*. 1994 USA[20]	Y	Radon	Convenience: undergraduate students (social science communications)	Brochures	1 brochure; 1 'filler task'	Knowledge[1]	+	A	Multiple evaluation methods: think-aloud protocols, multiple choice test, true/false test, open-ended recall questions, problem-solving questions	Y
						Positive/negative comments[2]	+			
Cousin *et al*. 2011 Switzerland[52]	N	Mobile phone masts	Public convenience sample (recruited at a variety of locations)	Booklet	1 neutral text; 1 newspaper article	Knowledge[1]	+	C	Between subject design experiment; data collected during interviews	Y
						Perception[1]	+			
						Mast decision evidence-based[1]	+			
De Bruin *et al*. 2009 USA[50]	Y	Xenotrans plantation	Primarily students (participants recruited at a university)	Graphic only; text only; graphic and text, written on sheets of paper	None. (Compared MMARC communication formats.)	Depth of understanding (knowledge)[1]	+	B	Between subject design experiment; open-book written test	Y
						Ease of processing[1]	=			
						Xenotransplantation evaluation[1]	+			
Downs *et al*. 2004 USA[51]	Y	Sexually transmitted disease	Urban adolescent girls	Video	1 book; 1 brochure	Knowledge[1]	=	A	Longitudinal experiment with a baseline assessment and 3 subsequent data collection points	Y
						Behaviour[1]	+			
						Disease acquisition[1]	+			
Fischman *et al*. 2010 USA[56]	Y	Energy technologies	Public convenience sample (recruited from community groups)	Brochures	None. (Compared to 50% correct answers).	Knowledge[1]	+	B	Participants received the communications prior to assessment.	N
						Technology ranking[1]	=			

									Assessment included individual written and group exercises	
Calada *et al*. 2009 Mexico[54]	N	Carbon monoxide	Public convenience sample (from target population)	Comic book	None (Pre post comparison.)	Knowledge[1] Attitude[1] Willingness to change behaviour[1]	+ + +	A	Pretest posttest survey of participants	n/a
Longstaff 2005 Canada[55]	N	Fish diets	Public convenience sample (from five locations in Vancouver)	Flow chart; case study; FAQs	None. (Compared MMARC communication formats.)	Knowledge[1] Confidence[1] Acceptance[1] Purchase intention[1]	+ + +/- +/-	B	Experimental survey comprising pretest posttest assessment.	N
Maharik and Fischhoff 1993 USA[48]	Y	Nuclear energy in space	Participants drawn from a nature society	Brochure	1 brochure; 1 unspecified	Knowledge[1] Attitude[1]	+ =	C	Posttest with participants completing a series of knowledge and attitude assessments	N
Niewöhner *et al*. 2004 Uk[34]	N	Workplace chemicals	Employees from target population	Text	1 leaflet; 1 brochure	Comprehension (knowledge)[2]	+	A	Postal questionnaires; discussion groups; verbal protocols	N
Read and Morgan 1998 USA[53]	Y	Electrical fields	Participants from university community (>80% students)	Graphic; picture (each with varying amounts of text)	no communication	Knowledge[1]	+	C	Pretest posttest with participants providing written responses	N
Vogt and Schaefer 2012 Germany[49]	N	Contraception	Participants from target population (young women)	Brochure	1 brochure	Knowledge[1] Attitude[1] Intention[1]	= = =	A	Double-blind, parallel-group randomised longitudinal trial with participants providing written responses	Y

Table V. Detail of knowledge measures applied in each study. Direct knowledge relates specifically to the study topic, whereas indirect/proxy knowledge is a measure of a participant's broader scientific literacy,

Study	Knowledge measures	
	Direct	**Indirect / proxy**
Atman *et al*. 1994[18]	n/a – did not assess participants' knowledge	
Bostrom *et al*. 1994[20]	58-item true/false test addressing both common radon misconceptions and expert concepts 7-item multiple choice test (including radon health effects, detection and mitigation)	None
Cousin *et al*. 2011[52]	13-item test relating to the technical functionality of mobile communications (true, wrong, don't know)	Two 3-item scales measuring mobile phone and base station health concerns
De Bruin *et al*. 2009[50]	14-item open-ended knowledge test of xenotransplantation cause and effect variables 7-item knowledge test with a 7-point response scale assessing direction of influence of xenotransplantation variables 5-item test assessing probability of several health outcomes following xenotransplantation (0-100%) Scenario (participants asked to draft a scenario based on information they were provided as part of the assessment)	None
Downs *et al*. 2004[51]	40-item true/false test of reproductive health 15-item multiple choice test relating to eight diseases	None
Fleishman *et al*. 2010[56]	15-item (true/false) test about energy technology, focussing on misconceptions	None
Galada *et al*. 2009[54]	22-item test including carbon monoxide production, health effects and mitigation	None
Longstaff 2005[55]	Multiple choice questions relating to salmon production misconceptions (number of items not provided)	None
Maharik and Fischhoff 1993[48]	66-item test with a 5-point response scale (true, maybe true, don't know, false, maybe false) covering expert model concepts and misconceptions	None
Niewöhner e*t al*. 2004[34]	Questionnaires assessing respondents' knowledge background (unclear whether this was actionable or indirect)	

	Discussion groups (user evaluation sessions) assessing participants understanding of chemical risk (unclear whether this was actionable or indirect) Verbal protocols assessing comprehension of the communication (unclear whether this was actionable or indirect)	
Read and Morgan 1998[53]	2 knowledge measures assessing respondents' knowledge of magnetic field strength and distance	None
Vogt and Schaefer 2012[49]	39-item knowledge test assessing participant's knowledge of the benefits and risks of combined oral contraceptives	None

FIGURES

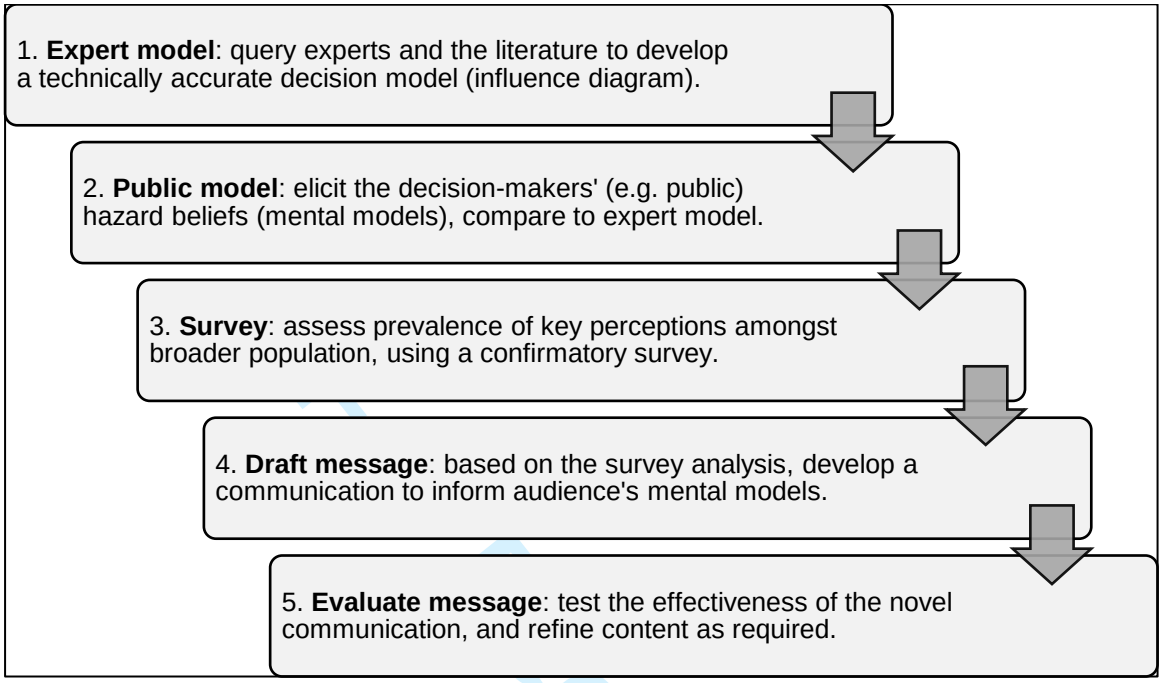

Figure 1. Outline of the Mental Models Approach to Risk Communication, adapted from Morgan *et al*. 2002.[10]

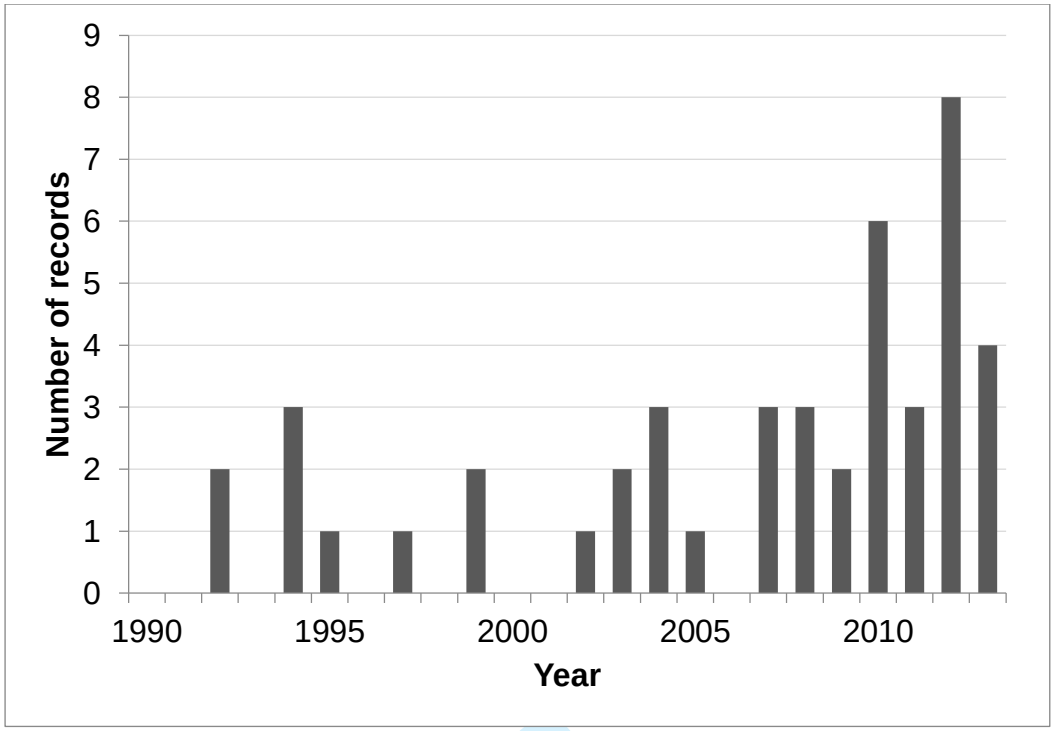

Figure 2. MMARC publications by year. Results from searches on Web of Science database on 4 July 2014 with search term "mental models approach" in topic field. This figure was intended to provide an indication of the pattern of MMARC publications over time, rather than an exhaustive assessment of the literature (importantly the searches described in Section 2.1 applied a broader range of searches).

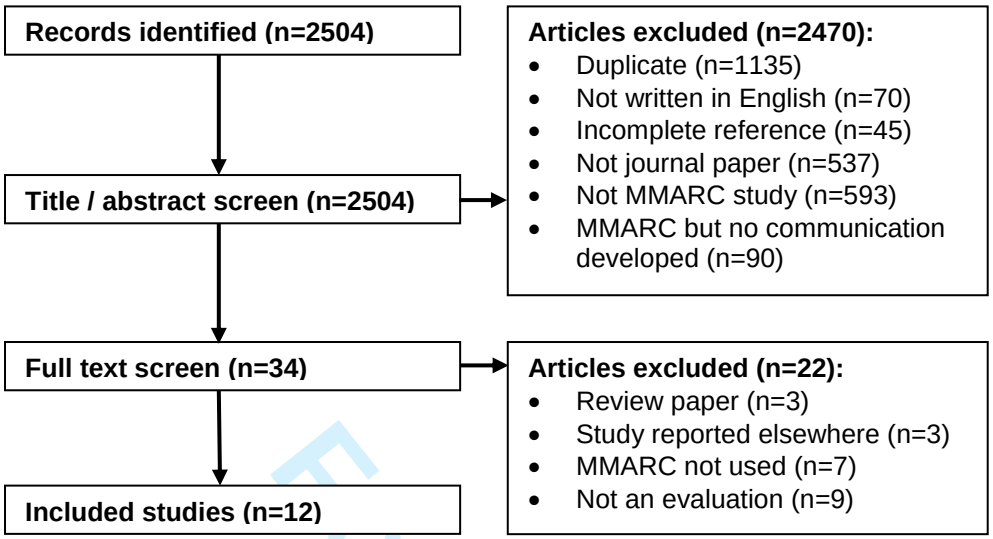

Figure 3. Screening procedure for search results.