TINNITUS HETEROGENEITY: Can big data reveal new insights?

Dr Theodore Kypraios & Eleni Genitsaridi
Copyright © British Tinnitus Association 2018.

Editors: Deborah Hall & Nic Wray

The editors would like to thank all the contributors for their support and enthusiasm. We would also like to thank the copyright owners of the images and figures reproduced in this publication for their permission to reproduce them either directly or via licence. We believe that all such work has been acknowledged, but please contact Nic Wray on nic@tinnitus.org.uk if any acknowledgements have been inadvertently omitted.

British Tinnitus Association
Unit 5 Acorn Business Park
Woodseats Close
Sheffield
S8 0TB

www.tinnitus.org.uk
info@tinnitus.org.uk
0114 250 9933

Registered charity no. 1011145. Company limited by guarantee no. 2709302, registered in England.
TINNITUS HETEROGENEITY: CAN BIG DATA REVEAL NEW INSIGHTS?

Dr Theodore Kypraios & Eleni Genitsaridi

Introduction

Characterising tinnitus heterogeneity remains one of the major challenges facing clinical practice and tinnitus research. Tinnitus heterogeneity refers to the variability in any feature of tinnitus and reactions to it that can be observed between different individuals (inter-individual variability) or at different times in the same individual (intra-individual variability). These various aspects of tinnitus can be measured and so from a statistical perspective they are therefore all potential ‘variables’. Developments in the fields of mathematics, statistics and computer science have provided the scientific community with many powerful tools to analyse multi-dimensional data in various ways. This short article considers how statistical analysis of large datasets can reveal insights into tinnitus heterogeneity.

The problem of heterogeneity

If we randomly select two people from the tinnitus population we will most likely find that they differ in the cause of tinnitus (e.g. ear pathologies), characteristics of tinnitus perception (e.g. pitch, loudness, temporal characteristics, localisation), reactions to tinnitus (e.g. annoyance, distress), co-occurring conditions (e.g. generalised anxiety), and other general characteristics such as age, hearing status and physical health. Furthermore, the same person can experience tinnitus differently at different times depending on internal factors such as being stressed or tired or external factors such as the level of environmental noise or the quality of social support.
Heterogeneity poses significant challenges for clinical practice. There is no agreement about what are the most relevant variables for assessing a tinnitus patient. Moreover, the various causal factors, characteristics, reactions, co-occurring conditions and general characteristics all mean that a treatment might only be optimally effective for a specific subgroup of patients.

Heterogeneity also poses significant challenges for research. Without agreement about assessment variables, different teams of researchers tend to focus on different characteristics, meaning that patient groups cannot be properly compared across studies. Moreover, interactions between different variables are not clearly defined.

Collecting data from people with tinnitus and analysing them by using state-of-the-art statistical methods could potentially provide essential insight into tinnitus heterogeneity and in particular to help identify an appropriate set of variables for the evaluation and classification of tinnitus patients.

**Figure 1:**
Overview of variables relevant to people with tinnitus. Aetiological mechanisms can be related to many of these variables including medical history (e.g. ear diseases, hearing loss, ototoxic medication), environmental factors (e.g. noise exposure), and associations between variables (e.g. onset-related events and modulating factors). Reactions to tinnitus (e.g. degree of annoyance) and the impact of tinnitus on mental or physical health (e.g. cognition, concentration or sleep) are clinically significant variables and can be the targets of clinical interventions.
‘Big data’: many variables from large populations

Researchers need to collect and analyse large amounts of data in order to produce useful scientific insights which capture all aspects of heterogeneity and the complex interactions between variables. All those variables that might be relevant for such analyses are shown in Figure 1. These variables can be separated into general characteristics that are relevant to anybody with or without tinnitus, and tinnitus-related characteristics that are relevant only to people with tinnitus. Most of these are time-dependent, meaning that they can vary over time contributing to intra-individual variability.

Evidence on associations between variables relevant to people with tinnitus has been accumulating for many decades. In 2017, researchers reported, among others, associations between tinnitus and neck complaints [1], headache [2], dizziness [3], hyperacusis [4], vestibular schwannoma [5], Ménière’s disease [6], cochlear implantation [7] [8], mental health at tinnitus onset [9], cardiometabolic risk factors [10], brain anatomy and activity [11][12][13] and genetic characteristics [14]. Some of these studies examine differences between people with or without tinnitus, while others delve into tinnitus heterogeneity and examine differences within the tinnitus population. The following sections describe the different statistical approaches that have been used to investigate these questions.

**Multivariate data analysis**

Multivariate techniques enable simultaneous analysis of multiple variables on different individuals. They can be used to examine interactions between those variables and to identify the underlying structure of the data. However, there is a wide variety of available techniques and it is often challenging for researchers to choose the most appropriate method for a given research question.

Multivariate data analysis techniques can be broadly classified into dependence and interdependence techniques [15]. In the first case, researchers must choose one or more variable as the ‘dependent variable(s)’ that they will try to predict or explain using the rest of the variables called ‘independent variables’. Such techniques can be used when trying to predict the severity of a condition or the treatment outcome [dependent variable] from other available characteristics such as gender, age or co-existing conditions [independent variables]. Examples of dependence multivariate techniques include multiple linear regression, multiple logistic regression, multivariate analysis of variance and structural equation modelling.

Interdependence techniques treat all variables the same way and analyse them simultaneously. They can be used to explore the underlying structure among all variables. Examples of such techniques include cluster analysis, latent class analysis and principal component analysis.

After deciding on an appropriate technique, more decisions need to be made while implementing it. For example, the choice of the set of variables to be included in an analysis is made by the researcher either based on their experience and/or previous findings. This choice of variables can be crucial for the final results and therefore, caution is needed while interpreting the results. For valid conclusions, the limitations of the methods and previous knowledge must be taken into consideration.

**Applications of multivariate techniques to unravel tinnitus heterogeneity**

In 2017, a number of publications addressed the issue of tinnitus heterogeneity using a range of multivariate techniques, as follows:

**Cluster analysis to reveal subgroups**

Cluster analysis is used to define groups where members within a group are more similar to each other, compared to members between different groups [15]. Researchers must decide on the set of variables that will be included in the analysis to compare members and create clusters. This technique has been used previously to define...
subgroups based on various characteristics [16] [17][18][19][20].

Van den Berge et al conducted a cluster analysis to identify subgroups in a dataset from 1,783 tinnitus patients [21]. The authors followed two approaches to decide which variables would be included in the analysis. In the first case, the choice was made by clinical experts based on their experience. In the second, it was based on the results of principal component analysis, which is a dimension-reduction statistical technique that is used to reduce a large set of variables to a smaller set that still contains most of the information in the large set. In both approaches, the final set of variables included tinnitus, audiological and other individual characteristics. Although the main finding of this study was that no clear subgroups could be formed, the authors discuss that ‘any cluster analysis outcome highly depends on the variables that we entered into the clustering algorithm’.

Latent class analysis to reveal patterns of hearing loss
Latent class analysis is another method to cluster categorical observations of high dimension that uses a probabilistic model to define classes (i.e. groups) based on an underlying latent variable [22]. The set of the included variables is again predetermined by the researcher.

Langguth et al [23] used this technique to classify patients based on their hearing profile. In other studies, hearing status has often been described with a single variable (e.g. severity of hearing loss or mean hearing threshold), but this fails to capture potentially important information on differences between the two ears or across frequencies. In this recent study, the authors described hearing status using a vector of 14 variables representing audiometric thresholds at seven frequencies (0.125, 0.25, 0.5, 1, 2, 4, and 8 kHz), separately for each ear. The variables included in the analysis were categorical and could take one of four discrete values (i.e. normal hearing, mild or moderate hearing loss, severe or profound hearing loss and no audiometric data available). Using latent class analysis, patients were grouped into eight discrete classes. Subsequent analysis showed that these eight classes also differed in several other clinical characteristics such as depressive state. Even though the hearing profiling was much more detailed than in many similar studies, the authors discuss the importance of including even more variables such as hearing function in frequencies above 8 kHz.

Regression analysis: a very popular method
Regression techniques are commonly used to predict the outcome of a predefined variable (dependent) from a set of other variables (independent). Different types of regression analysis are used depending on the type of the dependent variable, i.e. linear regression for a continuous variable and logistic regression for a binary one [22]. Many researchers have used these techniques to shed light on tinnitus heterogeneity. For example, Michiels et al used multiple linear regression to identify predictors for improvement of patients with cervicogenic somatic tinnitus after cervical physical therapy [1]. The same technique was used by Wielopolski et al to explore if tinnitus severity could be predicted by specific personality and mental health traits [24]. Wallhäusser-Franke et al applied multiple linear regression to longitudinal data to identify predictors of developing disabling chronic tinnitus [9].

Logistic regression was used by Ralli et al to predict the presence of hyperacusis in patients with somatic tinnitus from questionnaire scores and tinnitus and other characteristics [25]. House et al used the same method to predict high or low tinnitus severity from cardiometabolic variables or presence of depression [10].

Multilevel modelling to explore intra-individual variability
The wide use of smartphones and applications comes with interesting potentials for research and data gathering. Probst et al used data from the ‘TrackYourTinnitus’ mobile application...
to explore variations in tinnitus loudness and distress [26]. Using Multilevel Modelling (MLM), a technique that enables analysis of data in a hierarchical structure [22], they examined how tinnitus loudness, tinnitus distress and stress levels vary within a day, between different days or between different individuals.

The results indicated that both loudness and distress were more severe during the night and the early morning compared to the rest of the day. A limitation of this study, reported by the authors, is that some important relevant variables such as environmental noise and sleep disturbances were not included in the analysis.

Conclusions

Tinnitus is a heterogeneous condition. In order to characterise a person with tinnitus a large set of variables need to be assessed. Advances in data analysis techniques can, in principle, help us to identify the latent structure of the associations between all these variables and to find ways to subtype the tinnitus population. Selecting what variables should be collected and analysed is of crucial importance in determining the final results.

Of equal importance is the selection and implementation of the proper technique to analyse this information. Recent advances in statistical and machine learning provide powerful tools that can help scientists reveal the mysteries of tinnitus heterogeneity. However, experts in the fields of both data analysis and tinnitus need to collaborate in order to efficiently apply these methods and answer long-standing questions regarding the heterogeneous tinnitus population. Moreover, since many factors can influence the results of data analysis, caution is necessary when interpreting current findings.

References


Tinnitus heterogeneity: can big data reveal new insights?


Annual Tinnitus Research Review 2018