Citation for published item:

Pennington L, James P, McNally R, Pay H, McConachie H. Analysis of compositional data in communication disorders research. *Journal of Communication Disorders* 2009,**42** 1 18-28.

Analysis of Compositional Data in Communication Disorders Research

Lindsay Pennington, Peter James, Richard McNally, Helen Pay, and Helen McConachie

Newcastle University, UK

Analysis of Compositional Data in Communication Disorders Research

As clinicians and researchers in the field of communication and its disorders we are interested in the use of single behaviors, such as gaze monitoring, voice amplitude or percentage of syllables stuttered, between and within groups over time (Jones et al., 2005; Legerstee and Reddy, 2007; Ramig, Countryman, Thompson, and Horii, 1995). We also use multiple measures whose scores are independent of each other. For example, scores on several tests assessing different language domains may be used as dependent variables in experiments involving children with specific language impairment. When measuring single behaviors or multiple behaviors that are not interdependent we can use traditional parametric statistical tests, if the data are normally distributed.

In some areas of communication, however, we want to examine behaviors that are part of a set, and which are mutually exclusive. This is particularly true in the field of adult-child conversation, where utterances and nonverbal behaviors have been classified in terms of the function they play in developing discourse, for example initiating conversational exchanges, responding to the interlocutor (Bloom, Rocissano and Hood, 1976; Francis and Hunston, 1992; Sinclair and Coulthard, 1992); and their underlying communicative intent, such as providing information, directing a listener’s attention (Gallagher, 1981; Ninio, Snow, Pan and Rollins, 1994; Sinclair and Coulthard, 1992; Wetherby, Cain, Yonclas and Walker, 1988). If counts of behaviors are undertaken, the frequency with which individual behaviors are used can be compared across time and across groups. However, if behaviors are coded continuously, rather than at specific time points, people will vary in the total number of behaviors they produce. Using frequency counts in these situations can potentially mask underlying differences between groups. For example, two groups of children may both produce an average of 20 requests for joint attention in ten minutes. However, one group may produce approximately 60 communication behaviors in total and the other group may produce 100. The overall patterns of communication and distribution of individual behaviors are very different. Therefore, although frequencies may be appropriate for the comparison of key behaviors that are targeted in interventions (Fey et al., 2006; Yoder and Warren, 2002), they are less satisfactory for the analysis of differences between *patterns* of behaviors across entire data sets.

To standardize communication data sets across research participants, we can calculate proportions of behaviors from frequency counts (Nicholas, 2000; Nicholas and Geers, 1997; Rollins, 2003; Yont, Snow and Vernon-Feagans, 2003). However, as proportions the behaviors are no longer independent of each other. Unlike frequencies, high proportions of one behavior are necessarily accompanied by lower proportions of others. Furthermore, when contrasting groups, if one group produces higher proportions of one behavior, the comparison group is likely to produce higher proportions of at least one different behavior. Testing both of these behaviors independently, without taking into account the interdependence of the proportions, may give two statistically significant results. However, this may be an over estimation of the true group differences because the two proportions tested are related. Independence of behaviors is a key assumption of standard parametric statistical tests and to be safe one should analyze only one behavior at a time. Yet analyzing only one behavior in a composite data set is wasteful of data and does not allow comparison of the entire patterns of behaviors between groups. In an attempt to circumvent these difficulties researchers have calculated mean proportions of behaviors within data sets and described the differences they see between groups (Nicholas, 2000; Nicholas and Geers, 1997; Rollins, 2003). But, we cannot be certain whether differences between groups are true by visual inspection alone. For example, we cannot fully account for within group variation. As researchers we are left with a dilemma. Should we use frequency counts to analyze all behaviors, ignoring the differences in total communication behavior? Or, should we standardize data using proportions and test only key behaviors? Some researchers have handled multiple testing of proportions by means of Multiple Analysis of Variance (Yont et al, 2003). However, this technique assumes that the data are normally distributed and such an assumption may be invalid for proportions data (Aitchison, 1986).

Our own aims, as researchers in the field of early intervention and communication development, have been to standardize data sets from parents and children and to investigate patterns of behaviors within the data set across groups. As there is a dearth of research into the effects of interventions on multiple communication behaviors, we have not been able to predict reciprocal change, with increases in behavior X being accompanied by decreases in behavior Y. We have therefore needed to examine the use of all communication behaviors, rather than selecting a small number for analysis. In the late 1990's we began to use Coda (Aitchison, 1986). This is a statistical technique developed for compositional data; i.e. data where the component variables add up to a constant, such as percentages or proportions; which has been used in the study of medicine, marine biology, anthropology, and geology (Billheimer, Guttorp,and Fagan, 2001; Bracci, Bull,and Grynpas, 1998; Buccianti, Mateu-Figueras,and Pawlowsky-Glahn, 2006; Engas and Soldal, 1992). We found the technique useful in showing differences between clinical groups (Pennington and McConachie, 2001a), and have recently applied a development of the technique to a larger data set and to data from an intervention study. In this paper we describe our approach and give two examples of its use with data from two different clinical groups. We show how the technique can test differences in patterns of behaviors across entire data sets, both across groups and within groups across time. We suggest that this technique will be a useful addition to the toolbox of basic and applied communication researchers, allowing much more accurate and extensive testing of composite data than has been previously possible.

The statistical technique

The key feature of the statistical technique is that it involves deriving new variables, which are transformations of the proportions expressed as percentages, so that the new variables can be analysed as if they had a multivariate normal distribution. The transformations are chosen to suit the type of experiment. Groups can then be analysed using multivariate tests to compare two or more transformed variables at a time. Alternatively they can be analysed using univariate tests to compare the variables singly.

For example, when conducting an experiment in which we are comparing two clinical groups on a single occasion, we select the variables in which we predict change and we calculate the proportions of all behaviors in the data set. If the proportions in which we are interested are $W_1, W_2,\ldots W_r$ and W_s is a reference proportion i.e. the remaining proportion:

$$W_s=1 - W_1 - W_2 - \ldots - W_r.$$

Then we define new variables $Y_1, Y_2,\ldots,Y_r$ by:-

$$Y_1=\ln\{(W_1+1)/(W_s+1)\},\ Y_2=\ln\{(W_2+1)/(W_s+1)\},\ldots\ Y_r=\ln\{(W_r+1)/(W_s+1)\}$$

To put this simply, we add 1 to each of the proportions and then calculate Y_1 etc as the logarithm of the ratio of the adjusted proportion W_1+1 etc to the adjusted proportion W_s+1. Since we add 1 to the proportions this means that the ratios $(W_1,+1)/(W_s+1)$, $(W_2,+1)/(W_s+1)$, etc. are all positive and not less than 1, meaning that we can always take logarithms of them.

The transformed Y variables are known as ‘log-ratios’ and for convenience we refer to them in this article as *Type I Transformations*. They range in value from minus infinity to plus infinity (depending on the size of demoninator proportion W_s relative to the numerator proportions). For example, log(15 / 20) = -0.288 and log(25 / 20) = 0.223.

In practice we can fit statistical models to them on the assumption that they have a multivariate normal distribution. This means that any Y considered on its own has a normal distribution, while overall there is correlation between these observations for the same individual. The assumption of multivariate normality is, according to Aitchison (1986) valid for many compositional datasets. Furthermore, taking the logarithm of a variable in order to render its distribution less skew is a well known method.

It is also straightforward to carry out standard multivariate techniques for comparing samples of data on these Ys. Because the transformed variables are the logarithms of the ratio of proportions, it is then natural to compare ratios of the mean proportions for each sample. We might, for example, be interested in testing the hypothesis that the ratio of the mean value of W_1 to the mean value of W_s is larger for the intervention sample than for the control. This hypothesis is equivalent to saying that the mean of Y_1 is larger in the intervention than in the control sample. To test if there is a significant difference between the two groups in one or more of the means of the variables $Y_1, Y_2,\ldots,Y_r$ we use a multivariate version of the two-sample t test, the 'two-sample Hotelling T^2 test'. However, this test does not directly reveal *which* of the Y's has a difference. To reveal which individual Y's differ in their means across the two groups we can use standard t tests.

In experiments in which we have communication data on each participant before and after an intervention there may be significant within-participant correlation. In such cases we analyse paired differences in transformed proportions. More specifically, if the proportions measured as percentages before the intervention are $U_1, U_2,\ldots U_r$ and the proportions after the intervention are $V_1, V_2,\ldots V_r$ then the new variables $Y_1, Y_2,\ldots,Y_r$ are defined by:-

$$Y_1=ln\{(V_1,+1)/(U_1+1)\},\ Y_2=ln\{(V_2,+1)/(U_2+1)\},\ldots\ Y_r=ln\{(V_r,+1)/(U_s+1)\}$$

For convenience we refer to these new variables as *Type II Transformations*. Following this transformation, the analysis can then proceed on the assumption that the *Y*s have a multivariate normal distribution. The hypothesis that the mean of the first proportion is larger in the intervention sample than in the control is equivalent to saying that the mean of Y_1 is greater than zero. We would then use the one-sample Hotelling T^2 test to assess if there is a change in the mean value of one or more of the variables $Y_1, Y_2, \ldots, Y_r$.

The above description and the examples which follow concern predicted differences between groups and across times. Bonferroni adjustments may lead to *p* values that are extremely conservative and are not recommended for *a priori* predicted differences (Sankoh, D'Agostino and Huque, 2003; Schulz and Grimes 2005). However, if proportions are compared but no prior hypotheses predict the directions of differences, Bonferroni adjustments should be made to the significance level to ensure overall significance of such tests is within acceptable limits, such as 0.05 (Sankoh, D'Agostino and Huque, 2003; Schulz and Grimes 2005). As with other statistical test the power of the tests proposed in this paper is proportional to the square root of the sample sizes. Full technical details about these techniques are given in the Appendix. The computation can easily be done using STATA (Stata Corp 2001) or R (R is an open source software for statistical computing and graphics, which is freely available at http://www.r-project.org). Examples of their use with real data are given below.

Example 1

Children with motor impairments have been observed to take a respondent role in conversation, rarely introducing new topics or initiating exchanges. They have also been observed to use communication for a restricted range of functions (Harris, Jones, Brookes and Grant, 1986; Light, Collier and Parnes, 1985). A decade ago we began to examine how the acquisition of intelligible speech affects the development of conversation skills by children with motor impairments (Pennington and McConachie, 2001a, 2001b). We have continued to study children's speech and have added to the original data set. We hypothesize that children whose speech is intelligible to their parents will behave similarly to non-disabled children: they will take a more equal role in conversation, starting more conversational exchanges and producing fewer response turns than children without intelligible speech; and they will produce higher proportions of provision of information and request functions than will children whose speech is unintelligible to their parents.

To investigate these hypotheses we videotaped the conversation of two groups of children with motor impairments and their parents. Group One consisted of 28 children whose speech was not intelligible to their parents out of context. Children communicated using vocalizations, facial expression, body movement, limited gestures, and by reaching or leaning towards desired objects. Group Two comprised 25 children who had speech that was affected by their motor disorder, but who were intelligible to their parents out of context. They usually communicated via speech, but supplemented this with facial expression and limited gesture. All children had motor disorders that affected at least two of their limbs. For further information on the children see Pennington and McConachie (20001a, b) and Pennington, Thomson, James, Martin and McNally (submitted for publication).

We coded ten minutes of conversation, from each parent-child pair, at two levels. The first level showed the structure of conversation; the second showed the functions for which children used communication.

Discourse Moves: Method and Results

Five codes were used to show how conversation was structured. *Initiations* started conversational exchanges; *response-initiations* were produced in reply to another person's initiations, but demanded some response of their own; *responses* were produced in reply to an initiation or response-initiation; *follow ups* acknowledged a response; and *no response* coded a failure to respond when obliged to do so (Francis and Hunston, 1992; Sinclair and Coulthard, 1992). For example

Adult:	Hey, look at this.	Initiation
Child:	What is it?	Response-initiation
Adult:	It's a spaceship	Response
Child:	Oh yeah	Follow up

Each turn taken by the child and their parent was coded using one of the mutually exclusive structural codes. In this paper we report the interaction patterns of the children only, as an exemplar. We hypothesized that children in Group One would produce proportionally more responses and fewer initiations than children in Group Two. The mean proportions of the moves used by both groups are shown in Table 1.

Insert Table 1 about here

To reveal if our predictions may be correct, and to conduct an initial exploration of the differences in proportions across the two groups, we compared the proportions of all five codes between the groups using two-sample *t* tests. As these tests may be biased by the skewness of the data, and the interdependence of the codes, this is only a preliminary exploration. Results of the *t* tests, as shown in Table 1, suggest that initiations and responses may differ across the two groups, as hypothesized. However, response-initiations and no responses also appear to differ across the groups. For the analysis, we denoted these four respectively by X_1, X_2, X_3 and X_4 and we also define $X_r = 100 - X_1 - X_2 - X_3 - X_4$. We then analyze the Type I Transformations:- $Y_1 = \ln\{(X_1+1)/(X_r+1)\}$, $Y_2 = \ln\{(X_2+1)/(X_r+1)\}$, $Y_3 = \ln\{(X_3+1)/(X_r+1)\}$ and $Y_4 = \ln\{(X_4+1)/(X_r+1)\}$.

Insert Table 2 about here

There is strong evidence that the ratio of initiations and the ratio of response-initiations are larger in Group 2 (see Table 2). Overall, there is evidence that these two log-ratios differ between the groups (Hotelling 2-sample $F(2,50) = 6.97$, $p = .002$). This means that the children in Group Two produced proportionally more initiations and response-initiations, which demand some response from the parent. However, there is no direct reciprocal relationship between initiations and response-initiations, on the one hand, and responses and no responses, on the other. As initiations and response-initiations increase for Group Two, there is not a corresponding decrease in responses and no responses. The reduced proportions arising from higher initiations and response-initiations are spread across all other moves. Using all data in the compositional analysis, we partially support our hypotheses, but also gain further insight into the pattern of conversation behavior. The overall pattern of behaviors shows the mainly respondent role taken by all the children with motor disorders in the study, with higher rates of initiations by children whose speech is intelligible to their parents.

Communicative Functions: Method and Results

With only five codes it is easy to see patterns in proportions and differences between the two groups. However, when more behaviors are coded proportional differences between groups may not be as clear cut, and the need to use all codes in the analysis to show communication patterns becomes more pressing. We used ten codes to represent the functions for which children used communication: (a) requests for joint attention, (b) requests for information, (c) requests for objects or actions, (d) provisions of information, (e) requests for clarification, (f) provision of clarification (repetitions and revisions of a previous communication signal), (g) expressions of self (showing personality e.g. humor, sarcasm), (h) confirmations and denials, (i) simple acknowledgements of previous utterances (e.g. good boy), and (j) signals that were indecipherable. We hypothesized that the intelligible children in Group Two would produce proportionally more functions that add extra information to the conversation (provisions of information) and more functions that direct their partner and involve requests (requests for joint attention, information, object, actions, and clarification).

We hypothesized that children in Group One would use higher proportions of functions that added little extra information or exerted little control over the conversation (confirmations, denials, acknowledgements).

Insert Table 3 about here

As a first stage in the analysis we compared the differences in proportions of these ten codes between the groups using two-sample *t* tests. From Table 3, the codes which show group differences, and which need further examination, are:- (b) request for information, (d) provision of information, (f) provision of clarification, (g) expression of self, (h) confirmation or denial, and (i) acknowledgment, partially supporting our predictions. For convenience, we denote these six respectively by X_1, X_2, X_3, X_4, X_5 and X_6 and we also define $X_r = 100 - X_1 - X_2 - X_3 - X_4 - X_5 - X_6$. We then analyze the 'log-ratios' transformations:- $Y_1, Y_2, \ldots Y_6$ defined in the same way as before.

Insert Table 4 about here

There is evidence (Table 4) that all the first three ratios - (b) requests for information, (d) provisions of information and (f) provision of clarification - are higher in Group 2, whereas the other three ratios - (g) expressions of self , (h) confirmations and denials, (i) simple acknowledgements of previous utterances - are lower in Group 2. Overall, there is a significant difference between the groups (Hotelling 2-sample $F(6,46) = 7.00$, $p < .001$).

Discussion

Thus, the results show that we did not find all the differences we predicted between the two groups. Children whose speech was intelligible to their parents used a greater proportion of functions that developed the theme and that added content to the conversation (requests for information and provisions of information), and fewer functions which exerted little control, than those children whose speech was not intelligible. However, when the inter-dependence of the variables was taken into account the other request functions were not found to differ between the two groups. Analysis of all behaviors suggested that intelligible children were developing the discussion within the conversation and adding content, but were not using their increased intelligibility and ability to control conversation to simply direct the behavior of their partner.

Example 2

Children with autism have difficulties in the domains of communication and social interaction, and show restricted and repetitive behaviours. Identification of the disorder is increasingly possible at two and three years of age. Some of the predictors of better outcome include development of joint attention with a parent, and ability to imitate others (Charman, 2003). And, there is now some evidence for the effectiveness of early intervention that trains parents to develop strategies that facilitate their children’s communication, such as increasing joint attention, and using fun words, praise, imitation, expansion and pretend games (Aldred, Green and Adams, 2004; Kaiser and Hancock, 2003; McConachie, Randle, Hammal and Le Couteur, 2005). We conducted a further study, using a development of the coding scheme described in Example 1, to investigate if intervention also had an impact on the interaction patterns used by parents.

Method

In an evaluation study by McConachie and colleagues (2005) 52 parents attended group training sessions in which they were taught strategies to facilitate their child’s communication, with weekly group meetings over a three month period (McConachie et al, 2005; Sussman, 1999). In the controlled phase of the study (immediate versus delayed intervention), a positive treatment effect was observed in two outcome variables: child vocabulary size, and parents’ observed use of positive strategies to engage their children in interaction (McConachie et al, 2005). We conducted a further examination of the data from this study to investigate if the degree of change in parents’ use of these facilitative strategies was associated with change in conversation patterns. Some of the parents were found to use a high level of facilitative strategies prior to the training. These parents were excluded from the current analysis, and the remaining parents were ranked according to their degree of change in facilitative strategies from before to after training. The top one third of parents who showed a significantly positive increase in strategies following training were classed as 'treatment responders' (n = 10). The bottom one third, who changed least, were classed as 'treatment nonresponders' (n = 11). The groups are small, but the aim of the study was to test if parents may differ in the patterns of conversation used after training.

Parents and children were observed during a 7 to 10 minute play session with a standard set of toys, and 5 minutes of observed interaction were coded. In this study we coded the structure of the conversation according to the scheme used in Example 1, but extended the scheme to include the extent to which responses were compliant with preceding initiations, in order to investigate the synchrony of conversation. In the extended coding scheme the permissible structural moves were *initiations*, *compliant responses*, *other responses* (in which a request is acknowledged but not fulfilled), *no responses, noncompliant responses*, and *follow ups*. Pragmatic functions were not coded.

We compared move proportions for the two groups of parents (responders and nonresponders) before and after training (the children's data are not presented here). We predicted that following intervention parents in the responder group, who had begun to use more facilitating strategies, would also change their discourse patterns, becoming more engaged with their child's interaction by producing fewer no responses and noncompliant responses, and initiating fewer conversational exchanges themselves. We predicted no change for the nonresponder group. Mean proportions of moves produced by parents in each group are shown in Table 5.

Insert Table 5 about here

Results

As our comparisons involved differences between proportions of behaviors before and after training we tested our hypothesis using Type II transformations to reduce the skew of the data. Results did not show the predicted change. As shown in Table 6, parents in the responder group did not change from the first observation to the second. However, parents in the nonresponder group did change following intervention, producing proportionally fewer other responses ($t(10) = -2.27$, $p = .047$), in which they acknowledged their child's requests but did not comply with them, following intervention. There is also some evidence of a corresponding increase in non-compliant responses ($t(10) = 1.68$, $p = .125$) in this group. Overall, the One-Sample Hotelling statistic for these two moves indicates a significant difference ($F(2,9) = 4.46$, $p = .048$).

Insert table 6 about here

To help understand the changes in communication patterns better we conducted supplementary analyses. Visual inspection of the untransformed mean move proportions presented in Table 5 seemed to suggest that parents differed prior to intervention. Parents in the nonresponder group appeared to produce proportionally more initiations and fewer compliant responses than parents in the responder group before treatment. After intervention the groups seemed very similar in move proportions. This suggested that parents in the nonresponder group became more similar to those in the responder group following intervention, using a more facilitative pattern of interaction after training. However, as shown in Table 7, when we analysed logged data to reduce skewness, initiations and compliant responses did not differ between the two groups prior to intervention. But, parents in the nonresponder group were found to produce proportionally more other responses ($t(10) = -2.18$, $p = .042$) than the responder group prior to intervention. There was also evidence of a similar difference in non-compliant responses ($t(10) = -1.83$, $p = .083$). Thus, differences were found between the groups, but not in the pattern suggested by the untransformed mean proportions.

Insert Table 7 about here

Although Table 7 provides comparisons for logged data there is still some possibility of serious bias due to skewness of the data and so to compare the two groups Type I Transformations similar to the ones used for Example 1 should be used instead. Accordingly we derived a new variable as follows:-

Y_1 = log of ratio of sum of other responses and noncompliant responses proportions (%) plus 1 to initiations and follow-ups proportions (%) plus 1.

Univariate analysis using this new composite variable confirmed that parents in the nonresponder group produced more other responses and noncompliant responses than parents in the responder group ($t(10) = -3.53$, $p = .002$) prior to training.

Discussion

Findings of the current study suggest that prior to the intervention the parents who increased their use of facilitating strategies in the study McConachie et al (2005) differed in their interaction patterns from those who acquired fewer facilitating strategies. Parents who did not use more use facilitating strategies, the nonresponders, used more other responses in which they acknowledged their child’s interaction but did not act on it, and more noncompliant responses than the parents in the responder group before training. Following intervention the groups became more similar, with parents in the nonresponder group using fewer other responses. The results also suggest that there may also have been an increase in noncompliant responses for the nonresponder group. However, the moves that differed between groups and changed for the nonresponder group were used seldom. The most frequently used moves (initiations, compliant responses and follow ups) did not differ between groups, nor did they change following intervention. The question therefore arises of the clinical significance of the differences observed. Our feeling for the small sample here is that the differences observed are probably not clinically relevant. However, this supposition should be tested further in future studies of the intervention. Thus, our conclusions from the study of the data in Example 2 are that the intervention is effective in facilitating engagement strategies in some parents, but it may be difficult to identify in advance which parents are likely to respond from those who will find it harder to change their strategies. In addition, a relatively short intervention may not lead to major change in the structure of interaction between parents and children.

A similar training programme has recently been found to lead to positive change in interaction patterns for parents and their young children with motor disorders, with increases in responses and fewer directives observed for parents, and higher initiations and directives observed for children, following parent training (Pennington et al, submitted). It is possible that the lack of similar findings for with parents of children with autism is due to conversation between parents and children with motor disorders being more amendable to change. Children with motor disorders can initiate, take increased control, and engage in reciprocal dialogue if parents give them time and opportunity to do so. However, children with autism have a core difficulty with social-communication and may not be able to change within a short period of time to engage in reciprocal and complex conversation exchanges, even if parents change their discourse patterns to provide more opportunities for them to do so. Thus, parents' attempts to engage their children in conversation may be constrained. For parents and children with autism, outcome measures other than those depending on microanalysis of interaction structure may be more relevant; for example, coding percentage of time parents and children are sharing the same focus of attention (shown to increase significantly in the 'responder' group – unpublished data), and percentage of parents' utterances that pertain to the child's focus of attention.

Discussion

The above exemplars have shown that the statistical technique we described can be used to analyse differences in multiple, interdependent behaviors between groups and across time. In the exemplars mean proportions suggested that there were differences in several behaviors within a coding frame, either between groups or across time. However, when the interdependency of the data was taken into account in the analysis, the number of behaviors showing change reduced. Analysing untransformed mean proportions only may therefore lead to an over estimation of the change in behaviour patterns. The results of the exemplars also show that the techniques we describe are capable of detecting small differences between groups, which is extremely useful when comparing large numbers of behaviors. As with any statistical test used in clinical research, researchers and clinicians must decide whether statistically significant results are clinically significant (Wade, 2005).

The above examples illustrate how the statistical technique we have described allows the comparison of differences between composite data sets. It can be used to compare a small number of key variables. However, its main advantage is that it permits the analysis of all variables within a set, if needed. Thus, none of the data that potentially add to the differences between groups are ignored. The differences may be predicted and tested *a priori*, if research or theory allows. The technique can also be useful in the exploration of data sets about which few predictions can be made. In the latter case, data may be preliminarily examined through the usual route of means, standard deviations and *t* tests, and then tested thoroughly, taking into account the interdependence of the variables, using the technique above with Bonferroni adjustments to the significance level in multiple comparisons where we have no prior hypotheses. As with the design of any study researchers must guard against the chances of not detecting a true difference between groups, by ensuring adequate power through the inclusion of sufficient numbers of research participants.

Interaction data are often composite, being proportions or percentages. The technique described here could be a useful addition to the communication researcher's toolbox allowing the comparison of group proportional data to demonstrate the presence of patterns of disorder or the clinical effectiveness of interventions.

Acknowledgements

Lindsay Pennington was supported by the Health Foundation and the UK Department of Health National Health Service Research and Development Programme during the preparation of this manuscript. Peter James is funded by North of England Children's Cancer Research.

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. London: Chapman Hall.

Aldred, C., Green, J. and C., A. (2004). A new social communication intervention for children with autism: Pilot randomized controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry*, *45*, 1420-1430.

Billheimer, D., Guttorp, P. and Fagan, W. F. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, *96*, 1205-1214.

Bloom, L., Rocissano, L. and Hood, L. (1976). Adult-child discourse: Developmental interaction between information processing and linguistic knowledge. *Cognitive Psychology*, *8*, 521-552.

Bracci, P. M., Bull, S. B. and Grynpas, M. D. (1998). Analysis of compositional bone density data using log ratio transformations. *Biometrics*, *54*, 337-349.

Buccianti, A., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (Eds.). (2006). *Compositional data analysis in the geosciences: From theory to practice*. London: Geological Society.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? . *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 358*, 315-324.

Engas, A. and Soldal, A. V. (1992). Diurnal variations in bottom trawl catch rates of cod and haddock and their influence on abundance indices. *ICES Journal of Marine Science, 49*, 89-95.

Fey, M. E., Warren, S. F., Brady, N., Finestack, L. H., Bredin-Oja, S. L., Fairchild, M., Sokol, S. and Yoder, P.J. (2006). Early effects of responsivity education/prelinguistic milieu teaching for children with developmental delays and their parents. *Journal of Speech, Language and Hearing Research, 49*, 526-547.

Francis, G. and Hunston, S. (1992). Analysing everyday conversation. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 123-161). London: Routledge.

Gallagher, T. M. (1981). Contingent query sequences within adult-child discourse. *Journal of Child Language, 8,* 51-62.

Harris, M., Jones, D., Brookes, S. and Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British Journal of Developmental Psychology, 4*, 261-268.

Jones, M., Onslow, M., Packman, A., Williams, S., Ormond, T., Schwarz, I. and Gebski, V. (2005). Randomised controlled trial of the Lidcombe programme of early stuttering intervention. *British Medical Journal, 331,* 659-.

Kaiser, A. P., and Hancock, T. B. (2003). Teaching parents new skills to support their young children's development. *Infants and Young Children, 16*, 9-21.

Legerstee, M., and Reddy, V. (2007). What does it mean to communicate for infants? *Infant Behavior and Development 30*, 177-179.

Light, J., Collier, B. and Parnes, P. (1985). Communicative interaction between young nonspeaking physically disabled children and their primary caregivers: Part ii - communicative function. *AAC Augmentative and Alternative Communication, 1*, 98-107.

McConachie, H., Randle, V., Hammal, D., and Le Couteur, A. (2005). A controlled trial of a training course for parents of children with suspected autism spectrum disorder. *Journal of Pediatrics*, *147,* 335-340.

Nicholas, J. G. (2000). Age differences in the use of informative/heuristic communicative functions in young children with and without hearing loss who are learning spoken language. *Journal of Speech, Language and Hearing Research*, *43*, 380-394.

Nicholas, J. G. and Geers, A. E. (1997). Communication of oral deaf and normally hearing children at 36 months of age. *Journal of Speech, Language and Hearing Research, 40*, 1314-1327.

Ninio, A., Snow, C. E., Pan, B. A. and Rollins, P. R. (1994). Classifying communicative acts in children's interactions. *Journal of Communication Disorders*, *27,* 157-187.

Pennington, L., and McConachie, H. (2001a). Interaction between children with cerebral palsy and their mothers: The effects of speech intelligibility. *Journal of Language and Communication Disorders*, *36*, 371-394.

Pennington, L., and McConachie, H. (2001b). Predicting patterns of interaction between children with cerebral palsy and their mothers. *Developmental Medicine and Child Neurology*, *43*, 83-90.

Ramig, L. O., Countryman, S., Thompson, L. L. and Horii, Y. (1995). Comparison of two forms of intensive speech treatment for Parkinson disease. *Journal of Speech and Hearing Research, 38*, 1232-1251.

Rollins, P. R. (2003). Caregivers' contingent comments to 9-month-old infants: Relationships with later language. *Applied Psycholinguistics*, *24*, 221-234.

Sankoh, A. J., D'Agostino R. B. and Huque, M.F.(2003). Efficacy endpoint selection and multiplicity adjustment in methods in clinical trials with inherent multiple endpoints. *Statistics in Medicine* 22, 3133-3150.

Schulz, K. F. and Grimes D. A. (2005). Multiplicity in randomised trials I: endpoints and treatments. The Lancet 365: 1591-95.

Sinclair, J. and Coulthard, M. (1992). Towards an analysis of discourse. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 1-34). London: Routledge.

StataCorp. 2001. *Statistical Software: Release 9.0.* College Station, TX: Stata Corporation

Sussman, F. (1999). *More Than Words: Helping parents promote communication and social skills in children with autism spectrum disorder.* Toronto: Hanen Centre.

Wade, D. (2005). But is the difference clinically significant? *Clinical Rehabilitation, 19*, 349-353.

Wetherby, A. M., Cain, D. H., Yonclas, D. G. and Walker, V. G. (1988). Analysis of intentional communication of normal children from the prelinguistic to the multiword stage. *Journal of Speech and Hearing Research, 31*, 240-252.

Yoder, P. J. and Warren, S. F. (2002). Effects of prelinguistic milieu teaching and parent responsivity education on dyads involving children with intellectual disabilities. *Journal of Speech, Language and Hearing Research, 45*, 1297-1310.

Yont, K. M., Snow, C. E. and Vernon-Feagans, L. (2003). The role of context in mother-child interactions: An analysis of communicative intents expressed during toy play and book reading with 12-month-olds. *Journal of Pragmatics, 35*, 435-454.

Table 1.

Univariate Comparisons of Children's Move Proportions

Move	Group 1		Group 2				95%CI
	M	*SD*	*M*	*SD*	*t (51)*	*p*	
Initiation	8.83	7.20	24.73	10.41	6.53	<.001	11.01, 20.80
Response-initiation	1.46	2.92	4.06	3.37	3.01	.004	.87, 4.34
Response	74.00	16.80	60.62	14.67	-3.07	.003	- 22.13,-4.63
Follow up	.69	1.76	.81	1.80	026	.798	-.85, 1.11,
No response	14.90	14.92	9.23	8.57	-1.67	.101	-12.49, 1.15

Table 2

Group Differences in Logged Ratios of Move Proportions

Move[a]	*M*	*t(51)*	*p*	95% CI
Initiation	1.34	3.77	<.001	.63, 2.06
Response-initiation	1.06	3.14	.003	.38, 1.74
Response	.03	.10	.923	-.65, .72

Note. [a.] Log of the ratio of the move proportion plus 1 to the proportion plus 1 of follow up and no response

Table 3

Univariate Comparisons of Children's Functions Expressed as Percentages

	Group 1		Group 2				
Function	*M*	*SD*	*M*	*SD*	*t(51)*	*p*	95%CI
Req. joint attention	4.34	4.61	5.22	5.95	.61	.547	-2.04, -3.80,
Req. information	1.32	3.54	6.72	5.02	4.57	<.001	3.03, 7.78,
Req. object or action	1.95	3.33	3.42	3.33	1.60	.115	-.37, 3.31,
Prov. information	14.50	10.85	36.81	9.64	7.87	<.001	16.62, 28.00
Req. clarification	.75	2.74	1.03	2.24	.40	.689	-1.11, 1.67
Prov. clarification	1.48	2.90	3.73	3.59	2.52	.015	.455, 4.04,
Exp. self	5.18	6.63	2.27	3.01	-2.02	.049	-5.81, -.01
Conf. or denial	30.77	19.88	17.08	9.24	-3.15	.003	-22.41, -4.97
Acknowledgement	36.70	21.38	20.84	9.44	-3.42	.001	-25.17, -6.55
Unintelligible	3.80	4.93	3.04	3.03	-.67	.509	-3.05, 1.53

Table 4

Group Differences in Logged Ratios[a] of Function Percentages

	M	*t(51)*	*p*	95% CI
Req. information	1.29	5.33	<.001	.81, 1.78
Provisions of information	1.11	3.92	<.001	.54, 1.68
Provision of clarification	.58	1.76	.085	-.08, 1.24
Expressions of self	-.53	-1.40	.168	-1.28, .23
Confirmations or denials	-.41	-1.20	.236	-1.08, .27
Acknowledgement	-.51	-1.67	.101	-1.12, .10

[a.] Log of the ratio of the proportion plus 1 to the proportion plus 1 of requests for objects or actions, requests for information, requests for clarification and unintelligible signals

Table 5.

Parent Move Proportions Pre and Post Treatment Expressed as Percentages

	Nonresponder group				Responder group			
Move	Pre-intervention		Post-intervention		Pre-intervention		Post-intervention	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Initiations	77.20	11.50	73.70	11.30	74.20	15.30	74.10	13.60
Compliant responses	8.00	6.70	10.60	10.90	13.90	11.00	11.40	9.50
Other responses	1.10	1.60	.20	.80	.00	.00	.30	1.10
No responses	.60	1.40	1.40	2.10	2.30	3.30	.90	1.50
Noncompliant responses	.70	1.20	4.30	7.30	.00	.00	.90	1.90
Follow ups	12.50	6.50	9.70	7.60	9.60	5.20	12.30	9.50

Table 6.

Post- minus Pre-Intervention Differences for Parents in Logs of Proportions

	M	*t*	*df*	*p*	95%CI
<u>Non-Responder Group</u>					
Initiations	-.019	-.622	10	.548	-.089, .050
Compliant responses	.022	.861	10	.409	-.035, .080
Other responses	-.008	-2.266	10	.047	-.016, .000
No responses	.007	.876	10	.402	-.011, .026
Noncompliant responses	.034	1.676	10	.125	-.011, .078
Follow ups	-.026	-.910	10	.384	-.089, .038
<u>Responder Group</u>					
Initiations	.000	.009	9	.993	-.091, .092
Compliant responses	-.021	-.484	9	.640	-.118, .076
other responses	.003	1.000	9	.343	-.004, .011
No responses	-.013	-1.339	9	.213	-.036, .009
Noncompliant responses	.009	1.447	9	.182	-.005, .022
Follow ups	.023	.643	9	.536	-.056, .101

Table 7.

Differences between Groups of Parents in Logged Move Proportions Pre and Post Intervention

	M	*t(19)*	*p*	95%CI
Pre-Intervention				
Initiations	-.018	-.536	.598	-.090, .053
Compliant responses	.051	1.489	.153	-.021, .122
Other responses	-.011	-2.183	.042	-.021, .000
No responses	.016	1.558	.136	-.006, .038
Noncompliant responses	-.007	-1.831	.083	-.015, .001
Follow ups	-.025	-1.096	.287	-.074, .023
Post-Intervention				
Initiations	.001	.045	.965	-.065, .068
Compliant responses	.008	.199	.844	-.071, .086
Other responses	.001	.248	.807	-.007, .009
No responses	-.004	-.552	.587	-.021, .012
Noncompliant responses	-.032	-1.473	.157	-.077, .013
Follow ups	.023	.710	.486	-.045, .091