Deepest into the sea:
false information about CAM in Italian Web conversations

Mauro MAZZOCUT, Ivana TRUCCOLO, Carlo BIANCHINI

EAHIL 2016 (Seville), Thursday 9th June, 11.30 – 13.10

Parallel Session C, Education II

1

Introduction

- Internet is frequently used by patients to google and to exchange a relevant amount of medical information in oncology

- An undefined part of the available online information is *false*, and a major function of the medical library is to inform patients about risks of misinformation in the internet.

- Libraries and scholars have developed tools to evaluate web sites, but there are not researches that deal with deep content analysis

«False»

- The term "false": it can refer to the presence of wrong words or grammatical or syntactical errors, the absence of evidences or support to the information or to the statement of scientifically unproven facts

Linguistic	Methodologic	Scientific	Intentional
Erroneous	Unfounded	Untrue	Fraudulent

Aim of the study

- To prove that a lot of different false information is present in CAM (Complementary and Alternative Medicine) websites.

- To *qualify* the kind of falsehood present in websites

- To *quantify* the kind of falsehood within each website

- To *explore relations* among the kinds of falsehood and the general content of websites

Sample websites

- From a previous study ("How deep is the see: web intelligence for patient education") 15 CAM (Complementary and Alternative Medicine) websites were selected and analyzed

- 8 are health related websites (3 CAM, 2 Health and fitness, and 3 Oncology),

- 7 are not heath related (3 online newspaper, 2 pseudoscience and conspiracy, and 2 debunking).

- 1 authoritative Italian website was added to be evaluated and used as benchmark

Websites analysis

Analysis was conducted at two levels: *macro* analysis and *micro* analysis

1. Macro analysis: recording of data as author of the text, publication date, last access date, URL address, etc.;

2. HONcode grid was applied to score websites (authoritative, complementarity, privacy, attribution, justifiability, transparency, financial disclosure, and advertising policy)

3. Only 3 out of 15 (20%) website score at least 4 over 8 points assigned by HONcode.

Websites analysis

Micro analysis. To deepen analysis and get more data, website texts were subdivided in items (a sentence going from dot to dot)

1. All the texts were subdivided in totally 1355 items
2. Each item was examined in relation to its possible falsehood (erroneous, unfounded, untrue)
3. Items with relevant scientific content were evaluated by a team of experts chosen for their with different disciplinary profiles (+ a layman)
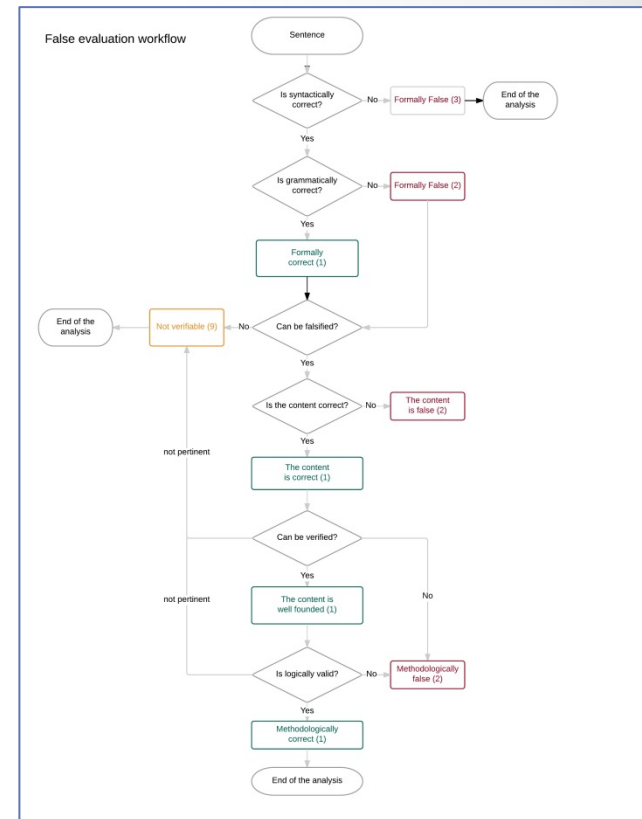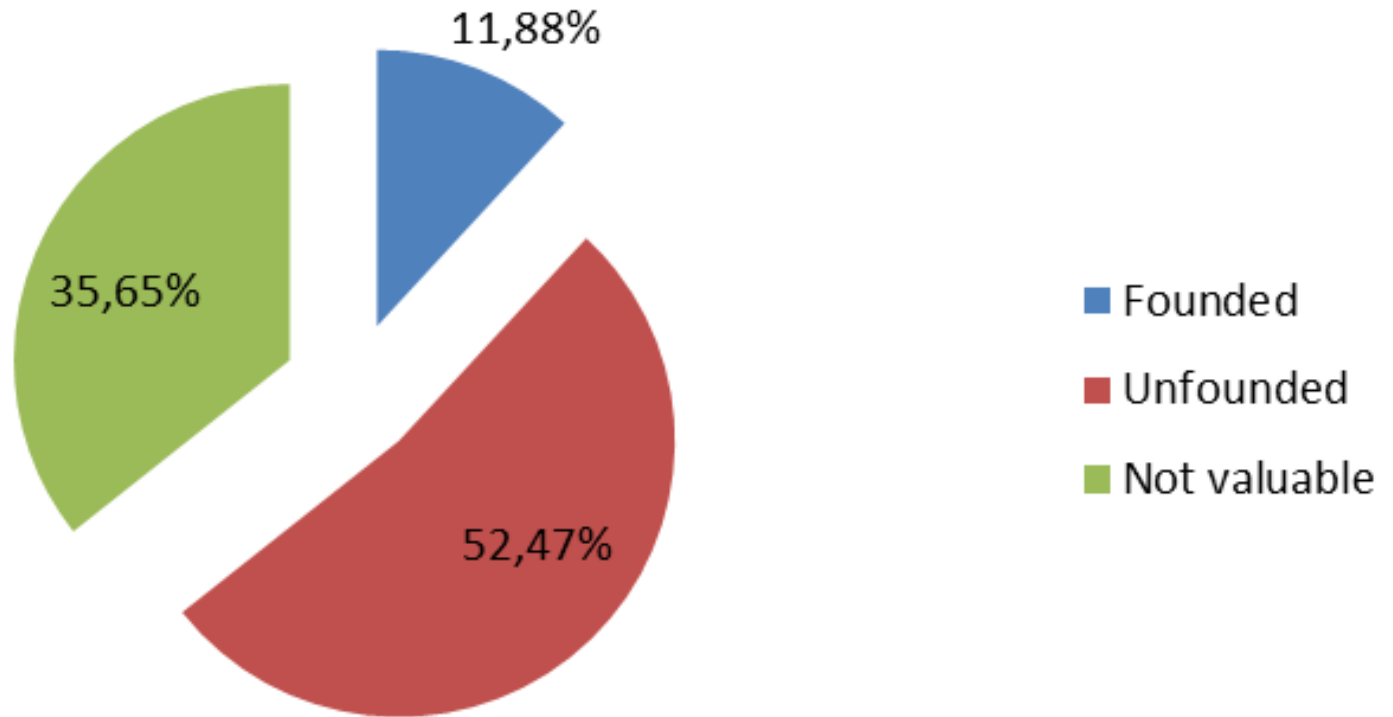
Results
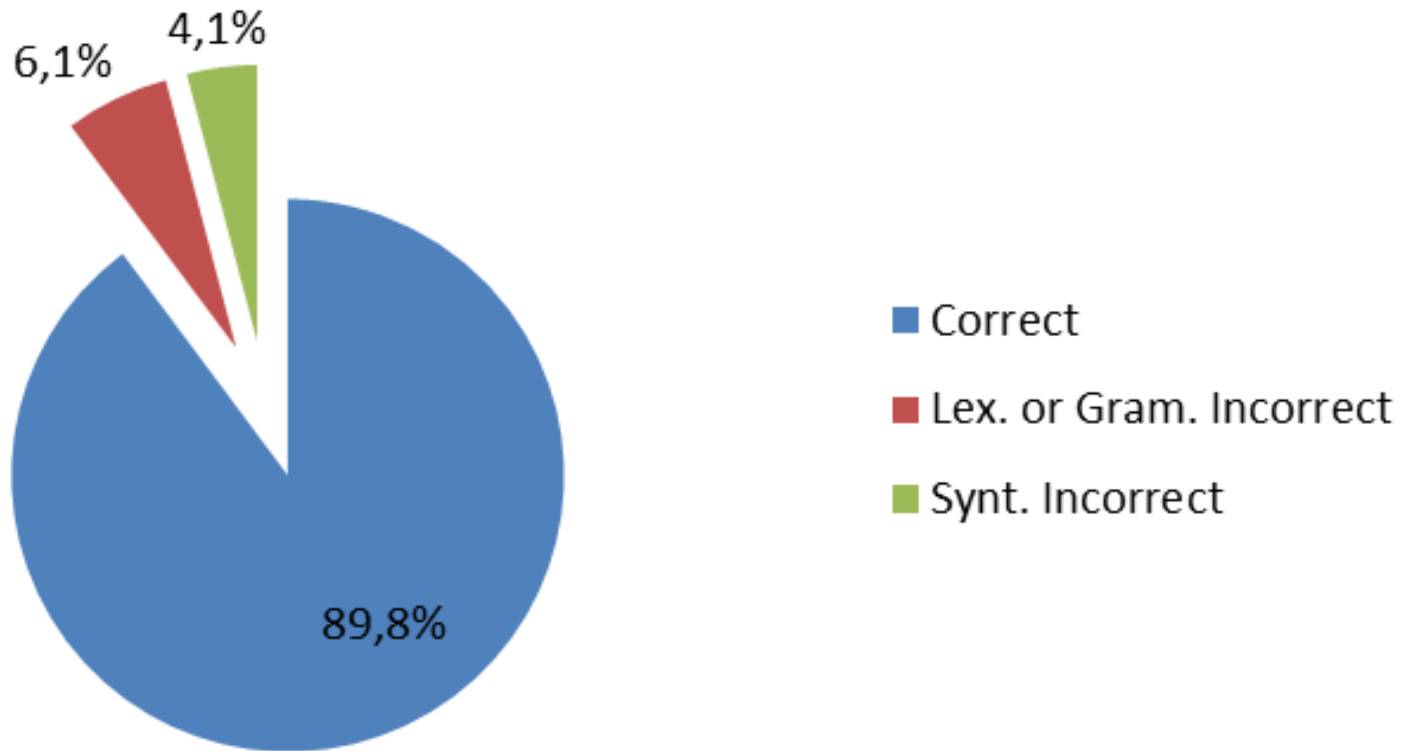
Results (2)
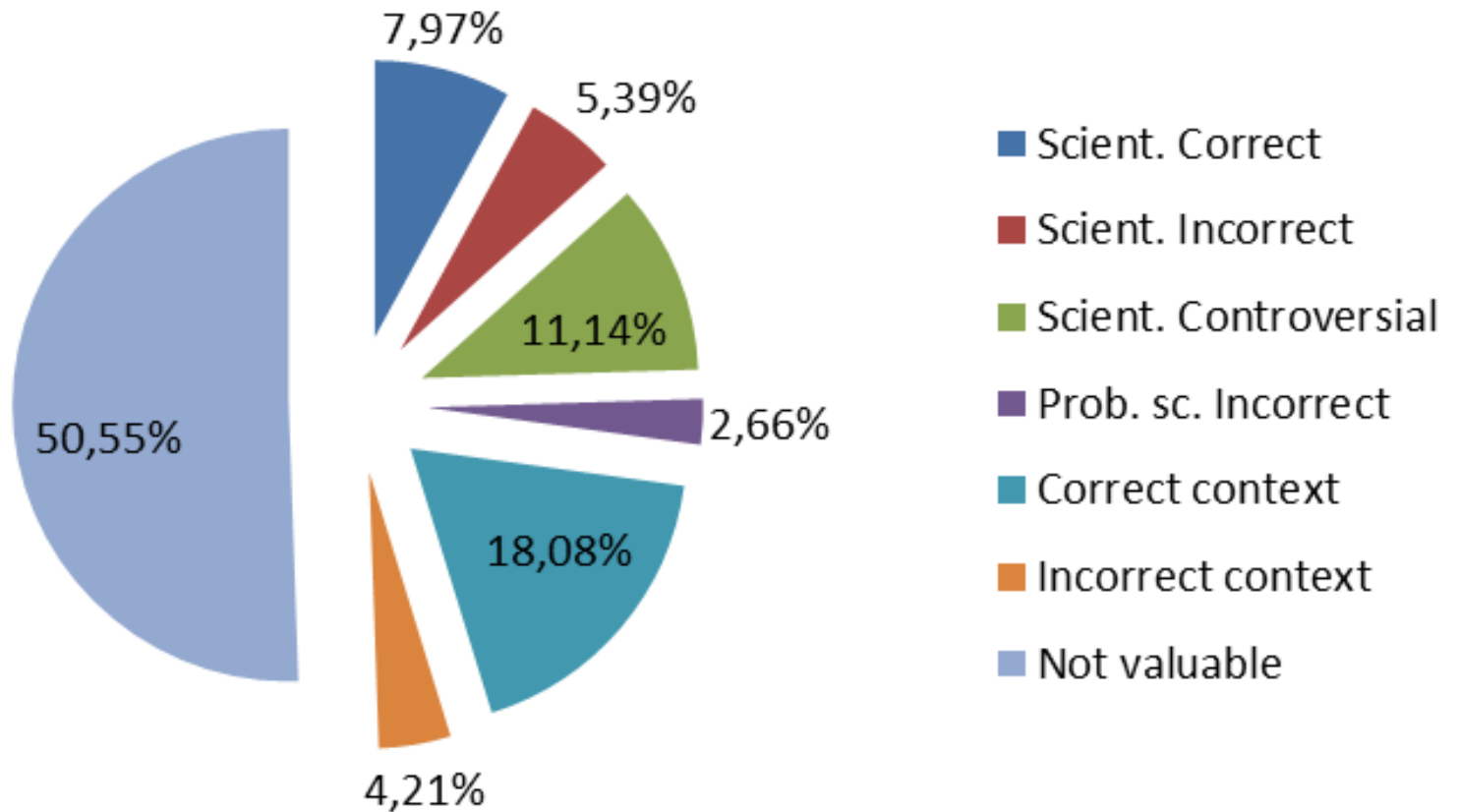
Results (3)

Discussion

Id. Doc	Title	Informative density (%)	True content (%)	Formal errors (No.)	Unfounded (%)	Scientifically Controversial (%)	Probably Scientifically Incorrect (%)	Scientifically Correct (%)	Scientifically Incorrect (%)	Score per doc
20111226	Tumore e rimedi natura	56,1	100	0	81,8	0	0	100	0	8
20140220	Piante e cancro: come	64,2	73,5	2	89,1	21,1	0	73,7	0	7
20140221	Spiritualità nella cura d	73,5	86	6	55,3	27,8	0	72,2	0	7
201312032	Bufala! La guanabana c	83,6	66,7	1	21,3	70,8	0	29,2	0	7
20140308	Dieta vegana, dieta alc	61	63,9	0	76,9	39,3	3,6	53,6	3,6	6
201303052	Funghi: alleati naturali	33,3	100	0	55,6	0	0	0	0	6
201312031	10 miti persistenti sul c	66,7	81	7	48,1	37	0	55,6	3,7	6
201303051	Cure alternative: "Il car	3,6	75	1	100	0	0	0	0	5
20131231	Il cancro si cura con l'es	46,4	53,8	0	31,6	44,4	0	33,3	11,1	4
20130129	Il cancro si cura con il b	42,1	31,3	1	96	62,5	0	31,3	6,3	3
20140126	Scappa di casa per evit	22,2	50	8	91,2	0	0	0	100	2
20140515	I veri metodi di cura cor	60,4	51,2	8	99,1	7,4	11,1	22,2	55,6	2
20131002	Ecco i prodotti, chi mi a	35	11,3	25	97,7	10,3	2,6	2,6	82,1	1
20140320	L'inganno melatonina	66,7	21,9	2	89,7	52,4	4,8	9,5	14,3	1
20130211	Cancro al seno rimedi n	35,7	15	37	100	68,4	10,5	15,8	5,3	0
20140520	Il Ganoderma Lucidum	51,5	16,3	40	98,3	46,3	25,9	7,4	14,8	0

Discussion (2)

- Some defined indicators are more relevant and more effective than others to rank websites based on their information quality.

- The most important indicator to establish information quality results to be the percentage of scientifically correct items (ScC) on the total of scientifically examined items.

$$ScS = \frac{Scientifically\ correct\ items}{Total\ of\ sc.examined\ items}$$

Discussion (3)

- Websites sorted by ScC (Scientifically Correct) indicator are ranked in a way that reflects expectations of information quality, but not completely.

- In fact, this ranking assigns a good position to documents that have also a relevant percentage of **Sc**ientifically *Incorrect* (**ScI**) content.

- **ScI** is the percentage of scientifically incorrect items on the total of scientifically examined items.

Discussion (4)

- Ideal ranking indicator should include values coming from both **Sc**ientifically **C**orrect (**ScC**) and **Sc**ientifically **I**ncorrect (**ScI**) contents.

- Best document = highest **ScC** and lowest **ScI**

- Documents should be ranked by indicator **SQ** (**S**cientific **Q**uality), calculated as follows

$$SQ = {ScS}/{100} \times (1 - {ScI}/{100})$$

SQ Ranking

Id. Doc	Title	Informative density (%)	True content (%)	Formal errors (No.)	Unfounded (%)	Scientifically Controversial (%)	Probably Scientifically Incorrect (%)	Scientifically Correct (%)	Scientifically Incorrect (%)	Doc Score	SQ
20111226	Tumore e rimedi natural	56,1	100	0	81,8	0	0	100	0	8	100,00%
20140220	Piante e cancro: come n	64,2	73,5	2	89,1	21,1	0	73,7	0	7	73,70%
20140221	Spiritualità nella cura de	73,5	86	6	55,3	27,8	0	72,2	0	7	72,20%
201312031	10 miti persistenti sul ca	66,7	81	7	48,1	37	0	55,6	3,7	6	53,54%
20140308	Dieta vegana, dieta alca	61	63,9	0	76,9	39,3	3,6	53,6	3,6	6	51,67%
20131231	Il cancro si cura con l'es	46,4	53,8	0	31,6	44,4	0	33,3	11,1	4	29,60%
20130129	Il cancro si cura con il bi	42,1	31,3	1	96	62,5	0	31,3	6,3	3	29,33%
201312032	Bufala! La guanabana cu	83,6	66,7	1	21,3	70,8	0	29,2	0	7	29,20%
20130211	Cancro al seno rimedi na	35,7	15	37	100	68,4	10,5	15,8	5,3	0	14,96%
20140515	I veri metodi di cura cor	60,4	51,2	8	99,1	7,4	11,1	22,2	55,6	2	9,86%
20140320	L'inganno melatonina	66,7	21,9	2	89,7	52,4	4,8	9,5	14,3	1	8,14%
20140520	Il Ganoderma Lucidum r	51,5	16,3	40	98,3	46,3	25,9	7,4	14,8	0	6,30%
20131002	Ecco i prodotti, chi mi ai	35	11,3	25	97,7	10,3	2,6	2,6	82,1	1	0,47%
201303052	Funghi: alleati naturali r	33,3	100	0	55,6	0	0	0	0	6	0,00%
201303051	Cure alternative: "Il can	3,6	75	1	100	0	0	0	0	5	0,00%
20140126	Scappa di casa per evita	22,2	50	8	91,2	0	0	0	100	2	0,00%

Further research

- This work should be continued along different research lines;

- 1) to analyze of a larger number of sample websites;

- 2) to investigate possible relationship of SQ and other online information evaluation methods;

- 3) to analyze all the items of the selected websites in order to identify possible logical fallacies and to explore relationships among fallacies and other measures (for example, IDE or TC) from a logical, not linguistic, point of view

Thank you
for your attention!

Mauro Mazzocut mmazzocut@cro.it

Ivana Truccolo itruccolo@cro.it

Carlo Bianchini carlo.bianchini@unipv.it