ELSEVIER

Journal of Neuroimmunology 170 (2005) 71 – 84

Journal of Neuroimmunology

www.elsevier.com/locate/jneuroim

Statistical analysis of data from studies on experimental autoimmune encephalomyelitis

Kandace K. Fleming [a], James A. Bovaird [a,1], Michael C. Mosier [b,2], Mitchell R. Emerson [b,3], Steven M. LeVine [b], Janet G. Marquis [a,*]

[a] *Research Design and Analysis Unit, Life Span Institute, 1052 Dole Building, University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045, USA*
[b] *Department of Molecular and Integrative Physiology, 3901 Rainbow Blvd., University of Kansas Medical Center, Kansas City, KS 66160, USA*

Received 3 June 2005; accepted 26 August 2005

Abstract

Research in multiple sclerosis often employs animal models of the disease, especially experimental autoimmune encephalomyelitis (EAE) in rodents. The statistical analysis procedures chosen for these studies are often suboptimal, either because of violations of the assumptions of the procedure or because the analysis selected is inappropriate for the research question. In this paper, we discuss the types of research questions frequently asked in EAE studies and suggest appropriate and useful research designs and statistical methods that will optimize the information contained within the data. We also discuss other troublesome issues such as missing data, atypical disease profiles, and power analysis.

Keywords: Multiple sclerosis; Experimental allergic encephalomyelitis; Statistics; Data analysis; Study design; Statistical significance

1. Introduction

Research in multiple sclerosis is often conducted using animal models of the disease. One common approach is the use of experimental autoimmune encephalomyelitis (EAE) in rodents. Our review of a sample of the relevant literature revealed that data from such studies are generated, analyzed, and reported in many different and sometimes incorrect ways. These inconsistencies in procedure create difficulty in interpreting and comparing the results of similar studies across different investigators. In this paper, we will discuss the types of questions addressed by the data commonly collected and analyzed in EAE studies and the issues associated with analyzing these data. We will also make suggestions for appropriate and useful methods of analyzing and presenting these data in a way that optimizes the information contained therein.

1.1. Common scales/measures in EAE research

Studies employing the EAE animal model for multiple sclerosis are conducted with a variety of species of rodents, the majority of which are a strain of mice. Although we realize that different species are utilized, in this paper we will generically refer to experimental units as mice although the issues discussed and the statistical analyses proposed may be applied to studies involving other rodents. Additionally, the data presented in the examples are based on real data but some values have been changed to fit the specifics of the analytical issues being discussed. Data and the syntax for conducting the analyses are available at http://www.lsi.ku.edu/lsi/researchdata/eae/. Also, we have noted in the

* Corresponding author. Tel.: +1 785 864 0561; fax: +1 785 864 5323.
E-mail address: jmarquis@ku.edu (J.G. Marquis).
[1] Current Address: Department of Educational Psychology, University of Nebraska, Lincoln, NE 68588, USA.
[2] Current Address: Department of Mathematics and Statistics, Washburn University, Topeka, KS 66621, USA.
[3] Current Address: Department of Pharmaceutical Sciences, College of Pharmacy-Glendale, Midwestern University, 19555 N. 59th Avenue, Glendale, AZ 85308, USA.

0165-5728/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneuroim.2005.08.020

reference list some introductory texts that may be useful for the interested reader (Kleinbaum et al., 1998; Lee and Wang, 2003; Siegel and Castellan, 1988; Tabachnick and Fidell, 2001).

Several types of clinical data are gathered from mice in EAE studies. Typically a clinical score is reported that rates the animals' daily signs during the clinical disease course. It should be noted that, in order to avoid possible bias, the investigator should perform the scoring of clinical signs in a blinded manner. Ideally, the investigator should be blinded as to the group (e.g., drug versus vehicle; genetically modified versus wild type) that each mouse belongs to in a study. Researchers have developed a variety of ordered categorical (ordinal) rating scales to provide clinical scores. Most common is a rating scale from 0 to 5, with 0 indicating no signs and 5 indicating either a moribund state or death (e.g. Fenyk-Melody et al., 1998; Kassiotis et al., 1999; Kennedy et al., 1990; Korner et al., 1995; Martin et al., 1995; Moore et al., 1984; Ruddle et al., 1990; Selmaj et al., 1991, 1998; Yu et al., 1996). Also found in the literature were studies using a 0-to-4 scale (Baker et al., 1994), a 0 to 6 scale (Korner et al., 1997; Riminton et al., 1998), a 0-to-7 scale (Hooper et al., 1998), a 0 to 8 scale (Chakrabarty et al., 2003, 2004), and a 0-to-9 scale (Gold et al., 2004). Among studies using a 0-to-5 scale, however, the intermediate steps 1 to 4 unfortunately often do not correspond to the same clinical signs. Some of these differences may be unavoidable due to the unique characteristics of the particular EAE model that is being studied. Table 1 provides some examples of the clinical scales used in published EAE mouse studies. Although these differences make interpretation from one study to another more difficult, the focus of this paper is not to address the clinical scale per se but to provide some suggestions regarding the analysis and presentation of whatever data have been collected. However, greater statistical power is generally achieved with a greater number of points in a scale, assuming that each point represents a distinct level of disease activity.

In addition to the clinical scale, researchers can use other interval level measures such as daily weight or relative strength, e.g., the length of time the animals can hang onto a rod by their forelimbs (Pedchenko and LeVine, 1998). Many of these measures meet the data assumptions necessary for parametric tests and the parametric techniques discussed in this paper can be used with these types of measures. However, the presentation in this paper is primarily concerned with clinical scale values.

2. Types of questions

EAE studies most often focus on three types of questions about differences between treatments or groups: (1) questions about differences in the level or extent of disease between treatment groups, (2) questions about differences in the length of time until the occurrence of a specified event, and (3) questions about differences in rates or percentages of the occurrence of certain specified events. To answer each of these types of questions, the researcher needs to consider the measurement characteristics of the

Table 1
Sample clinical scale variations

Score	Example 1	Example 2	Example 3
	Source: Pedchenko and LeVine (1998)	Source: Kassiotis et al. (1999)	Source: Chakrabarty et al. (2005)
0	Normal	Normal/no disease	Normal
1	Limp tail and weight loss	Tail limpness	Flaccid tail, piloerection, and/or weight loss (>.4 g the first day, >.1 g thereafter)
2	Hind limb weakness causing righting impairment	Paraparesis with clumsy gait	Hind limb weakness causing righting difficulty from a supine position
3	Hind limb partial paralysis, incontinence	Hind limb paralysis	Hind limb weakness causing righting inability within 8 s from a supine position
4	Hind limb paralysis	Hind and fore limb paralysis	Hind limb weakness causing limping and abnormal gait and/or incontinence
5	Death or moribund requiring sacrifice	Death	Partial (one limb) hind limb paralysis
6	–	–	Total (both) hind limb paralysis plus forelimb weakness
7	–	–	Hind limb paralysis and forelimb weakness or paralysis resulting in a side resting position
8	–	–	Death or moribund requiring sacrifice