The case for development of an E-cigarette Ontology (E-CigO) to improve quality, efficiency and clarity in the conduct and interpretation of research

Sharon Cox[1], Janna Hastings[1], Robert West[1], Caitlin Notley[2]

1 University College London, University of London
2 University of East Anglia

Abstract

There is an urgent need for more clarity and consistency in the use of terms in the field of e-cigarette research. Ontologies are computer artifacts that are increasingly widely used in science to represent knowledge in terms of uniquely defined entities and their relationships with other entities. These are constructed in a way that promotes clarity of thinking, much more effective searching, inference and interoperability across domains of study. We are constructing an E-cigarette Ontology (ECig-O) covering all the types of entity that are referred to in reports of e-cigarette research. It is part a larger Addiction Ontology (AddictO) covering all aspects of addiction.

Definitions

Disclosure of potential competing interest
Defined by Robert West

Sharon Cox[1], **Janna Hastings**[2], **Robert West**[2], **Caitlin Notley**[3]

[1] *Centre for Addictive Behaviours Research, School of Applied Sciences, London South Bank University*

[2] *Department of Behavioural Science and Health, University College London*

[3] *Norwich Medical School, Faculty of Medicine and Health Sciences, University of East Anglia*

The rapid growth in the popularity of e-cigarettes in many countries has sparked controversy within the tobacco control community, with some taking the view that e-cigarettes provide an opportunity to save thousands if not millions of lives if smokers switch to them[1], while others take the position that e-cigarettes are acting as a gateway to smoking in young people, pose serious health risks, and that their use detracts from, rather than helps with, quitting smoking[2].

Currently, e-cigarette research findings are being interpreted selectively as supporting or opposing already established positions. This can be achieved by ignoring contradictory evidence, by being willing to make generalisations where these support a given view but not if they do not, or by applying different quality criteria to views one supports than those one opposes. This has important implications for policy and practice, which are susceptible to influence from media reporting that may foreground sensationalist interpretations of research findings. Several jurisdictions in the US have

decided to ban the sale of e-cigarettes[3]. A recent cluster of illnesses and deaths in the US, which the US Food and Drug Administration now acknowledges are mostly likely arising from use of illicit products, had been attributed by some authorities such as the US Centres for Disease Control to the use of e-cigarettes per se before the evidence was made available. This contrasts with the regulatory approaches taken by the UK, Canada and New Zealand[4]. The field of tobacco control appears to be becoming increasingly polarised with no sign of the accumulating evidence leading to a rapprochement[5].

These tendencies, as well as gaps in the research literature, may underpin reasons for adopting restrictive policies[6]. More broadly, from a public health perspective, misrepresenting evidence to the general public can be very damaging. Examples of this in the history of public health evidence reporting are currently being played out with rates of childhood vaccinations dropping across Europe since misreporting of the potential harms of the MMR vaccine[7].

The importance and urgency of reducing tobacco-related disease, the popularity of e-cigarettes where they are available as a cessation aid, and the potential to significantly reduce the harm caused by combustible tobacco smoke, have the consequence that it is extremely important to undertake research to address the key issues of contention in such a way that conclusions are well-founded. The findings from such research need to be interpreted and communicated accurately and objectively in order for policy to be grounded in correctly interpreted evidence and for clinicians to give correct advice. There is an urgent need for consistency and transparency in defining concepts, methods for testing hypotheses, and reporting of study findings, including open access preregistration of hypotheses and planned analyses[8].

Like many other fields in addiction and health research, e-cigarette research encompasses experts from a wide range of disciplines. However, with this diversity of scientific disciplines comes variation in the use of language, terminology and the meaning ascribed to terms. Research priorities also differ. Though not all difference can be attributed to use of language, divergence nonetheless impacts both individual level understanding and broader scientific knowledge. As an example, how e-cigarettes are used – frequency and duration – relates to a range of behaviours that are important for measuring impact; but the umbrella term 'e-cigarette use' is ambiguous and has been used by researchers to mean anything from experimentation through to daily and exclusive use. Indeed, not being sufficiently accurate on the definition of 'use' can lead to inflation of estimates of use by young people in whom experimentation is much higher than daily use[9]. Other examples also exist; for example, what constitutes a tobacco product, or whether even independent e-cigarette companies are part of the tobacco industry, is an area of debate globally, with discernible lack of consensus and much confusion. Nonetheless, whether an e-cigarette is or isn't defined as a tobacco product has significant consequences for policy, through taxation, regulation and product availability, impacting the most deprived smokers if not carefully understood by policy makers.

Interpreting findings and building a consensus is difficult when terms are used with different meanings. This is further hampered when there is limited knowledge by researchers of the core constructs relating to e-cigarette use and the relationships between them, and it can be damaging when there is an interaction between this limited knowledge and media interest. Poor or inconsistent understanding of terminology limits the extent to which some findings can contribute to the development of new theoretical and applied knowledge of e-cigarette use.

In order to resolve some of these issues we are developing an e-cigarette 'ontology' (E-CigO). The aim of E-CigO is to provide a coherent and systematic way of defining terms and representing research methods and findings in the field. It will enable much more accurate, complete and detailed searching of relevant literature than is currently possible, and provide a basis for evidence integration and interpretation. This in turn will help i) identify sources of disagreement in the field, ii) prioritise topics for research funding, iii) provide guidance to researchers to improve the reporting and interpretation of findings, and iv) help the end user obtain an accurate impression of research findings taking account of possible sources of bias.

Formal ontologies are becoming increasingly widely used across science and medicine[10][11][12][13][14]. For example, the Gene Ontology[10] is an ontology for the biological domain that standardises the description of gene functions in a way that allows unified annotations across species and has revolutionised large-scale interpretation of genomic functional data. The Cochrane collaboration has developed its PICO ontology to capture key components of clinical trials and their reporting, including in the field of addiction[15]. The New England Journal of Medicine published an article showing how clinical ontologies can provide a basis for precision medicine, the benefits of which better serve medical decision making[11]. This approach could equally be of great value in nicotine and tobacco research.

In information science, ontologies represent knowledge in terms of entities and their properties. Each entity is given a clear, unambiguous definition. Properties consist of relationships between entities, and different types of relationship are also given clear, unambiguous definitions. Thus, a body of knowledge can be expressed as a simple yet comprehensive set of interconnected, clearly defined entities and relations, in turn supported by a computational framework for knowledge representation based on logic-based artificial intelligence technology. The ontology we are developing will enable accurate interpretation of data from multiple sources through the explicit definition of terms and relationships. Examples of this may include: the clear delineation of nicotine and tobacco containing products; understanding how nicotine strength relates to user behaviour; or understanding user behaviour in context. Users will be able to query the ontology as a searchable system. Users will also be able to identify how entities are being applied by e-cigarette researchers across different disciplines, including where divergence exists.

Having a repository of knowledge about the constructs, models and theories that are relevant for an entire field allows for sophisticated search and automated inferences. The logical model on which the ontology is built is able to detect inconsistencies and contradictions automatically.

Annotations are used to link the entities defined in the ontology to the research findings and broader literature about those entities in a way that enables systematic evidence synthesis and facilitates further research design, scientific communication and open debate. This will enable all literature that bears on a given contested question to be retrieved in one query. Thus, the mappings of ontology entities to findings and literature can be used to automatically highlight when a given finding is inconsistent with the remainder of the available evidence and suggest when errors of interpretation might have arisen.

Taken together, the result is that the ontology serves as an externalised understanding of the conceptual progress in the

field, creating a hub around which conceptual disagreements and disputes can be identified and resolved.

The ontology is a 'live' representation of the field, requiring ongoing and transparent scrutiny, adjustment and additions as the field evolves. Users will also have the opportunity to feed back to the core research team and suggest alterations as well as additional terms, relationships and annotations. These suggestions will be reviewed and decisions made based on the strength and validity of the suggested addition or amendment, with feedback (where necessary) from a team of experts.

Importantly, this project does not seek to dictate the language of users of vaping products, but instead the expectation is that real-world user behaviour will inform the ontology. This is however, a resource for use by academics, policy makers and clinicians, seeking to resolve ambiguity and dispute through scientific and evidence-based definitions, and clear understanding of constructs. Ontologies may be perceived as restrictive, limiting academic creativity, but alternatively we suggest that they simply promote clarity and interoperability. Consensus amongst the scientific community will be clearly evident through the ontology, and where there is divergence this will direct the field to reach consensus or to acknowledge the need for divergence in some instances. The e-cigarette ontology will address the current divide in tobacco control through facilitating the use of clear and unambiguous shared terminology and supporting an objective interpretation of evidence.

Disclosure of potential competing

interests

- SC has provided expert consultancy to UK Life Insurance providers on issues relating to smoking cessation and the use of e-cigarettes.
- CN and JH have no interests to declare.
- RW undertakes research and consultancy for and receives travel funds and hospitality from manufacturers of smoking cessation medications (Pfizer, GlaxoSmithKline and Johnson and Johnson).

All authors declare no financial links with tobacco companies or e-cigarette manufacturers or their representatives.

References

1. ^David T Levy, Ron Borland, Eric N Lindblom, Maciej L Goniewicz, Rafael Meza, Theodore R Holford. (2017).Potential deaths averted in USA by replacing cigarettes with e-cigarettes. Tob Control, vol. 27 (1), 18-25. doi:10.1136/tobaccocontrol-2017-053759.

2. ^*Sara Kalkhoran, Stanton A Glantz. (2016). E-cigarettes and smoking cessation in real-world and clinical settings: a systematic review and meta-analysis. The Lancet Respiratory Medicine, vol. 4 (2), 116-128. doi:10.1016/s2213-2600(15)00521-4.*

3. ^*Howard K. Koh, Clifford E. Douglas. (2019). The San Francisco Ban and the Future of e-Cigarettes. JAMA, vol. 322 (16), 1540. doi:10.1001/jama.2019.16181.*

4. ^*Benjamin R. Brady, Jennifer S. De La Rosa, Uma S. Nair, Scott J. Leischow. (2019). Electronic Cigarette Policy Recommendations: A Scoping Review. am j health behav, vol. 43 (1), 88-104. doi:10.5993/ajhb.43.1.8.*

5. ^*Sharon Cox, Lynne Dawkins. (2018). Global and local perspectives on tobacco harm reduction: what are the issues and where do we go from here?. Harm Reduct J, vol. 15 (1). doi:10.1186/s12954-018-0239-5.*

6. ^*Amy Lauren Fairchild, Ronald Bayer, Ju Sung Lee. (2019). The E-Cigarette Debate: What Counts as Evidence?. Am J Public Health, vol. 109 (7), 1000-1006. doi:10.2105/ajph.2019.305107.*

7. ^*World Health Organisation. Vaccine-preventable diseases: monitoring system. 2019 global summary. 2019.*

8. ^*Marcus R. Munafò, Robert West. (2019). E-cigarette research needs to adopt open science practices to improve quality. Addiction, vol. 115 (1), 3-4. doi:10.1111/add.14749.*

9. ^*Robert West, Jamie Brown, Martin Jarvis. (2019). Epidemic of youth nicotine addiction? What does the National Youth Tobacco Survey reveal about high school e-cigarette use in the USA? (Preprint). Qeios. doi:10.32388/745076.3.*

10. a, b *Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry. (2000). Gene Ontology: tool for the unification of biology. Nat Genet, vol. 25 (1), 25-29. doi:10.1038/75556.*

11. a, b *Melissa A. Haendel, Christopher G. Chute, Peter N. Robinson. (2018). Classification, Ontology, and Precision Medicine. N Engl J Med, vol. 379 (15), 1452-1462. doi:10.1056/nejmra1615014.*

12. ^*Robert West, John Marsden, Janna Hastings. (2019). Addiction Theories and Constructs: a new series. Addiction, vol. 114 (6), 955-956. doi:10.1111/add.14554.*

13. ^*Susan Michie, James Thomas, Marie Johnston, Pol Mac Aonghusa, John Shawe-Taylor, Michael P. Kelly. (2017). The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implementation Sci, vol. 12 (1). doi:10.1186/s13012-017-0641-5.*

14. ^*Janna Hastings, Stefan Schulz. (2012). Ontologies for Human Behavior Analysis and Their Application to Clinical Data. doi:10.1016/b978-0-12-388408-4.00005-8.*

15. ^*Cochrane Collaboration. PICO Ontology.*