HEALTH PSYCHOLOGY OPEN

Report of empirical study

Health Psychology Open
July-December 2015: 1–9

DOI: 10.1177/2055102915615338
hpo.sagepub.com
SAGE

Using computerized text analysis to assess communication within an Italian type 1 diabetes Facebook group

Alda Troncone[1], Crescenzo Cascella[1], Antonietta Chianese[1] and Dario Iafusco[2]

Abstract

The purpose of this study was to assess messages posted by mothers of children with type 1 diabetes in the Italian Facebook group "Mamme e diabete" using computerized text analysis. The data suggest that these mothers use online discussion boards as a place to seek and provide information to better manage the disease's daily demands—especially those tasks linked to insulin correction and administration, control of food intake, and bureaucratic duties, as well as to seek and give encouragement and to share experiences regarding diabetes and related impact on their life. The implications of these findings for the management of diabetes are discussed.

Keywords

children, computerized text analysis, Facebook group, mothers, type 1 diabetes

Introduction

Thanks to a general trust in its anonymity, the Internet, especially trough its forums and blogs, has become a place where people are comfortable to reveal their feelings (Pounders et al., 2015) and their psychological needs (Ogden and Bennett, 2015), as well as being a rich source of health related data, collected via online surveys and questionnaires (Faber and Dubé, 2015; Lyons et al., 2015; Rosenbaum and White, 2015; Strodl et al., 2015).

In addition, for people who suffer from chronic diseases, social networking websites are a way to exchange information and find a virtual place to talk freely about difficult issues and share their experience in writing (Farmer et al., 2004; Hawn, 2009; Shaw et al., 2000).

Although online chronic disease groups are widely available on the Internet, research exploring the information that patients or their relatives request and share is limited. Little is known about the nature of the virtual communities that congregate on Facebook.

Some chronic disease group researchers have used content analysis to understand the typology and nature of responses that are directly affected by the illness (Alpers et al., 2005; Davison and Pennebaker, 1997; Klemm et al., 1999; Mazzoni and Cicognani, 2014; Thompson et al., 2015).

Among studies examining individuals suffering from type 1 diabetes, Ravert et al. (2003) performed content analysis of the messages posted by adolescents in public web-based forums via qualitative data analysis. Social support and information were the two prominent aspects of these messages. Life tasks (queries relative to transitions or activities regarding normal adolescent development), medical care (queries related to specific concern about health care, the medical system, service, etc.), management (queries related to problems in managing diabetes), and intrapsychic (queries related to affective issues such as state of mind and attitude) were the other request topics identified by this analysis.

Greene et al. (2011) qualitatively evaluated a sample of discussions on the most popular Facebook groups dedicated to diabetes. Two investigators evaluated the posts by

[1]Second University of Naples, Naples, Italy
[2]Department of the Woman, of the Child and of the General and Specialized Surgery, Second University of Naples, Naples, Italy

Corresponding author:
Alda Troncone, Department of Psychology, Second University of Naples, Viale Ellittico 31, 81100 Caserta, Italy.
Email: alda.troncone@unina2.it

developing a thematic coding scheme and identified four prominent topics: clinical information-sharing, patient-center management (sensitive aspects of diabetes management unlikely to be revealed to doctor or available through professional consultation), interpersonal support/community-building, and marketing activity and data collection functions unique to the Facebook diabetes group (advertising, promotional posts, etc.).

De la Torre-Díez et al. (2012) analyzed the main purposes and use of the Facebook and Twitter groups concerning chronic diseases. The 527 diabetes groups found on Facebook and Twitter focused primarily on research issues. Content analysis showed that the majority of users were involved in social networks to support research and find cures. Other users include users who had the aim of supporting patients and their relatives, who want to report and share information about the illness, who are dedicated to fundraising to pay for people's treatment, to prevent the disease, to encourage people to use boluses of insulin or promoting the new technologies for the care of the disease.

In order to provide further data to the existing research and to quantitatively describe what Facebook users share, the present study was designed to evaluate a sample of discussion on a specific Italian Facebook group dedicated to type 1 diabetes. "Mamme e diabete" is a Facebook group where the caregivers of children with type 1 diabetes—mostly mothers—exchange messages with one another. Libreri and Graffigna (2012) have already provided a preliminary description of the main features of this Facebook group. However, as stated by these authors in their original paper, it is important to deepen the understanding of the roles and needs of the actors participating in these online exchanges.

Messages posted by users of "Mamme e diabete" were assessed using computerized text analysis to better understand the nature of interactions taking place among participants.

This study sought to address the following questions:

1. Which concepts are associated with the word diabetes?
2. What are the most frequent topics of the messages?
3. Are there longitudinal changes in discourse structure so that differences will be found among messages posted during different months?

Methods and data

A representative sample of text serves as the raw input data. It was collected from June to September 2014. This text constitutes the text corpus. The participants' writings were copied and corrected for grammatical errors. The names and surnames of the writers as well as the dates and times of the posts and messages were deleted. There were 40,355 posts copied and pasted into a text file.

Textual analysis of the participants' writing was performed via a content analysis software package called the T-Lab 5.1 (Lancia, 2004). The T-Lab operates through a distinction between the context units (CUs—fixed chunks of text that divide the total body of the input text, for example, paragraphs) and lexical units (LUs—the different words or categorizations of words, for example, lemmas, semantic classes, or dictionary categorizations).

This software produces matrices representing relationships between the analysis units (CU and LU), which are the basis of all of the T-Lab's operations. T-Lab can identify and categorize the semantic content in the text. This procedure is grounded in the assumption that the semantic content of a text can be depicted in terms of how the words associate with each other. The meaning of a word consists of the associations it has with the other words in the text (Salvatore et al., 2010). According to the same assumption, a semantic content in the text corresponds to a pattern of co-occurrences of words (i.e. a set of words that tend to be present together across the text).

Two of T-Lab's functions were employed in this research: co-occurrence analysis (computation of word association and sequence analysis of key words) and thematic analyses of the CUs (thematic analyses of elementary contexts (ECs)). The computation of word association is used to present an association chart on a single word that gives an impression of the contextual use of that word in terms of which other words the term under consideration often co-occurs. To analyze the contextual meaning of a word, the software reviews that word's co-occurrence profile with respect to all other sample words. The relative closeness of one word to another word in the chart is devised according to a parameter called the cosine coefficient (Salton, 1989). The function sequence analysis allows a Markovian analysis of a sequence of key words. These items are LUs that are present in the corpus.

This tool constructs and analyzes two co-occurrence matrices whose respective values are the count of how many times within the analyzed corpus each LU precedes or follows the other in the linear (sequential) structure of the texts (Lancia, 2002, 2012). Beginning from a matrix in which all the predecessors and all the successors of each item (i.e. LU or theme) are recorded, T-Lab calculates the transition probabilities (Markov chains) between predecessors and successors. In other words, this tool highlights the transition (i.e. what comes before or after any "x") via a Markovian approach. Thematic analyses of ECs (ECs are a type of CU) provide a representation of the corpus' content through a small number of significant thematic clusters. The clusters divide the media discourse into a number of "themes" that represent the content of the media discourse. Thematic clusters are internally homogeneous and externally heterogeneous. One can plot the evolution of thematic clusters along a longitudinal dimension. In particular, the evolution of thematic clusters, identified in

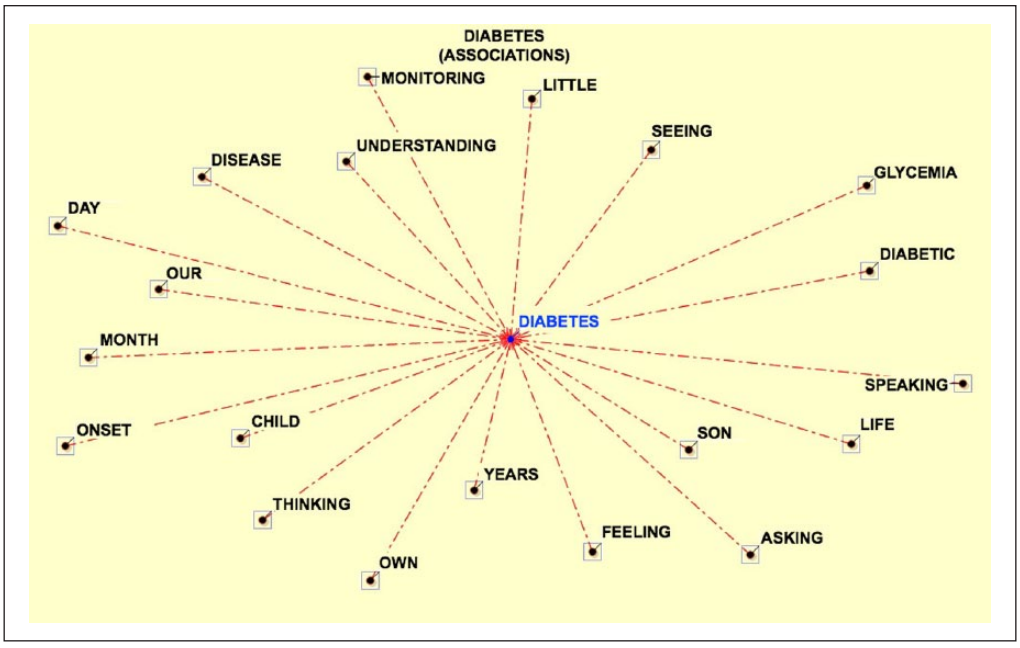

Figure 1. "Diabetes" association chart.

Table 1. List of words associated to the word diabetes and related occurrence and cosine coefficient values.

Word	Occurrence	Cosine coefficient
Son	1018	0.454
Years	901	0.436
Child	675	0.377
Ours	508	0.377
Understanding	565	0.334
Seeing	564	0.325
Diabetic	433	0.322
Life	347	0.313
Feeling	442	0.307
Thinking	459	0.296
Month	406	0.289
Illness	243	0.285

terms of their weight in each month during the study period, was analyzed.

All studies were approved by local ethics committees and were performed according to the principles of the Helsinki Declaration II.

Results

Preliminary data from this study have already been communicated in Troncone et al. (in press).

An association chart of the word "diabetes" was constructed with T-Lab to identify which concepts are associated with diabetes in the entire text (Figure 1). Table 1 shows occurrence and cosine coefficient values of the first 12 words related to the word "diabetes." All words in the graph are closely related to the word "diabetes" in the text corpus. The higher a word's cosine coefficient with respect to the word "diabetes," the closer that word is positioned to the center.

The results of the sequence analysis of key words that are closer to the word diabetes are showed in Figures 2 and 3 and Table 2.

Figure 2 presents items that have the highest probability of coming before (predecessors) the word "diabetes."

Figure 3 shows items that have the highest probability of coming after (successors) the word "diabetes."

Table 2 shows the sorted list of predecessors and successors for each selected item. The list is in descending order according to the probability values ("Prob."). For example, Table 2 shows that the probability that "type" will follow the word "diabetes" is 0.034 or 3.4 percent.

Thematic analyses of ECs were used to identify the most frequent topics. The output of the ECs thematic analyses converged on a structure of six thematic clusters. Each cluster received a label based on the qualitative interpretation performed by analyzing the ECs grouped in each theme as well as the words connected to each cluster. In Table 3, the six thematic clusters are presented listing the chi-square test values of the most important LUs in each cluster. A high chi-square value means that the corresponding LU is central to a cluster.

The six clusters were interpreted as follows: "correction" is characterized by issues concerning correction of insulin dose and monitoring of blood glucose level; "diabetes and life" focuses on emotional correlates related to diabetes and its management; "hi group!" refers to all contents related to the group including participation, encouragement, and

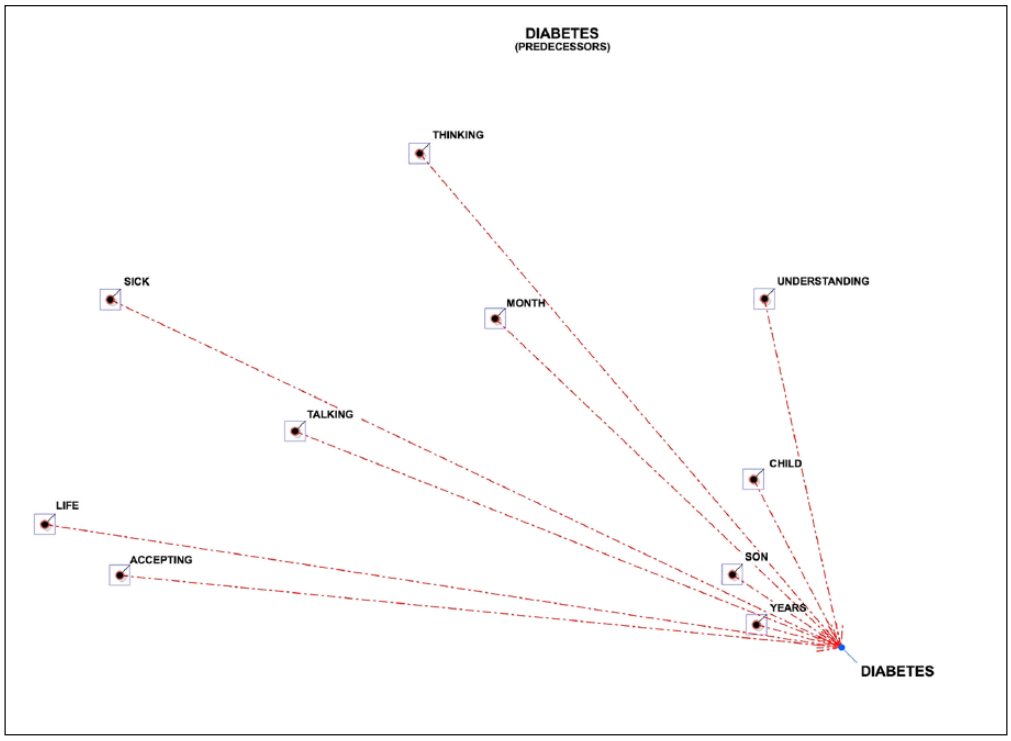

Figure 2. Predecessors of the word "diabetes."

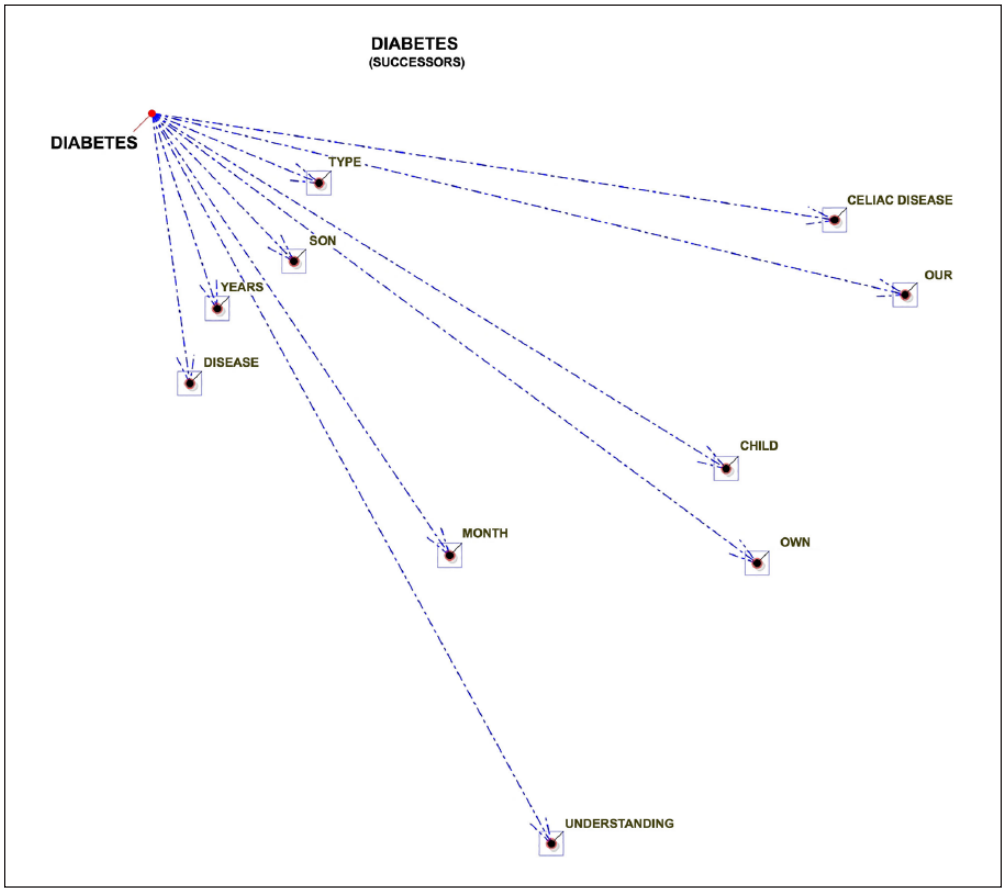

Figure 3. Successors of the word "diabetes."

Table 2. List of predecessors and successors to "diabetes."

Predecessors	Prob.	Successors	Prob.
Years	0.071	Type	0.034
Son	0.04	Son	0.029
Child	0.022	Years	0.028
Understanding	0.011	Disease	0.021
Talking	0.01	Month	0.011
Month	0.01	Child	0.009
Accepting	0.009	Celiac disease	0.009
Life	0.008	Our	0.008
Thinking	0.007	Own	0.008
Sick	0.007	Understanding	0.007
Finding Out	0.007	Feeling	0.007
Onset	0.007	Putting	0.006
Living With	0.006	Seeing	0.006

emotional support; "bureaucracy" contains issues regarding compensation, bureaucracy procedures, and practices; "needle" is a cluster characterized by words closely related to all the procedures associated with the insulin administration such as injection, prick, and insulin pump; "food" is characterized by issues concerning food, eating, and carbohydrate level.

Considering the relative weight of each thematic cluster in the overall number of ECs, the largest cluster was "correction" (20.21%) followed by "diabetes and life" (18.47%), "hi group" (16.78%), "bureaucracy" (15.82%), "needle" (15.32%), and "food" (13.39%).

Figure 4 uses a two-dimensional (2D) space to show the positions of the thematic clusters and relevant words with respect to the two factors. The axes are a result of a formal statistical analysis and are a way of making clear the structure of the discourse under examination. These factors are calculated by lumping together similar items to distinguish groups of different things (Sengers et al., 2010).

Regarding the distribution in the semantic space of the thematic clusters, four main spaces grouped the clusters. The first space is shared by three clusters ("correction," "needle," and "food"). These clusters are all linked in that they are all themes of monitoring blood glucose values. They are all procedures that relate to insulin administration and related procedures including adjusting insulin intake based on blood glucose values and matching insulin dose to the carbohydrate intake. Their proximity in the factorial space represents their thematic closeness. The "bureaucracy," "diabetes and life," "hi-group" clusters are in a different factorial space—they are distant from the previous three clusters and are relatively distant from each other as well. These three clusters do not share the factorial space because of the distance among the themes that they describe.

Finally, the six thematic clusters were crossed with the month of publication to identify the longitudinal changes in the discourse structure. Figure 5 shows the longitudinal dimension of the six thematic clusters and their relative distributions in each month.

The relative weight of "correction," "diabetes and life," and "food" was prominent in the first month, but lost prominence in July and August; it increased in September. On the other hand, "bureaucracy" and "needle" themes become progressively more salient especially in July and August. The "hi-group" loses relevance from June to September.

Discussion

Our first goal was to define representations of diabetes via co-occurrence analysis (computation of word association and sequence analysis of key words). The data showed that "son," "type," and "years" are the most relevant words surrounding the word "diabetes." The words "child," "ours," "understanding," "disease," "seeing," "diabetic," "life," "feeling," "thinking," "speaking," and "accepting" are also closely related. These results suggest diabetes-related words and likely represent the perspective of mothers who focus on the burden associated with the illness and its duration, as well as the effects of illness on the child and family, the understanding of the mechanisms of disease, and acceptance of the illness.

Evaluation of the most frequent topics suggests that mothers of children with type 1 diabetes use online discussion boards to seek and provide information about food and insulin correction, to compare and share experiences regarding diabetes and related impact on their lives, to seek and give encouragement, to share coping strategies regarding bureaucratic tasks and duties, and to better manage the disease's daily demands. This is especially true for those tasks linked to insulin administration (multiple daily insulin injections, blood glucose monitoring by finger pricking, subcutaneous administration of insulin with portable infusion pump, etc.). By analyzing the LU connected to each cluster, it should be noted that the word hypoglycemia is unexpectedly included in the cluster "food" and not "correction." This association is likely explained by the use of food as a hypoglycemia treatment. Most episodes of symptomatic hypoglycemia can be effectively self-treated by ingestion of glucose tablets or carbohydrate in the form of juice, soft drinks, milk, crackers, or a full meal.

Evaluating the longitudinal changes in discourse structure, these themes are relatively stable over the study period, with some changes in the relative weight of some clusters. In particular, the changes in the relative weight of the themes "correction," "diabetes and life," and "food" are a probable consequence of the challenging management of diabetes when children are at school or away from home. In June and September, children usually spend 4–8 hours per day at school. Thus, more attention is given to diabetes care at school. Examples include blood glucose testing, eating appropriately, administering insulin as needed, the presence of knowledgeable individuals to assist the student during the school, variations in activity and food intake, recognizing symptoms and treatment of hypoglycemia, and so on. On the other hand,

Table 3. Six clusters obtained from the thematic clustering of elementary contexts from the corpus.

Cluster no. 1	Correction	Cluster no. 2	Diabetes-life	Cluster no. 3	Hi-group
Word	χ^2	Word	χ^2	Word	χ^2
Correction	1532.502	Life	918.595	Strength	3462.731
Glycemia	1187.721	Diabetes	496.939	Hug	2999.255
High	994.218	Son	461.941	Giada	2552.92
Night	714.158	Our	427.762	Praying	1757.411
Monitoring	663.816	Living	333.729	Strong	1711.66
Dose	602.984	Friend	332.713	Welcome	1225.376
Lantus	599.567	Thinking	229.167	Good	1005.919
Decreasing	501.973	Feeling	219.828	Hello	867.462
Values	495.099	Understanding	215.519	Courage	794.185
Lunch	383.939	Anger	191.823	Mariagrazia	727.5
Basal	331.478	Crying	187.985	Big kisses	682.22
Fast	278.162	Talking	186.773	You	615.922
Sleeping	267.773	Parent	178.19	Closeness	485.764
Dinner	276.773	People	164.315	God	458.526
Lowering	269.251	Accepting	160.367	Mommy	441.454
Bed	265.248	Disease	157.318	News	433.669
Morning	222.459	Child	155.1	Heart	363.917
Moon	220.763	Worse	151.948	Giada	327.39
Pizza	218.831	Reasoning	151.228	Big	310.252
Insulin	201.96	Moment	139.223	Compliments	289.166
Cluster no. 4	Bureaucracy	Cluster no. 5	Needle	Cluster no. 6	Food
Word	χ^2	Word	χ^2	Word	χ^2
Benefit	1027.592	Insulin pump	1755.574	Carbohydrates	2964.981
Allowance	1000.225	Using	1488.553	Sugar	1571.01
Month	678.803	Needle	1272.155	Eating	1254.91
Inps[a]	655.866	Pen (injection)	1105.69	Grams	899.231
Paying	609.71	Chemist	597.854	Hypoglycemia	872.817
Law	534.378	Putting	579.181	Milk	851.212
Disability	495.038	Refrigerator	381.553	Fruit	786.094
Request	481.578	Test strips	373.385	Juice	727.374
Diabetic	423.386	I-port	353.797	Pasta	677.177
Exemption	388.467	Pricking	346.37	Biscuit	659.103
Municipality	362.756	Belly	332.194	Bread	605.976
Welfare agency	352.351	Injection	321.918	Ice cream	448.3
Years	344.696	Sensor	307.482	Counting	431.529
Certificate	342.931	Finger	272.703	Chilling	430.84
Medical examination	342.917	Use	261.046	Slice	407.379
To breastfeed	337.479	Band-aid	255.399	Snack	405.239
Commission	334.575	To change	228.647	Insulin	353.991
Benefit for carers	310.034	To buy	193.729	Fat	351.727
To benefit from	309.405	Diabetologist	190.217	Potato	282.255
Vaccine	298.031	Cartridge[b]	183.775	Calculation	258.142

[a]Inps: social security service.
[b]Preparations of pen insulin with a built-in glass cartridge.

the increase in the "needle" theme observed in July and August is likely due to increased occurrences of dealing with diabetes in public (e.g. beach) rather than at home. The "bureaucracy" theme also becomes more salient during the summer perhaps because of the interruption of scholastic indemnity with accompanying problems and complaints.

In conclusion, these findings are a useful contribution that enhances knowledge about the typology and nature of responses posted by participants in a social networking

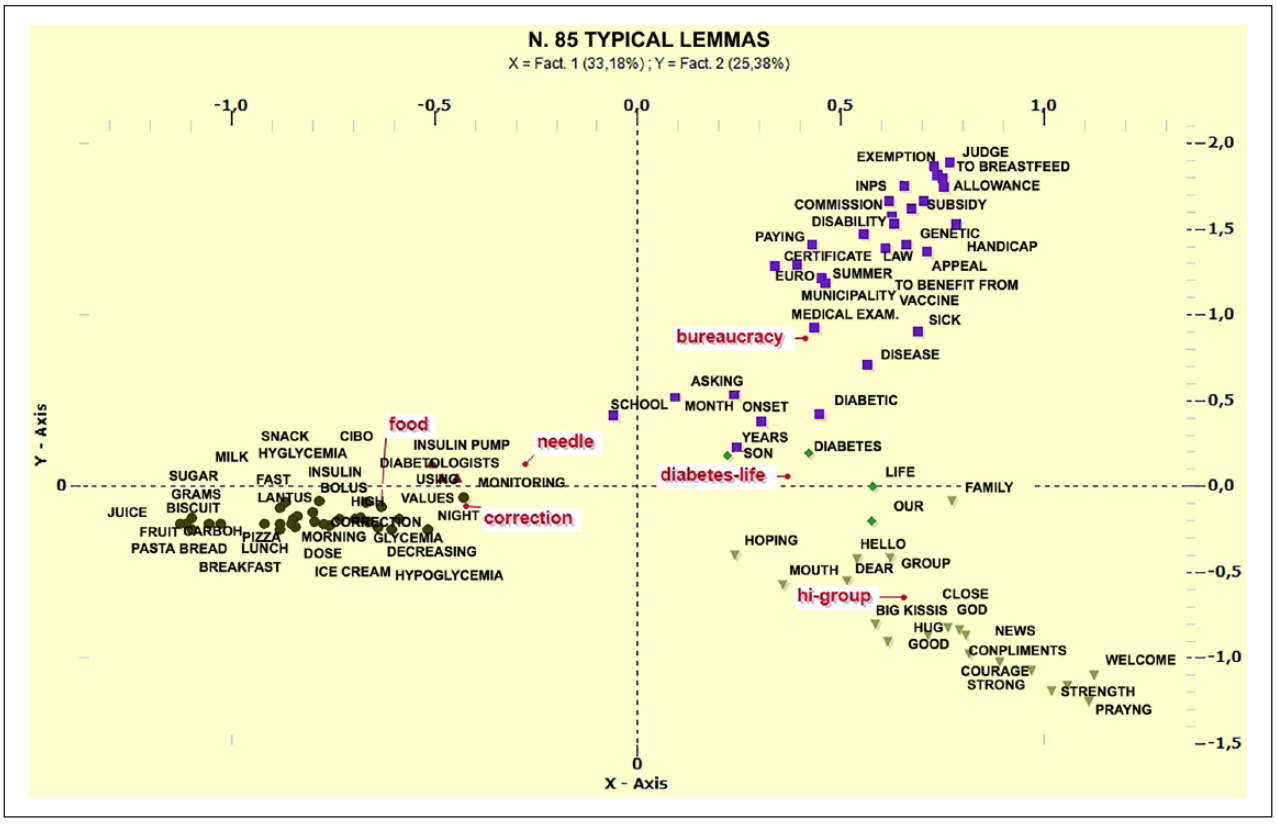

Figure 4. Thematic cluster chart.

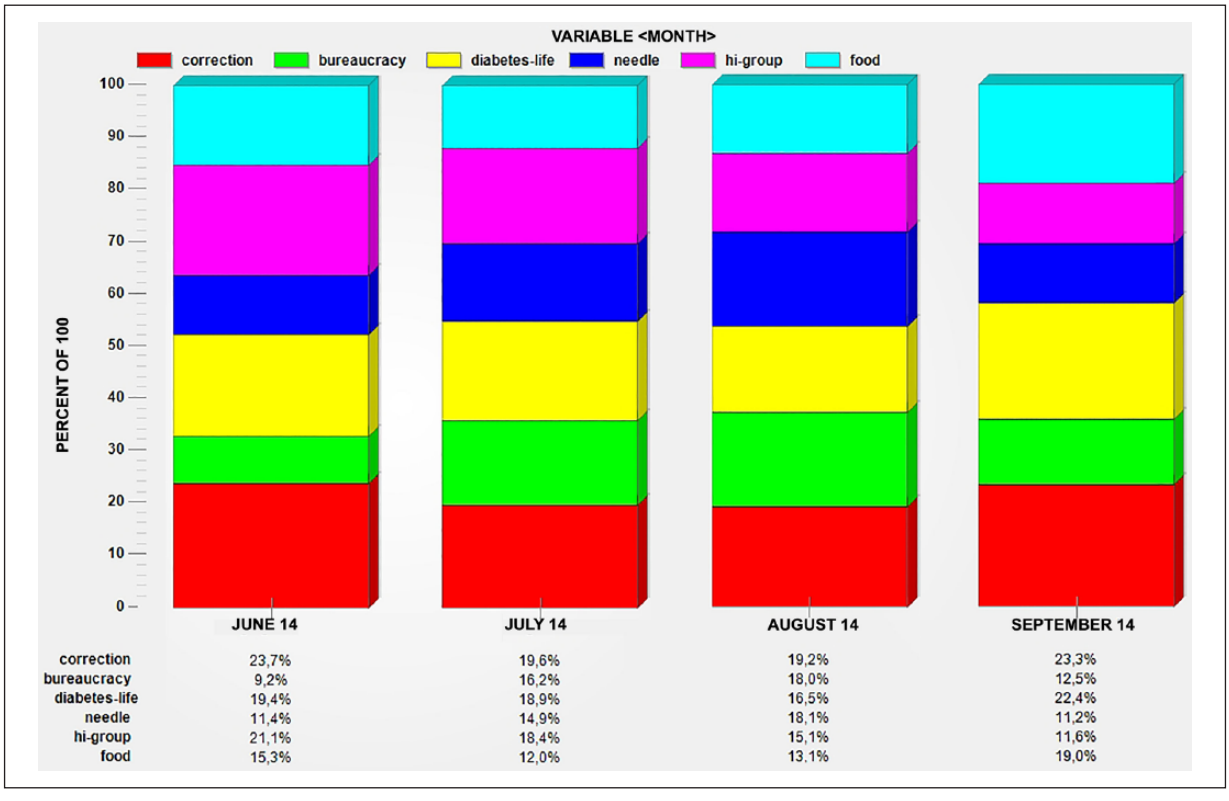

Figure 5. Relative weight of the thematic clusters across time.

website. This knowledge is critical in improving the clinical management of diabetes. The knowledge and nature of responses posted by participants can help health care providers to be cognizant of the main issues that give caregivers concerns. This knowledge is essential for improving the doctor–patient communication. In addition, within the larger effort to use computer technology to address problems and improve life, the comprehension of needs of social networking users can help better configure the Web context (e.g. creation of other virtual communities through online groups) to assist people and achieve improved health and well-being.

As emphasized by Barrera et al. (2002), with additional outcome research, online groups could become validated resources to help the patients with type 1 diabetes better manage their health and cope with chronic disease by promoting and supporting daily self-management. To realize this aim, it will be necessary to collect additional data to deepen the understanding of what ingredients allow the configuration of different Web contexts to reach different aims (e.g. to diffuse information or to provide peer support) as well as different targets (e.g. patients, caregivers, and the general public) (Libreri and Graffigna, 2012).

Acknowledgements

The authors thank the Staff of the outpatient clinics at the pediatric diabetes unit "*G. Stoppoloni*" of the Second University of Naples for their help in data collection, in particular Dr Valeria Taiani, Dr Alessia Piscopo, Dr Alessandra Cocca, Dr Lisa Caredda, Dr Stefania Picariello, Dr Emilia Frongione, Dr Serena Rollato, Dr Fortunato Cotugno, Dr Angela Zanfardino and Dr Santino Confetto. Special thanks to all members of the Facebook group "Mamme e diabete," especially to the founder Mrs Rosanna Sannino.

Declaration of conflicting interests

The authors have no conflicts of interest statement to disclose. No author or related institution has received any financial benefit from research in this study.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Alpers GW, Winzelberg AJ, Classen C, et al. (2005) Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior* 21(2): 361–376.

Barrera M Jr, Glasgow RE, Mckay HG, et al. (2002) Do Internet-based support interventions change perceptions of social support? An experimental trial of approaches for supporting diabetes self-management. *American Journal of Community Psychology* 30(5): 637–654.

Davison KP and Pennebaker JW (1997) Virtual narratives: Illness representations in online support groups. In: Petrie KJ and Weinman J (eds) *Perceptions of Health and Illness: Current Research and Applications*. London: Taylor & Francis, pp. 463–486.

De la Torre-Díez I, Díaz-Pernas FJ and Antón-Rodríguez M (2012) A content analysis of chronic diseases social groups on Facebook and Twitter. *Telemedicine and e-Health* 18(6): 404–408.

Faber A and Dubé L (2015) Parental attachment insecurity predicts child and adult high-caloric food consumption. *Journal of Health Psychology* 20(5): 511–524.

Farmer AD, Bruckner Holt CE, Cook MJ, et al. (2009) Social networking sites: A novel portal form communication. *Postgraduate Medical Journal* 85: 455–459.

Greene JA, Choudhry NK, Kilabuk E, et al. (2011) Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook. *Journal of General Internal Medicine* 26(3): 287–292.

Hawn C (2009) Take two aspirin and tweet me in the morning: How Twitter. Facebook, and other social media are reshaping health care. *Health Affairs* 28: 361–368.

Klemm P, Hurst M, Dearholt SL, et al. (1999) Gender differences on Internet cancer support groups. *Computers in Nursing* 17(2): 65–72.

Lancia F (2002) The logic of a text-scope. Available at: http://www.mytlab.com/textscope.pdf (accessed August 2007).

Lancia F (2004) *Strumenti per l'analisi dei testi. Introduzione all'uso di T-LAB* (Tools for text analysis: Introduction to the use of T-LAB). Milan: Franco Angeli. Available at: http://www.tlab.it

Lancia F (2012) The logic of the T-LAB tools explained. Available at: http://www.tlab.it (accessed September 2012).

Libreri C and Graffigna G (2012) Mapping online peer exchanges on diabetes. *Neuropsychological Trends* 12: 125–134.

Lyons C, Kaufman AR and Rima B (2015) Implicit theories of the body among college women: Implications for physical activity. *Journal of Health Psychology* 20(9): 1142–1153.

Mazzoni D and Cicognani E (2014) Sharing experiences and social support requests in an Internet forum for patients with systemic lupus erythematosus. *Journal of Health Psychology* 19(5): 689–696.

Ogden J and Bennett A (2015) Self-harm as a means to manage the public and private selves: A qualitative study of help seeking by adults. *Health Psychology Open*. Epub ahead of print 29 September. DOI: 10.1177/2055102915605987.

Pounders K, Stowers K, Wilcox G, et al. (2015) Exploring gender and identity issues among female adolescent and young adults who connect in an anonymous platform. *Journal of Health Psychology*. Epub ahead of print 30 September. DOI: 10.1177/1359105315605657.

Ravert RD, Hancock MD and Ingersoll GM (2003) Online forum messages posted by adolescents with type 1 diabetes. *The Diabetes Educator* 30(5): 827–834.

Rosenbaum DL and White KS (2015) The relation of anxiety, depression, and stress to binge eating behavior. *Journal of Health Psychology* 20(6): 887–898.

Salton G (1989) *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer*. Reading, MA: Addison-Wesley.

Salvatore S, Gelo O, Gennaro A, et al. (2010) Looking at the psychotherapy process as an intersubjective dynamic of meaning-making: A case study with discourse flow analysis. *Journal of Constructivist Psychology* 23(3): 195–230.

Sengers F, Raven RP and Van Venrooij AHTM (2010) From riches to rags: Biofuels, media discourses, and resistance to sustainable energy technologies. *Energy Policy* 38: 5013–5027.

Shaw BR, McTavish F, Hawkins R, et al. (2000) Experiences of women with breast cancer: Exchanging social support over the CHESS computer network. *Journal of Health Community* 5: 135–139.

Strodl E, Stewart L, Mullens AB, et al. (2015) Metacognitions mediate HIV stigma and depression/anxiety in men who have sex with men living with HIV. *Health Psychology Open*. Epub ahead of print 30 April. DOI: 10.1177/2055102915581562.

Thompson CM, Crook B, Love B, et al. (2015) Understanding how adolescents and young adults with cancer talk about needs in online and face-to-face support groups. *Journal of Health Psychology*. Epub ahead of print 27 April. DOI: 10.1177/1359105315581515.

Troncone A, Cascella C, Chianese A, et al. (in press) What relatives and caregivers of children with type 1 diabetes talk about: Preliminary results from computerized text analysis of messages posted on the Italian Facebook diabetes group. In: Bassis S, Esposito A, Morabito CF, Pasero E (eds) *Advances in Neural Networks: Computational and Theoretical Issues. Computational Intelligence for Information Communication Technologies.* New York: Springer International Publishing.