Chapter 13
He Wasn’t There Again Today

Richard A. O’Keefe
University of Otago, New Zealand

Nathan Rountree
University of Otago, New Zealand

ABSTRACT
In this chapter, the authors discuss the characteristics of data collected by the New Zealand Centre for Adverse Drug Reaction Monitoring (CARM) over a five-year period. The authors begin by noting the ways in which adverse reaction data are similar to market basket data, and the ways in which they are different. They go on to develop a model for estimating the amount of missing data in the dataset, and another to decide whether a drug is rare simply because it was only available for a short time. They also discuss the notion of “rarity” with respect to drugs, and with respect to reactions. Although the discussion is confined to the CARM data, the models and techniques presented here are useful to anyone who is about to embark on an association mining project, or who needs to interpret association rules in the context of a particular database.

INTRODUCTION
Having collected data on medical conditions and their treatments, it seems to make sense to derive association rules as pioneered by Agrawal et al. (1993). It would be nice to be able to report such things as “when conditions A and B are present with treatment C, then we have an unacceptably high confidence of seeing adverse reaction X.” Before we embark on such an analysis, it is wise to consider the characteristics of the database: does the support/confidence framework make sense here? Are we going to be affected by rare cases and missing data? And to what extent? It should be remembered that association mining was developed in the context of market basket analysis; in what way is our data like or unlike a set of market baskets?

These questions became pertinent to us as we began to analyse data from New Zealand’s Centre for Adverse Reaction Monitoring (CARM). CARM “…collects and evaluates spontaneous reports of adverse reactions to medicines, vaccines, herbal

DOI: 10.4018/978-1-60566-754-6.ch013
products and dietary supplements from health professionals in New Zealand. Currently the CARM database holds over 50,000 reports and provides New Zealand-specific information on adverse reactions to these products, and serves to support clinical decision making when unusual symptoms are thought to be therapy related.” (http://carm.otago.ac.nz/index.asp?link=carm) CARM is believed to produce the highest reporting rate of adverse drug reactions in the world, relative to both number of doctors and size of population (Zolezzi & Parsotam, 2005).

**MARKET BASKETS VS. ADVERSE REACTIONS**

Let’s start by considering how we get our cases in market basket analysis.

1. A shopper decides to buy some items on a shopping list and makes a shopping trip to get them. During this trip,
2. s/he enters a shop,
3. selects some items,
4. is unable to locate others,
5. and chooses others on impulse.
6. S/he pays at the cash register and leaves.
7. The register record is later analysed.

There are demographic facts about the shopper like age, sex, and home district that are normally missing from the cash register record. It is not that such data are not available in principle, it is that they are not measured or recorded. Loyalty cards may permit such information to be gathered. Researchers might still not have access to it for privacy reasons.

Market basket analysis is sometimes presented as giving us information about shoppers. Without loyalty card or banking information to link one shopping episode to another, what we actually have is information about trips, not people. If someone habitually shops for food on Mondays and cleaning products on Wednesdays, we will never learn about associations between food products and cleaning products for that shopper.

We do not learn about people who wish to buy goods we have for sale but who never enter the shop, or who enter and leave without making any purchase. Perhaps they do not like the lighting, or the scents of the cleaning products make them feel sick. (Such scents are often unpleasantly strong.)

We do not learn about the things people wanted to buy but could not find. The free market does not work very well when there are too many intermediaries: manufacturers cannot set prices to suit customers who are never allowed to see those products by supermarkets. If someone wants to buy brand X guacamole whenever he buys mashed potato flakes, but only brand Y, a sort of watery slime, is on offer, market basket analysis will never discover the association that wasn’t allowed to occur. In particular, we may fail to discover an association between two products because the two are not offered at the same time.

We cannot easily tell which purchases were intentional and which were the result of manipulation (“impulse purchases”). To some extent that can be discovered by experimental manipulation.

There is important information about customer response to prices that could be mined from cash register records, but in market basket association mining that information is ignored.

There are three characteristics of the market basket setup:

1. Every purchase is recorded and available for analysis.
2. Purchase numbers are generally high.
3. The aim of analysis is to increase these numbers, and other people are continually working to increase them anyway.

Now let’s think about adverse reaction monitoring, the area that we are interested in.