bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Science podcasts: analysis of global production and output from 2004 to 2018

Dr Lewis E MacKenzie*

Department of Chemistry, Durham University, Durham, UK. DH1 3LE.

https://orcid.org/0000-0002-8151-0525

Email*:* lewis.e.mackenzie@durham.ac.uk

Abstracts

Podcasts have emerged as a new decentralised global medium for science communication to the global public. However, despite their popularity and proliferation, there have been no studies of the production and dissemination of science podcasts. To address this need, this study identified 952 English language science podcasts and associated promotional websites. Key indicators of podcast production and output were classified from textual and visual data. It was found that between 2004 and 2010, the number of science podcasts grew linearly, but has subsequently grown exponentially. 65% of science were hosted by scientists and 77% were targeted to public audiences. 38% of science podcasts were created by independent producers, compared to 62% produced with an organisational affiliation. 34% of science podcasts focused on general science topics, but a diverse range of science podcast topics are available. Notably, chemistry is under-represented in science podcasts when compared to physics and biology. The USA and UK dominate English language science podcast production. Podcasts affiliated to organisations released more episodes compared to independent podcasts (median = 24 and 16 respectively). Only 24% of science podcasts had some form of overt supplementary income. It is anticipated that these results will inform future science communication strategy.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1. Introduction

Since 2004, podcasts have emerged as a decentralised medium for free and independent communication to global audiences. Podcasts are hosted on websites and provided to audiences via direct, on-demand audio and video downloads to devices such as personal computers, MP3 players, and smartphones.[1] Although podcasts are a decentralised medium, the primary database for podcasts is the '*iTunes*' podcast directory'.[2] Compared to conventional broadcast or print media, podcasts are convenient for listeners because they can be listened to/watched at any time.[3,4] Further, due to their global distribution, podcasts can cater for niche and highly specialised audiences; for example the '*This Week in Virology'* podcast, which serves the virology community.[5] Podcasts are also a convenient medium for content creators because podcasts may be created with widely available or low cost audio equipment.

Science podcasts have become a varied and abundant avenue for science communication, with hundreds of English language science podcasts currently available to the public. Science podcasts are produced by a wide variety of people, including professional scientists, media and broadcasting professionals, other professionals associated with host organisations, and amateur groups. Science podcasts are typically promoted online via websites and social media, but may also be promoted organisations that they are affiliated to, for example, scientific journals, universities, research institutions, podcast networks, or traditional media networks (e.g. magazines and radio stations). Science podcasts have also been targeted at a wide variety of audiences, including the public, children; students; and scientists/specialists. Podcasts are unconstrained by the style and time demands that typically restrict traditional broadcast TV and radio media. Consequently, many diverse styles of science podcasts have been created, including monologues, informal chat-shows, professional science-news, dramas, and comedy panel shows. Because of this varied and decentralised production, it could be said that there is a science podcast for everyone.

Despite the rise of podcasts as a medium for science communication, the peer-reviewed literature on podcasts to date has been rather limited in scope: all podcast studies have focused on the various pedagogical benefits of podcasts in education, factual retention, or promotion of science. Podcasts have been demonstrated to be a beneficial media improving scientific information uptake in students, medical patients, and the public.[6,7] Prakash et al., (2017) provide an excellent review of such studies.[8] However, to date, there have been no formal studies of the production of science podcasts. Thus, there is a large and fundamental gap in our knowledge in the rise of science podcasts, the fundamental demographics of the science podcast medium, and how science podcasts are used to communicate science globally.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This study provides the first quantitative insight into the global production and dissemination of science podcasts. This has been achieved by analysing 952 scientific podcasts for key production variables, including: science podcast topics, target audiences, hosts, organisational affiliations, supplementary income, number of episodes released, and overall podcast lifespan.

2. Results

952 science podcasts met the inclusion criteria for this study (see Section 3.3). The number of excluded podcasts was not recorded. Between 2004 and 2010, the total cumulative number of science podcasts available grew in a linear manner (see linear fit in Figure 1A, $R^2 = 0.99$). In contrast, between 2010 and 2018 the cumulative total number of available science podcast grew exponentially (see Figure 1A, $R^2 = 0.99$), rising to 952 podcasts by the sampling period (5th January – 5th February 2018). Prior to 2004, 11 science podcasts were available as internet radio shows before subsequent listing as podcasts

As of their individual sampling dates,[a] 46% of total science podcasts were 'active', meaning that they released an episode in the three months prior to their specific sampling date. Of the remaining 'inactive' podcasts, 14% released an episode between three to twelve months prior their sampling date, and 40% had been inactive for over a year prior to their sampling date (see Figure 1B).

The number of episodes released by each science podcast was found to be highly variable (see Figure 1C and Table 1); 33% of science podcasts produced fewer than 10 episodes, and 72% of science podcasts produced less than 50 episodes. The lifespan of podcasts (defined as the time between the first episode and most recent/last episode) shows a clear exponential decay trend (see Figure 1D). This exponential decay was only moderately well-fitted by a single term exponential fit ($R^2 = 0.92$), but a two-term exponential decay produced an excellent fit ($R^2 = 0.99$). The two-term exponential fit is indicative of a bimodal distribution of "short lived" and "long lived" podcasts; i.e. populations of podcasts with lifespans of a few months and several years respectively.

[a] The exact sampling date for each podcast is provided in the associated supplementary dataset.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

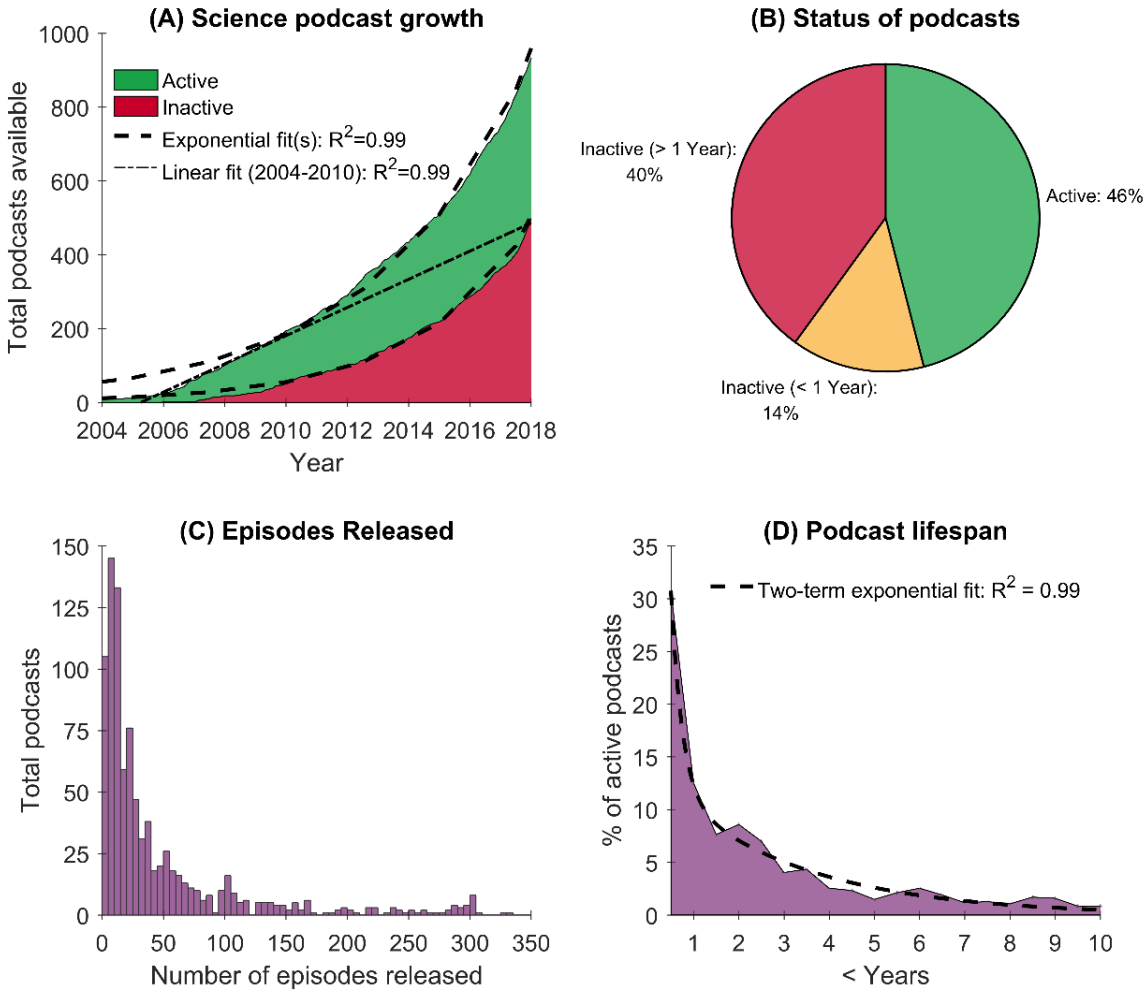

Figure 1: The growth and lifespan of science podcasts. (A) The total number of science podcasts shows linear growth between 2004 and 2010, followed by exponential growth to from 2010-2018. Dashed lines are lines of best fit (R^2 >0.99 in all cases). **(B)** The proportion of active/inactive science podcast shows as of the sampling period. **(C)** The total number of episodes of each podcast show available. Note that the x-axis is constrained to 350 episodes for clarity due to a small number of extreme outliers. **(D)** A clear exponential decay in the lifespan of podcasts is apparent. The dashed line is a two-term exponential decay fit (R^2 = 0.99). This two-term fit is indicative of a bimodal distribution of podcast lifespans, representing populations of short-lived and long-lived podcasts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1. The total number of episodes released by science podcasts

Number of Episodes released	Number of Podcasts	Percentage
1 Episode	25	2.6
≤ 10 Episodes	250	33.0
≤ 50 Episodes	685	72.0
≤ 100 Episodes	802	84.2
≤ 300 Episodes	913	95.9
> 300 Episodes	39	4.1
> 500 Episodes	17	1.8
> 1000 Episodes	5	0.5

Statistical Descriptor	Number of Episodes Released
Modal	10
Median	20
Mean	73

A wide variety of science podcast themes were recorded, with 66% of science podcasts themed around discipline-specific topics (see Figure 2A). Of note, 'Chemistry' seems to have been under-represented in comparison to the other traditional "physical sciences": Chemistry was the focus of only 3% of science podcasts, compared to 18% for 'Physics and Astronomy', and 14% for 'Biology'. 34% of science podcasts were categorised as 'General Science', i.e. science podcasts focusing on no single discipline-specific theme.

The majority of science podcasts (77%) have been targeted to public audiences, 16% were targeted towards scientists or specialists, and 6% were provided as academic lectures, research seminars/conferences, or as secondary education learning aids (see Figure 2B).

Nearly 2/3rds (65%) of science podcasts were hosted by 'scientists'; 10% were hosted by 'media professionals', 7% by 'other professionals', and 5% by 'amateurs' (see Figure 3A). Host categories could not be identified for 13% of science podcasts.

38% of science podcasts were produced independently, and 62% were produced with some explicitly acknowledged affiliation to an organisation (see Figure 3B). 'Professional Organisations' produced 17% of science podcasts; 'Universities' 14%; 'Conventional Media Networks' 13%; 'Other Research Bodies' 6%; 'Podcast Networks' 5%; 'Scientific Journals' 3%, and 'Amateur Organisations' 2%. How podcast affiliation, or lack thereof, affects various production outputs is explored further, later in this manuscript.[b]

57% of science podcasts did not follow a regular episode release schedule (see Figure 3C). The most popular release schedule was 'Weekly' (15%), followed by 'Monthly' (8%), and 'Fortnightly' (6%). Only 3% of science podcasts released more than one episode per week,

[b] See Figure 7 and Table 2.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and 1% released an episode daily. Only 2% of science podcasts explicitly acknowledged a seasonal release format, i.e. periods of scheduled episode releases followed by an extended period where no episodes are released.

Whilst podcasts can contain both audio and visual information, 87% of science podcast were audio-only, with the remaining 13% being video podcasts (so-called *vodcasts*) (see Figure 4A). 51% of science podcasts provided additional non-audio supplementary material in the form of show notes (e.g. hyperlinks, images, references, etc.) (see Figure 4B).

Global production of science podcasts to date is shown in Figure 5. 57% of the available English language science podcasts were produced in the United States of America (USA), 17% were produced in the United Kingdom (UK), 5% in Australia, 3% in Canada, and 1% in the Republic of Ireland. Other various countries produce a total of 7% of English language science podcasts. A country of production could not be identified for 10% of science podcasts.

76% of science podcasts were observed to have no overt supplementary income mechanisms and are thus seemingly independently financed by their producers (see Figure 6A). 'Advertising' was the least commonly utilised supplementary income mechanism (see Figure 6B), but it was common for science podcasts to mix 'Voluntary Donations', 'Merchandise', and 'Advertising' to various degrees.

The differences between 'Independent' science podcasts and 'Affiliated' science podcasts in relation to various production outputs is shown in Figure 7. In terms of podcast activity, there is only a marginal difference between the percentage of active 'Affiliated' and 'Independent' science podcasts (48% and 45% respectively) (see Figure 7A). However, a larger proportion of 'Independent' podcasts (84%) are targeted to the public, compared to 'Affiliated' podcasts (73%) (see Figure 7B). A slightly smaller proportion of 'Independent' podcasts (14%) are targeted towards 'scientist/specialist' audiences compared with 'Affiliated podcasts' (17%) (see Figure 7B). Nearly all science podcasts billed as academic seminars, student lectures, or secondary education aids are produced as 'Affiliated' podcasts (see Figure 7B). Roughly 75% of both 'Independent' and 'Affiliated' podcasts had no overt supplementary income (see Figure 7C). However, a considerably greater proportion of 'Independent' podcasts solicited for 'Voluntary Donations' and sold 'Merchandise' (see Figure 7C), and 'Advertising' was much more prevalent for 'Affiliated' podcasts (25%) compared to 'Independent' podcasts (11%) (see Figure 7C).

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

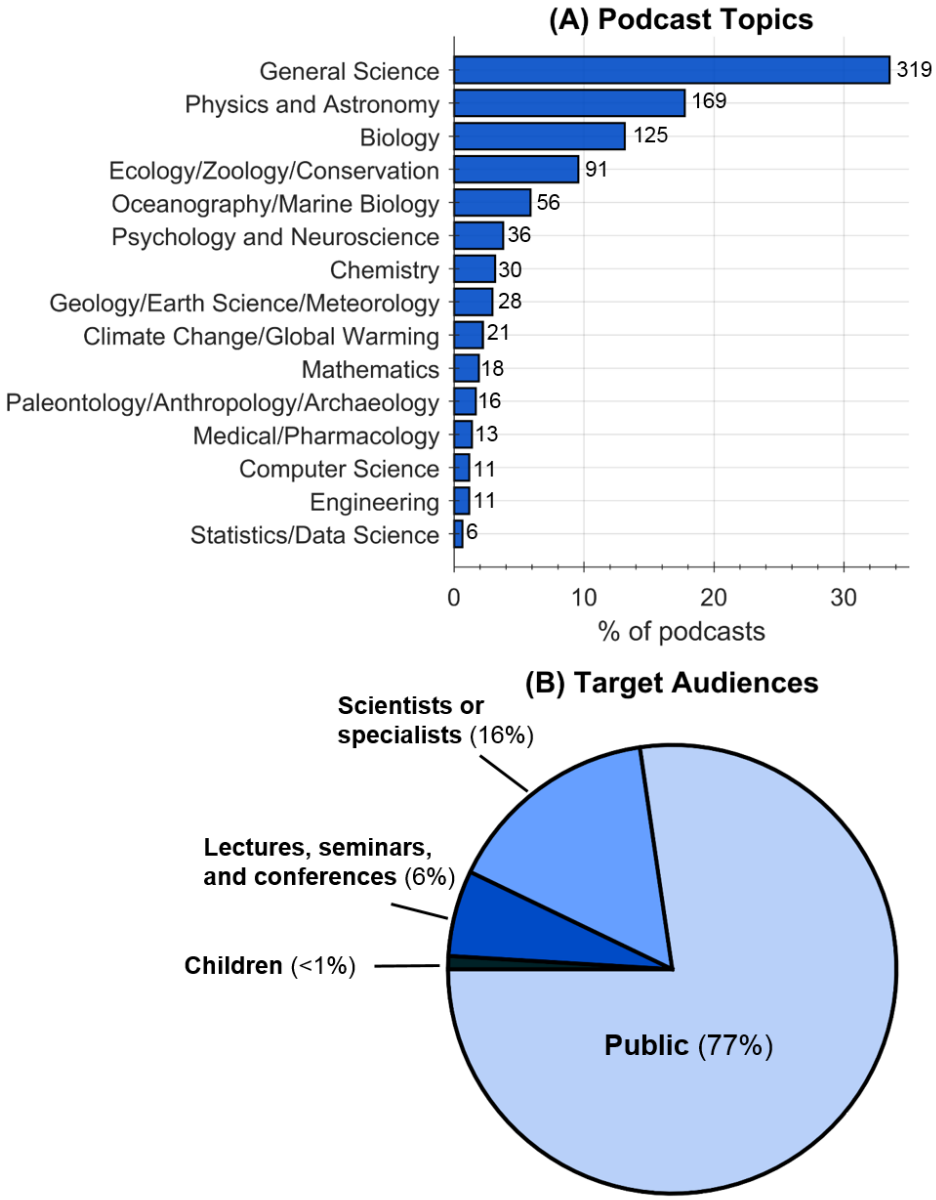

Figure 2: What are the scope and aims of science podcasts? (A) The proportion of science podcasts dedicated to various scientific topics. **(B)** The target audiences of science podcasts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

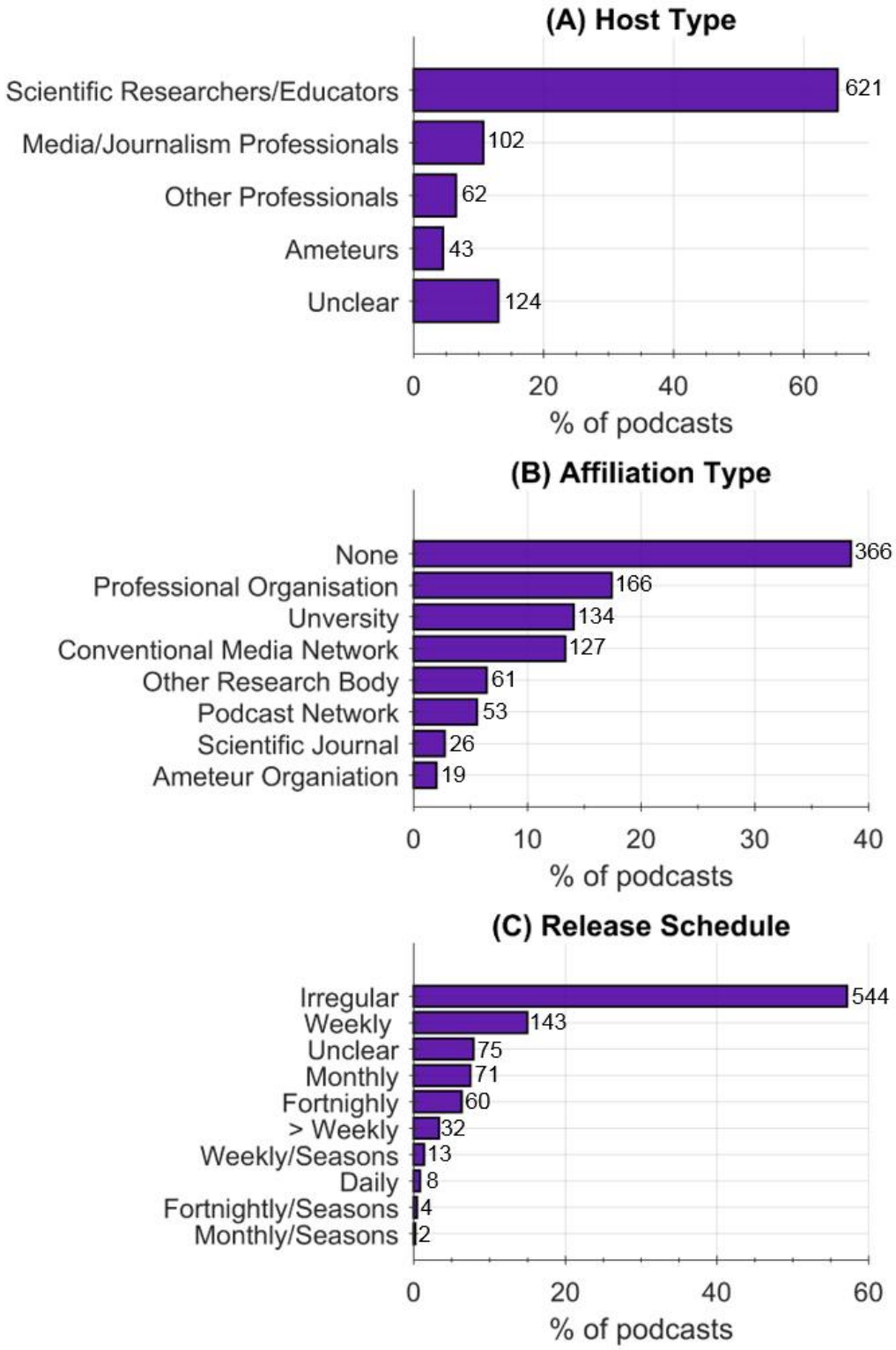

Figure 3: Who produces science podcasts? (A) The backgrounds of science podcast hosts. **(B)** The organisational affiliations of science podcasts. **(C)** The release schedule of science podcasts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

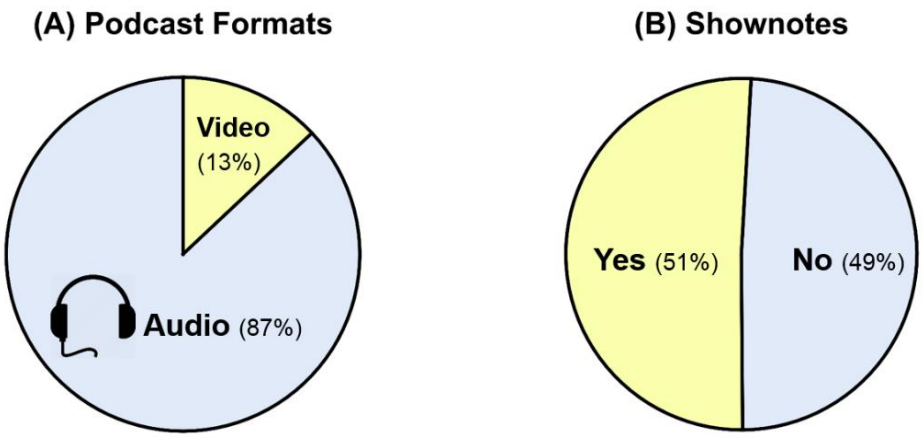

Figure 4: Adoption of non-audio media in science podcasts. (A) The proportion of audio-only science podcasts compared to video format science podcasts. **(B)** The usage of show notes by science podcasts.

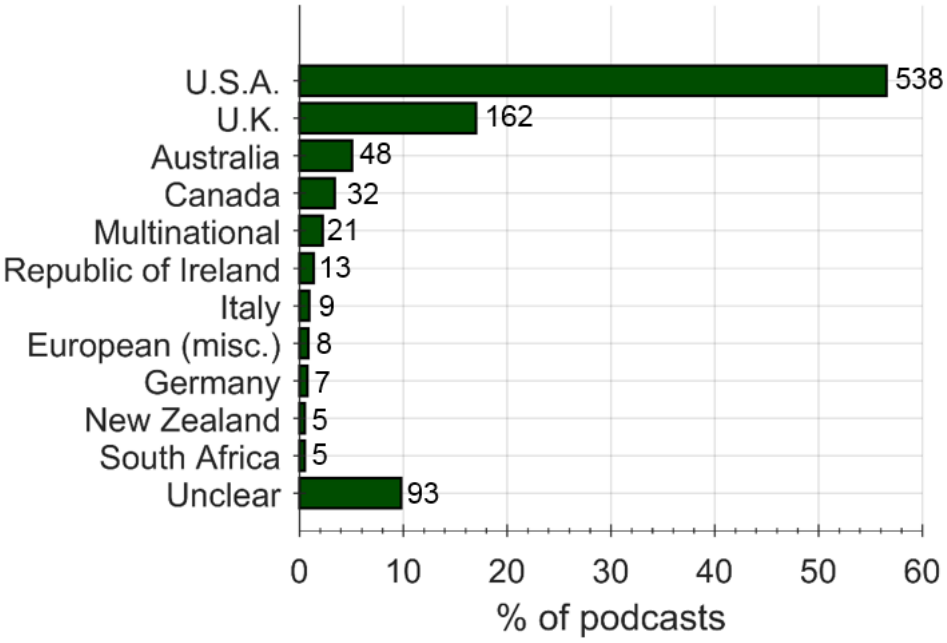

Figure 5: Global production of English language science podcasts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

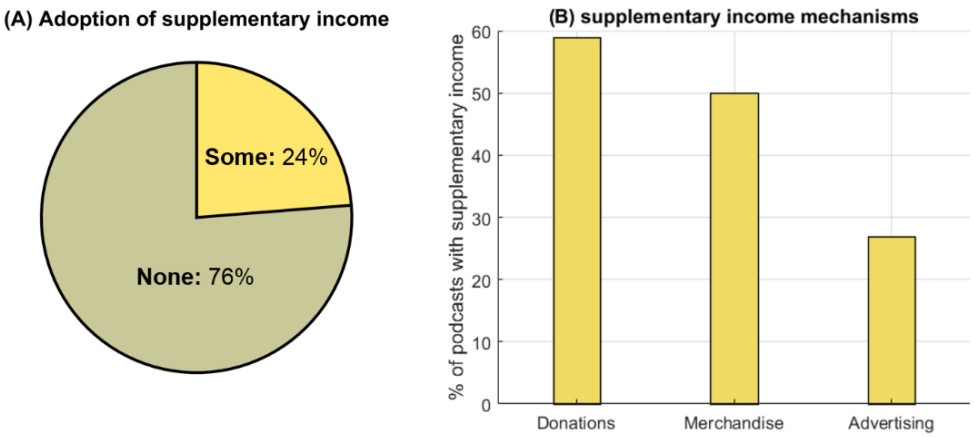

Figure 6: Do science podcasts generate overt supplementary income? (A) The proportion of podcasts with some supplementary income mechanism vs the proportion that have none. **(B)** The percentage of the subset of science podcasts with a supplementary income, that use each type of supplementary income mechanism. N.b. Some podcasts use more than one supplementary income mechanism.

The number of episodes released by 'Independent' and 'Affiliated' podcasts follows a broadly similar distribution (see Figure 7D); however, the distribution for 'Affiliated' podcasts is offset in favour of releasing a greater number of episodes in comparison to 'Independent' podcasts. In statistical terms, 'Affiliated' podcasts produced a median of 24 episodes, whereas 'Independent' podcast produce a median of 16 episodes.

The lifespan of both 'Independent' and 'Affiliated' podcast groupings is best-fitted by a two-term exponential. This indicates that both the groupings of 'Affiliated' and 'Independent' podcasts contain bimodal distribution of "short lifespan" and "long lifespan" podcast populations (see Figure 7D and Figure 7E). Extraction of fit parameters enables the estimation the podcast '*mean lifespan*' (τ), analogous to the concept of '*mean lifespan'* in radioactive decay; i.e. τ is the elapsed time span in which, 50% of the podcasts in a population become inactive. Estimated values for (τ) are summarised Table 2. For the population of 'Independent' podcasts, τ = 4.5 months, compared to τ = 3 months for the population of 'Affiliated' podcasts. This indicates that, on average, 'Independent' podcasts are active for a greater lifespan than 'Affiliated' podcasts in the short-term. However, for long-term trends, 'Affiliated' podcasts have a greater τ = 5 years and 6 months, compared to τ = 4 years and 4 months for the population of 'Independent' podcasts. This shows that, for the population of podcasts that run for a year or more, 'Affiliated' podcasts will, on average, last longer than 'Independent' podcasts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

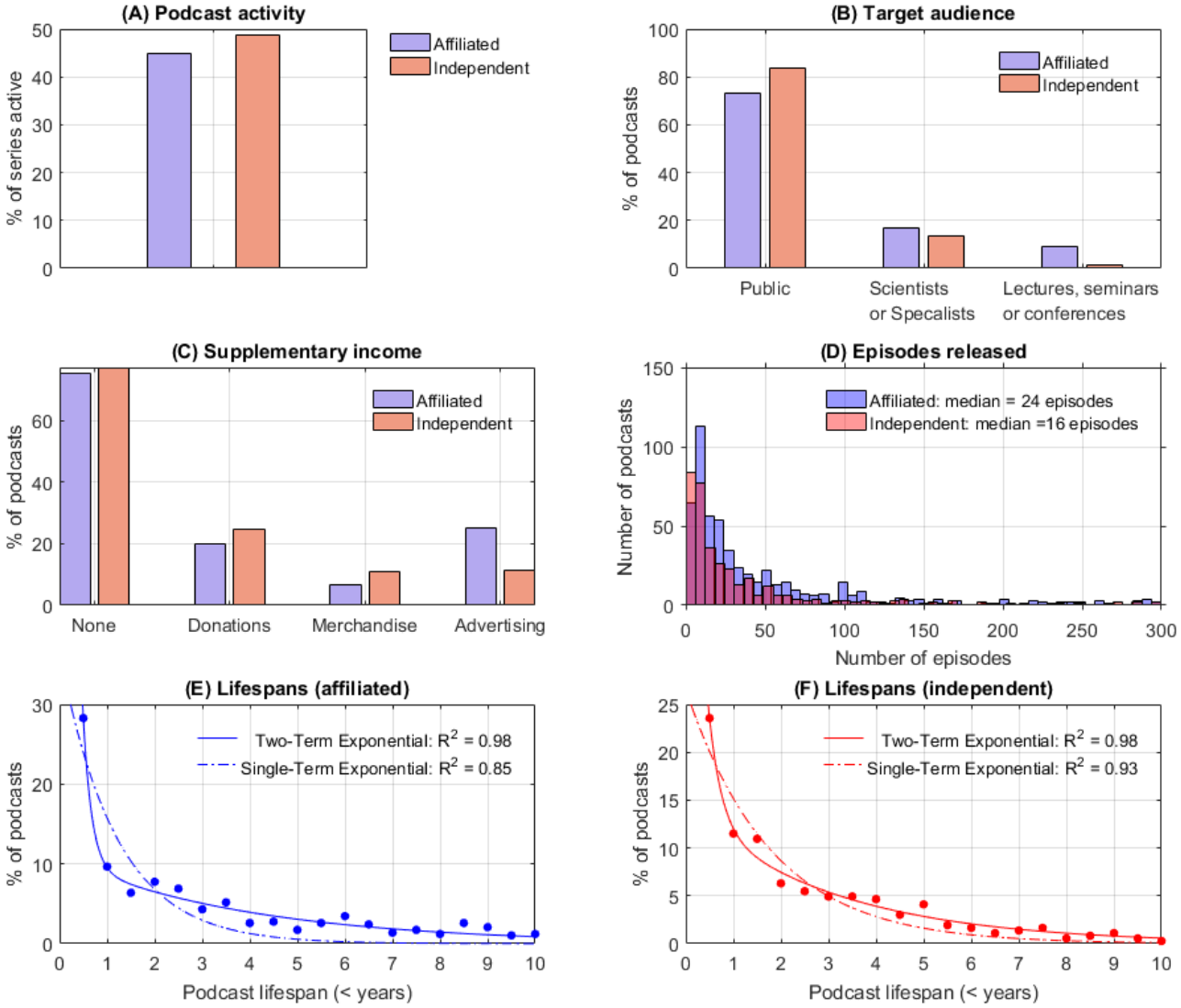

Figure 7: What are the effects of podcast show affiliation on production outcomes? (A) Podcast affiliation vs. podcast activity. **(B)** Podcast affiliation vs. target audience. **(C)** Podcast affiliation vs. supplementary income mechanisms. **(D)** Podcast affiliation vs. total number of podcast episodes produced for each show. N.b. x-axis limited to 300 episodes for clarity due to extreme outliers. **(E,F)** Podcast affiliation vs. overall lifespan for each population. The improved fits for two-term exponential functions over single-term exponential functions is indicative of a bimodal distribution of podcast lifespans, i.e. a “short-lived” and a “long-lived” podcast populations.

5

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2. Estimated mean lifespan for 'Independent' and 'Affiliated' podcasts. Values quoted for upper and lower bounds represent 95% confidence intervals.

	Mean lifespan (τ)					
	Short lifespan (months)			Long lifespan (years & months)		
	Lower bound	Best fit estimate	Upper bound	Lower bound	Best fit estimate	Upper bound
'Affiliated' Podcasts	2 months	3 months	6 months	4 years, 9 months	5 years, 6 months	6 years, 7 months
'Independent 'Podcasts	3 months	4 1/2 months	9 months	3 years, 9 months	4 years, 4 months	5 years, 2 months

3. Methods

3.1. Information Sources

All information used in this study was sourced from public websites that were dedicated to the promotion of podcasts. Information was gleaned exclusively from visual and textual "metadata" relating to each podcast, including the description of each podcast on '*iTunes*'; the websites of podcasts; and the social media content associated with podcasts on '*Twitter*',[9] '*Facebook'*,[10] and *'Patreon'*.[11] The audio and video content of podcasts themselves was not utilized due to the impracticalities associated with listening and transcribing the tens of thousands of hours of audio content that science podcasts provide.[12] Producers and other individuals associated with podcasts were not contacted for information relating to this study. In all cases, information was accessed between the 5th of January and 5th of February 2018. The associated supplementary database contains all the specific dates of when each website URL was accessed.

3.2. Identification of Podcasts for Study

Initially, 165 science podcasts were identified from a blog post previously compiled by the author.[13] However, this resource does not represent a comprehensive or systematic search of science podcasts. Further, due to the decentralised nature of the podcast medium, there is not a single podcast database or website that lists all podcasts. However, the closest thing to a de-facto centralised podcast database is the '*iTunes*' podcast directory, which as of 2015, was estimated to list over 200,000 podcasts.[14] The popularity of '*iTunes*' as the default directory for podcasts can be accredited to the fact that many podcast apps run on the '*iTunes*' directory engine. Therefore if a podcast is not listed on the '*iTunes*' podcast directory, then it is considerably less likely to be found by listeners.[15] A such, the '*iTunes*' podcast represents the best source to use as a basis for this study of science podcasts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A systematic review of the '*iTunes*' podcasts[c] '*Natural Sciences'* directory was conducted to identify potential podcasts for inclusion in this study.[2] All podcasts in the '*Natural Sciences*' section were examined in reverse alphabetical order, between the 5th of January 2018 and the 5th of February 2018. However, podcast listings within the '*iTunes*' podcast directory are based entirely on self-reporting by the uploaders of individual podcasts, so there are many non-scientific podcasts spuriously listed in the '*Natural Sciences*' '*iTunes*' category.[2] Therefore, to ensure all include in this study were valid podcasts covering scientific topics, a stringent set of inclusion criteria were developed and applied (see Section 3.3). The inclusion criteria were applied upon analysis of textual and visual "meta-data" associated with each podcast, as defined in Section 3.1.

In addition, during the study, some podcasts were found that were not listed on the '*iTunes*' podcast directory. These podcasts were found via superfluous search results for other podcasts, or within the websites and descriptions of other podcasts. Rather than immediately excluding these podcasts from this study, they were also considered for inclusion. Of these 'non-*iTunes*' listed podcasts, 18 met the inclusion criteria. These podcasts represent ~2% of the total podcasts reported in this study.

3.3. Inclusion Criteria

To ensure that only valid podcasts were included in this study of science podcasts, the following set of inclusion/exclusion criteria were developed and applied by the author:

- Podcasts must be English language podcasts.
- If a podcast was available in both English and another language as separate podcast feeds, only the English language podcast feed was included for analysis.
- If a podcast was available as both an audio-only feed and a video-feed, then only the video-feed was included for analysis.
- To be included in this study, the primary focus of a podcast must be on covering scientific topics, for example, scientific research, science news, scientific careers, scientific seminars, lectures, or similar.
- Podcasts focusing on non-science topics were excluded. This included (but is not limited to) podcasts covering technology; business; gardening, bird-watching (not including ornithology); food/cooking; religion; life-coaching; weather; sustainability; environmental activism; pseudo-science; occult and paranormal; nerd culture; and hobby podcasts focused on consumer products (e.g. tropical fish keeping or hobby telescopes).

[c] '*iTunes*' may also be referred to elsewhere as 'Apple Podcasts'.[47]

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

- If the scientific nature of a podcast was unclear, then that podcast was excluded.
- Podcasts with no episodes available to stream or download were excluded.
- To be included for analysis, episodes of a podcast had to be freely available for listeners to stream or download from a source at the time of sampling. For example, if a podcast had 100 episodes available on '*iTunes*', yet had 250 episodes available to stream on their own website, then the 250 episodes were included for analysis.
- If podcasts started as an internet or broadcast radio show, before subsequently becoming available as a podcast, then the original air date of the first episode (now freely available for streaming or download) was used as the date of the first episode.

3.4. Categorical Definitions

Podcasts, their production methods, and their production outputs were manually classified by the author in accordance with the definitions in Table 3.

Table 3. Categorical definitions used for classifying podcasts.

Category	Definition
<u>Podcast Activity</u> (see Figure 1)	
Episode	A single instalment of a podcast, which may be downloaded or streamed.
Podcast	A collection of podcast episodes released under the same podcast name/podcast feed.
Active podcast	A podcast that has released at least one episode within the three months immediately prior to the sampling date.
Inactive podcast (< 1 year)	A podcast that has released at least one episode in the period between twelve and three months immediately prior to sampling date.
Inactive podcast (> 1 year)	A podcast that has not released an episode in the twelve months immediately prior to the sampling date.
Podcast lifespan	The time elapsed between the release dates of the first and last episode of a podcast.
Number of episodes	The total number of episodes available to the public to freely download or stream, either via '*iTunes*' or another website.
<u>Audiences</u> (see Figure 2B)	
Public	The primary audience of this podcast are the general public, who are not assumed to have extensive scientific expertise or to be familiar with the topics covered. *Examples include 'BBC Inside Science',*[16] *'Science Vs.',*[17] *'Science Brunch',*[18]*and 'The Naked Scientists'.*[19]
Scientists or specialists	The primary audience of this podcast are scientists or specialists in fields related to science, who are assumed to have relevant specialist knowledge and specialist interests. *Examples include 'This Week in Virology',*[5] *'ExoCast',*[20] *and 'The Black Goat'.*[21]

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Lectures, seminars, or conferences.	This podcast is intended to deliver the contents of a scientific lecture, seminar, or conference presentation; i.e. it is intended to an audience listening to it for educational or professional learning purposes.
Children	The primary audience of this podcast is intended to be children. N.b. Age of children is not strictly defined in this study. *Examples include 'Brains On',*[22] *'Wow in the World',*[23] *'Tumble',*[24] *and 'The Show About Science'.*[25]

<u>Hosts</u> (see Figure 23B)

Scientific Researchers/Educators	Podcast hosts whose occupation is/was primarily based on science research, science education, or science communication.
Media/Journalism Professionals	Podcast hosts whose occupation is/was primarily focused on producing conventional media, such as radio shows or newspaper articles.
Other Professionals	Podcast hosts that have an acknowledged professional capacity that is not media production or scientific education/research. For example, comedians and musicians.
Amateurs	Podcast hosts that are hosting in an amateur capacity, for example as part of local astronomy or "sceptics" groups.
Unclear	Host category could not be identified with available information.

<u>Podcast Affiliations</u> (see Figure 3 and Figure 7)

Independent	A podcast with no explicit or direct affiliation to any organisation. N.b. this does not include paid advertisements or sponsorships.
Affiliated	A podcast which explicitly acknowledges a direct affiliation to an organisation, as per one of the categories below.
University	A university which is directly involved in education and research. *Examples: 'The University of California TV',*[26] *and 'The University of Wisconsin Sea Grant Institute'.*[27] N.b. For simplicity, secondary education organisations (i.e. high schools) are included within this category because they are not numerous enough to warrant separate categorisation.
Other Research Body	A non-university organisation which conducts scientific research. *For example: 'NASA',*[28] *and the 'Centres for Disease Control and Prevention'.*[29]
Professional Organisation	A professional organisation or body that does not directly conduct scientific research. *For example: 'The American Chemical Society',*[30] *'The American Society for Microbiology',*[31] *and 'The Institute of Physics'.*[32]
Scientific Journal	An organisation that mainly produces peer-reviewed scientific journals. *For example: 'Nature',*[33] *'PLOS',*[34] *and 'SAGE'.*[35]
Conventional Media Body	An organisation which primarily disseminates conventional media, such as TV/radio broadcasts, or print media. *For example: 'BBC Radio 4',*[36] *'ABC Radio National,*[37] *'Scientific American',*[38] *and 'NPR'.*[39]
Podcast Network	An internet-only media organisation solely dedicated to releasing podcasts. *For example, 'The Naked Scientists',*[19] *'Relay FM',*[40] *and the 'StarTalk Radio'.*[41]
Amateur Organisation	Any amateur organisation. For example, local astronomy groups and "sceptics" societies.

<u>Non-audio media</u> (see Figure 4)

Audio podcast	A podcast that directly incorporates only audio information (not including media in show notes).
Video podcast	A podcast that directly incorporates visual and audio information (not including media in show notes).
Show notes	Media or information, hosted on podcast websites, which is supplementary to a podcast episode. 'Show notes' may include information such as: images, videos, hyperlinks, scientific references, and show transcripts.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

	However, simple descriptions of a podcast episode are not classified as 'show notes'.

<u>Countries</u> (see Figure 5)

Country of podcast production	The country primarily associated with a podcast and it's hosts. N.b. If a podcast is clearly associated with two or more countries, then that podcast is classified as "multinational".

<u>Supplementary Income</u> (see Figure 6)

Donations	Requests for voluntary donations from listeners.
Merchandise	Goods or services associated with the podcast.
Advertising/Sponsorship	Explicitly acknowledged sponsorship or advertisement from an organisation other than the organisation the podcast is affiliated with. *N.b. Where podcasts are directly affiliated to advertiser-supported commercial radio, TV, or podcast networks, then advertising is assumed as default.*

<u>Podcast Lifespans</u> (see Figure 1, Figure 7, and Table 2)

Mean lifespan (τ)	The timespan in which 50% of a given population of podcasts will be become 'inactive'. The mean lifespan is estimated by fitting an exponential decay to the lifespan data of a population of podcasts, and it is analogous to the concept of 'mean lifetime' in the context of radioactive decay.
Short lifespan podcasts	The population of podcasts with a 'mean lifespan' of less than one year.
Long lifespan podcasts	The population of podcasts with a 'mean lifespan' of more than one year.

3.5. Data analysis and plotting

All relevant information and resultant categorical analysis was recorded within a spreadsheet database (Microsoft Excel 2016, .xlsx format), which is available as a supplementary dataset to this manuscript. Basic categorical analysis was undertaken with Microsoft Excel, however, advanced categorical and data analysis (such as analysis of podcast lifespan) was carried out using custom-written MATLAB scripts (MATLAB 2017b, Mathworks).

To estimate mean lifespan of podcasts, single-term and two-term exponential decays were fitted to podcast lifespan data by least-squares regression. The equations describing these fits are respectively:

$$y = ae^{bx}, \qquad \text{Equation 1}$$

$$y = ae^{bx} + ce^{dx}. \qquad \text{Equation 2}$$

Where *a, b, c, and d,* are recovered best-fit parameters. The mean lifespan (τ) was then calculated by:

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$$\mathrm{T} = -ln(2)/b. \qquad \text{Equation 3}$$

Where $\ln(2)$ is the natural logarithm of *2* (approximately 0.693). For estimation of long and short lifespans from two-term exponential decay fits, d can be substituted for b in Equation 3. Upper and lower bounds for T are estimated from upper and lower 95% confidence from fit data. Figures were created from data by plotting in MATLAB with some minor annotations added in PowerPoint (Microsoft PowerPoint 2016).

4. Discussion

4.1 Methodology and associated limitations

This is the first study to analyse the global production and outputs of a large group of science podcasts. As such, the findings here provide fundamental and novel insight into who is producing science podcasts and to what target audience. However, before detailed discussion of results, it is important to acknowledge the limitations of the methodology employed in this study.

Firstly, only English language science podcasts were surveyed and analysed. It is possible that non-English language science podcasts may have intrinsically different demographics.

Secondly, science podcasts were identified by exhaustive survey of the '*Natural Sciences*' category of '*iTunes*'.[2] Although listing on '*iTunes*' is highly beneficial for the visibility of podcast to listeners, listing on '*iTunes'* is self-reported by podcast producers and is not mandatory. 'Therefore, this study will not have analysed all science podcasts available in the sample period. Indeed, an unknown number of science podcasts are listed in the '*Science & Medicine'* '*iTunes*' umbrella-category but not be listed in the '*Natural Sciences*' sub-category.[42] However, exhaustive review of this category was not possible due to practical constraints. Moreover, some science podcasts are not listed on '*iTunes*' at all. An example of this is the '*BioLogic Podcast'*, which is hosted on the video sharing website '*YouTube*'.[43] Additionally, it should be noted that some podcasts may voluntarily restrict the number of podcast episodes that are freely available to the public via '*iTunes*' or other websites; however only freely-available episodes were included for analysis within this study. Despite these limitations, is expected that the sample size of 952 science podcasts is sufficient to provide a representative sample of science podcasts.

Thirdly, this study exclusively examined the visual and textual online presence of podcasts. For practicality, it was not possible to examine the extensive audio data associated with science podcasts. Thus, it is possible that various aspects of podcast production were not fully categorised; this could affect all studied podcast categories, but most likely affects the capture of any audio-only advertisements or sponsorships that were not acknowledged

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in textual or visual web content of science podcasts. Therefore, it is possible that a greater proportion of science podcasts contain advertisements or sponsorships than is explicitly reported by this study. In future, it may be possible to utilise machine learning to automatically transcribe, analyse, and categorize audio content without human supervision. Such techniques could enable data-mining of the extremely large amount of podcast audio content available.[44,45]

Fourthly, podcast episode length data and podcast download statistics were not available for analysis via the '*iTunes*' podcast listings. This data would be desirable for a complete analysis of science podcast demographics, but is not strictly necessary. In future, it may be possible to use automated computational analysis to ascertain the lengths of science podcast episodes if so desired.

4.2 Science podcasts demographics versus wider podcast demographics

The literature on podcasts is somewhat sparse: to-date, all peer-reviewed literature has focused on the use of podcasts for educational or promotional purposes. For studies of podcast production and outcomes, it is necessary to look beyond peer-reviewed literature. In 2015, Josh Morgan published a semi-formal study of podcast demographics as a blog post on '*medium.com'*.[14] Whilst not published in a peer-reviewed scientific journal, all data associated with Morgan's study is publicly available for scrutiny. Morgan's study sampled a subset of podcasts available on the '*iTunes*' podcast listings in June 2015. Morgan estimated that there were 206,000 unique podcasts available of '*iTunes'* at that time. Morgan than selected a random subset of podcasts for further analysis. This subset consisted of a total of 2500 podcasts, with 100 random podcasts drawn from the 25 "most popular" '*iTunes*' categories (N.b. this did not include any category dedicated to or focusing on science). Morgan's sampling and analysis was fully-automated, so manual categorisation of podcast production outputs was not conducted.[14]

Morgan found that the number of podcasts available on '*iTunes*' had grown from ~10,000 in 2007 to ~206,000 in 2015. When graphed, the trends in growth of total number of podcasts calculated by Morgan (not shown in this study) appear broadly like the trends shown in Figure 1A, i.e. displaying distinct linear growth from till 2010, and distinct exponential growth from 2010 onwards. This indicates that trends in the growth of science podcasts reflects the overall growth of the podcast medium. Additionally, Morgan found that roughly 40% of podcasts were 'active'.[14] This is lower than the comparable population of 'active' science podcasts (46%) (see Figure 1B). Furthermore, this difference in 'activity' is greater than the numbers would suggest, because the present study defined 'active' podcasts as podcast that released an

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

episode in the three-months prior to sampling, whereas Morgan's study deemed podcasts 'active' in the six months prior to sampling. This comparison suggests that science podcasts are significantly more inclined to continue to release episodes compared to the wider population of podcasts. However, this may also be due to the fact that Morgan did not exclude podcasts that had not released a single episode. Therefore, the comparison is not necessarily valid. Further, Morgan found that the average lifespan of podcasts was around 6 months, and that podcasts, on average, released 12 episodes, at a rate of 2 episodes per month. In addition, Morgan estimated that around 20% of podcasts listed on '*iTunes*' were not English-language podcasts.

4.3. Insights into the production of science podcasts

The predominance of scientists as hosts for science podcasts (see Figure 3A) combined with fact that most science podcasts (57%) are released on an irregular schedule (see Figure 3C) may indicate that a significant majority of science podcasts are being produced by scientists as an extra commitment beyond their regular duties as a scientific researcher, educator, or communicator. Further, most science podcasts do not have any overt supplementary income mechanisms. Considering that there can be non-trivial costs associated with hosting a podcast (audio equipment, editing software, promotional website hosting, and podcast hosting), this fact suggests that independent science podcast hosts are paying these costs "out of their own pocket". These results combine to give a broad impression that many science podcasts are being produced by scientists with no financial recompense. The obvious exception being the science podcasts 'Affiliated' to organisations that can provide undisclosed financial support. However, the fundamental validity of this interpretation requires further research and study before firm conclusions can be made.

Although a wide variety of topics are covered by science podcasts, Chemistry – accounting for only 3% of science podcasts - appears to be somewhat under-represented in comparison to the other "traditional" science subjects of 'Physics and Astronomy' (18% of total science podcasts) and 'Biology' (13% of total science podcasts) (see Figure 2A). It is currently unclear as to why this may be.

There are significant and notable differences between podcast production outputs for 'Independent' and 'Affiliated' podcasts, most notably in the lifespan of podcasts. From Table 2, it can be seen that for the population of short lifespan podcasts, podcasts with 'Independent' status have a greater mean lifespan when compared to 'Affiliated' status podcasts. This could possibly be explained by a hypothesis that many 'Affiliated' podcasts may have a fixed short-term brief or mission. In contrast, for the population of 'long lifespan' podcasts, 'Independent' status results in a shorter mean lifespan than 'Affiliated' status. This could be explained by a hypothesis that independent podcasts are produced by a smaller

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

group of peoples, with limited resources, whereas 'Affiliated' podcasts are produced by organisations with dedicated staff with defined duties. Such dedicated staff could take-over podcasting duties, thus extending the overall lifespan of an 'Affiliated' podcast in comparison to an 'Independent' podcast. However, it should be noted that there are exceptionally long-running 'Independent' and 'Affiliated' podcasts; therefore, further study is required on the matter of what influences podcast lifetime.

The ratio of podcasts produced by the USA and UK is 3.32 (57% USA / 17% UK). This is broadly similar to the 3.44 ratio of high-quality scientific papers produced by the USA & UK, as measured by the Nature Index (in 2011).[46] This similarity does not hold for other countries in relation to the USA. However, due to the low number of science podcasts from other countries, it is not possible to form any robust insights based upon the ratio of podcasts produced by countries.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

5. Conclusions

This is the first study to analyse the decentralised and growing production of science podcasts. In the period between 2004 and 2010, the total number of English language science podcasts grew linearly, and between 2010 and 2018, the total number of science podcasts grew exponentially. The growth in science podcasts likely reflects larger trends in the growth of the podcast medium, but peer-reviewed data is lacking. A total of 952 science podcasts were available in the sampling period between the January 5th and February 5th, 2018.

Although science podcasts have been created by a diverse range of independent producers and organisations, it was found that science podcasts are mainly produced and hosted by scientists. Further, the majority of science podcasts receive no overt financial support. This highlights that science podcasts are being utilised as a decentralised and independent avenue of science communication to the public.

Most science podcasts (77%) are targeted at public audiences. The greatest single portion of science podcasts cover 'General Science' topics (34%). Many specific topics, such as: 'Physics and Astronomy', 'Biology', 'Ecology/Zoology/Conservation', 'Oceanography/Marine-Biology', and 'Psychology/Neuroscience' are all well-represented by science podcasts. However, it is arguable that Chemistry (3%) is under-represented in science podcasts when compared to the other science subjects of Physics and Astronomy (18%) or Biology (13%).

Although science podcast production is global, 57 of science podcasts have been produced in the USA and 17% have been produced in the UK. The reasons for this predominance of podcast production are currently unclear.

The population of science podcasts was found to contain 'short lifespan' and 'long lifespan' podcasts, corresponding to podcasts with mean lifespans of less than one year, and greater than one year respectively. Independently produced podcasts had a greater short-duration mean lifespan than podcasts produced in affiliation with an organisation. However, independently produced podcasts had a reduced long-duration mean lifespan in comparison to affiliated podcasts.

It was found that 87% of science podcasts are audio-only, but only 51% feature supplementary information in the form of 'show notes'. Therefore, there is considerable potential for science podcasts to provide more information to their audiences via supplementary show notes, e.g. by providing images, hyperlinks, transcripts, and scientific references.

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

These insights combine to provide the first ever 'snapshot' of how science podcasts are being produced and disseminated as a new, global, and decentralised medium for direct public science communication.

References

1. Condayan, C. Culture Media. *Nat. Rev. Microbiol.* **6,** 646 (2008).
2. iTunes 'Natural Sciences'. (2018). Available at: https://itunes.apple.com/us/genre/podcasts-science-medicine-natural-sciences/id1477?mt=2. (Accessed: 3rd March 2018)
3. Skiba, D. J. The 2005 Word of the Year: Podcast. *Nurs. Educ. Perspect.* **27,** 54–55 (2006).
4. Dunn, M. J. Launch of the Proteomics Podcast. *Proteomics* **6,** 5983 (2006).
5. This Week In Virology. *Podcast* (2018). Available at: http://www.microbe.tv/twiv/. (Accessed: 3rd March 2018)
6. Johnson, J., Ross, L., Iwanenko, W., Schiffert, J. & Sen, A. Are Podcasts Effective at Educating African American Men About Diabetes? *Am. J. Mens. Health* **6,** 365–367 (2012).
7. Thoma, B. *et al.* The impact of social media promotion with infographics and podcasts on research dissemination and readership. *Can. J. Emerg. Med.* 1–7 (2017). doi:10.1017/cem.2017.394
8. Prakash, S. S., Muthuraman, N. & Anand, R. Short-duration podcasts as a supplementary learning tool: Perceptions of medical students and impact on assessment performance. *BMC Med. Educ.* **17,** 1–14 (2017).
9. Twitter. *Social Media Platform* (2018). Available at: https://twitter.com/. (Accessed: 3rd March 2018)
10. Social Media Platform. Facebook. (2018). Available at: https://www.facebook.com/. (Accessed: 3rd March 2018)
11. Patreon. *Social Media Platform,* (2018). Available at: https://www.patreon.com/. (Accessed: 3rd March 2018)
12. Bailey, J. First steps in qualitative data analysis: Transcribing. *Fam. Pract.* **25,** 127–131 (2008).
13. MacKenzie, L. E. A big list of science podcasts. *Blog* (2018). Available at: http://www.lewismackenzie.science/blog-1/2017/2/1/a-big-list-of-science-podcasts. (Accessed: 5th January 2018)
14. Morgan, J. How Podcasts Have Changed in Ten Years: By the Numbers. *Medium.com* (2015). Available at: https://medium.com/@monarchjogs/how-podcasts-have-changed-in-ten-years-by-the-numbers-720a6e984e4e. (Accessed: 3rd March 2018)
15. Squarespace. Setting up a blog to host your podcast. (2018). Available at: https://support.squarespace.com/hc/en-us/articles/115001990668-Setting-up-

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

a-blog-to-host-your-podcast. (Accessed: 3rd March 2018)

16. BBC Radio 4. BBC Inside Science. *Podcast/Radio Show* (2018). Available at: https://www.bbc.co.uk/programmes/b036f7w2/episodes/downloads. (Accessed: 3rd March 2018)

17. Gimlet Media. Science Vs. *Podcast* (2018). Available at: https://www.gimletmedia.com/science-vs. (Accessed: 3rd March 2018)

18. Science Brunch. *Podcast* (2018). Available at: http://www.sciencebrunch.org/. (Accessed: 3rd March 2018)

19. The Naked Scientists. *Podcast* (2018). Available at: https://www.thenakedscientists.com/. (Accessed: 3rd March 2018)

20. Exocast: The Exoplanet Podcast. *Podcast* (2018). Available at: http://www.exocast.org/. (Accessed: 3rd March 2018)

21. The Black Goat: A Podcast About Doing Science. *Podcast* (2018). Available at: http://www.theblackgoatpodcast.com/.

22. News, M. & Southern California Public Radio. Brains On: A Podcast For Kids and Curious Adults. *Podcast* (2018). Available at: https://www.brainson.org/. (Accessed: 3rd March 2018)

23. NPR. Wow In The World. *Podcast* (2018). Available at: https://www.npr.org/podcasts/510321/wow-in-the-world. (Accessed: 3rd March 2018)

24. Tumble: A Science Podcast For Kids. *Podcast* (2018). Available at: http://www.sciencepodcastforkids.com/. (Accessed: 3rd March 2018)

25. The Show About Science. *Podcast* (2018). Available at: http://theshowaboutscience.com/. (Accessed: 3rd March 2018)

26. University of Calafornia. University of Calafornia TV. *Podcast* (2018). Available at: https://www.uctv.tv/ondemand/. (Accessed: 3rd March 2018)

27. University of Wisconsin Sea Grant Institute: Podcasts. (2018). Available at: http://seagrant.wisc.edu/home/default.aspx?tabid=600. (Accessed: 3rd March 2018)

28. National Aeronautics and Space Administration (NASA). NASA Podcasts. *Podcast* (2018). Available at: https://www.nasa.gov/podcasts. (Accessed: 3rd March 2018)

29. Centers for Disease Control and Prevention. Podcasts at CDC. *Podcast* (2018). Available at: https://www2c.cdc.gov/podcasts/. (Accessed: 3rd March 2018)

30. Science Podcasts: The American Chemical Society. (2018). Available at: https://www.acs.org/content/acs/en/pressroom/podcasts.html. (Accessed: 3rd March 2018)

31. The American Society for Microbiology: Podcasts. (2018). Available at: https://www.asm.org/index.php/podcasts. (Accessed: 3rd March 2018)

32. The Institute of Physics. (2018). Available at: http://www.iop.org/. (Accessed: 3rd March 2018)

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

33. Nature Publishing Group. Nature Podcast. *Podcast* (2018). Available at: https://www.nature.com/nature/podcast/index.html. (Accessed: 3rd March 2018)

34. PLOS. PLOScast. *Podcast* (2018). Available at: http://blogs.plos.org/plospodcasts/. (Accessed: 3rd March 2018)

35. SAGE Publishing. Podcasts and Vodcasts. *Podcast* (2018). Available at: https://uk.sagepub.com/en-gb/afr/podcasts. (Accessed: 3rd March 2018)

36. BBC Radio 4 Podcasts. (2018). Available at: https://www.bbc.co.uk/podcasts/radio4. (Accessed: 3rd March 2018)

37. ABC Radio National. Podcasts and feeds: by program. (2018). Available at: http://www.abc.net.au/radionational/podcasts/program/. (Accessed: 3rd March 2018)

38. Scientific American. Podcasts. (2018). Available at: https://www.scientificamerican.com/podcasts/. (Accessed: 3rd March 2018)

39. National Public Radio (NPR). Podcasts. (2018). Available at: https://www.npr.org/podcasts/. (Accessed: 3rd March 2018)

40. Relay FM Podcast Network. (2018). Available at: https://www.relay.fm/. (Accessed: 3rd March 2018)

41. Star Talk Radio. *Podcast* (2018). Available at: https://www.startalkradio.net/. (Accessed: 3rd March 2018)

42. iTunes 'Science & Medicine'. (2018).

43. BioLogic Podcast. *Podcast* (2018). Available at: https://www.youtube.com/channel/UCO3Mka2Esro-bqVRDQnvmaQ. (Accessed: 3rd March 2018)

44. Brynjolfsson, E. & Mitchell, T. What can machine learning do? Workforce implications. *2Science* **358,** (2017).

45. Amodei, D. *et al.* Deep Speech 2: End-to-End Speech Recognition in English and Mandarin. **48,** (2015).

46. Van Noorden, R. 365 days: 2011 in review. *Nature.com* (2011). Available at: https://www.nature.com/news/365-days-2011-in-review-1.9684. (Accessed: 3rd March 2018)

47. Apple Inc. Apple Podcasts Identity Guidelines. (2017). Available at: https://www.apple.com/itunes/marketing-on-podcasts/identity-guidelines.html. (Accessed: 3rd March 2018)

Supplementary Dataset

All data associated with this study is available as a supplementary Microsoft Excel spreadsheet. This dataset contains all classification information and URLs associated with each podcast.

Acknowledgements

bioRxiv preprint doi: https://doi.org/10.1101/298356; this version posted May 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Thanks to the many science podcasts and podcast creators who inspired this work. Additional thanks to Cecilie A. Osnes for proof-reading. L.E. MacKenzie was supported by an Engineering and Physical Sciences Research Council (EPSRC) grant: EP/P025013/1.

Author Contributions

Dr MacKenzie is solely responsible for all work herein.

Competing Interests

The author declares no competing interests.

Electronic Supplementary Material

Available at: [DOI TBC]