Applying Textometirc Analysis to a Description of Cochrane Medical Abstracts and their Plain Language versions: Quantitative Characterisation of Plain Language in Medical Discourse
Introduction
Cochrane Systematic reviews (SR)

- Constant increase in published research (22,000 biomedical journals)
- Synthesise existing evidence → making decisions about individual patients, policy-making, research funding
- Cochrane SRs are high-level overviews of primary research on various questions
  - Electronic cigarettes for smoking cessation
  - House dust mite control measures for asthma
  - Cognitive stimulation to improve cognitive functioning in people with dementia
  - Comparison of different regimens of proton pump inhibitors for acute peptic ulcer bleeding
- Identify, select, synthesise and appraise all high quality research evidence
Objectives

- Linguistic comparison (two genres)
- Cochrane approached us in 2016: PLS linguistic guidelines
  - need to improve certain aspects of SRs: readability in ‘conclusions’ sections.
  - need for an accessible and unambiguous set of ‘approved phrases’.
- Define ‘Plain language’ in linguistic terms
- Provide guidelines starting from a corpus-based description
Cochrane SR structure
ABS / PLS (example 1)

ABS – Objectives

To examine the effects of direct composite resin fillings versus amalgam fillings for permanent posterior teeth, primarily on restoration failure.

PLS – Review question

This review, carried out by the Cochrane Oral Health Group, addressed the question of how effective tooth-colored (composite resin) fillings are compared with conventional amalgam fillings when placed directly into cavities in permanent teeth in the back of the mouth.
**ABS – Conclusions**

While antibiotic treatment is effective in reducing the risk of pyelonephritis in pregnancy, the estimate of the effect is very uncertain because of the very low quality of the evidence.

The reduction in low birthweight and preterm birth with antibiotic treatment is consistent with theories (…), but this association should be interpreted with caution given the very poor quality of the included studies.

**PLS – What does this mean?**

Antibiotic treatment can reduce the risk of kidney infections in pregnant women who have a urine infection but show no symptoms of infection.

Antibiotics may also reduce the chance a baby will be born too early or have a low birthweight. However, the low quality of the evidence makes it hard to know for certain what the effect of treatment will be.
Hypotheses

- Cochrane corpus analysis may reveal a number of features of simplified medical language that SR editors are subconsciously developing.

- Textometric analysis of Cochrane data can reveal regular linguistic features that are possible candidate structures to be included in a more formal set of guidelines.
Methods
Cochrane Data (4540 texts: ABS / PLS)

**ABS sub-corpus**
- 2173 000 OCC.

**PLS sub-corpus**
- 1105 000 OCC.

**ABS**
- Title
- Background
- Objectives
- Search Strategy
- Selection Criteria
- Data Collection
- Conclusions

**PLS**
- Title
- Body
Correspondence Analysis (CA) of contingency tables

Positions of corpus parts in the principal plane of a CA after POS-tagging
Characteristic elements (Lebart et al., 1998)

For each cell $K_{ij}$, we assess the sub-frequency using a probabilistic model.

If $P_{inf}K_{ij}$ is smaller that the fixed threshold, the unit $i$ is a **negative characteristic element**.

If $P_{sup}K_{ij}$ is less that the fixed threshold, the unit $i$ is a **positive characteristic element**.

Lexical table (matrix)

- $K$, size of corpus
- $F_i$, frequency of unit $i$ in entire corpus
- $K_{ij}$, sub-frequency of unit $i$ in part $j$
- $t_j$, size of part $j$
POS underrepresented in PLS

Positive specificity

Negative specificity
## Characteristic elements applied to Repeated Segments (Part Of Speech)

<table>
<thead>
<tr>
<th>Repeated Segment (POS)</th>
<th>Freq ABS</th>
<th>Specificity index</th>
<th>Freq PLS</th>
<th>Specificity index</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN IN DT NN</td>
<td>12 063</td>
<td>-**</td>
<td>8 916</td>
<td>**</td>
</tr>
<tr>
<td>NN IN DT NNS</td>
<td>2 597</td>
<td>-**</td>
<td>2 136</td>
<td>**</td>
</tr>
<tr>
<td>IN DT NN IN DT</td>
<td>1 804</td>
<td>-**</td>
<td>1 665</td>
<td>**</td>
</tr>
<tr>
<td>IN DT NN CC</td>
<td>1 649</td>
<td>-**</td>
<td>1 535</td>
<td>**</td>
</tr>
<tr>
<td>MD VB VVN IN</td>
<td>1 389</td>
<td>-**</td>
<td>1 358</td>
<td>**</td>
</tr>
<tr>
<td>VBZ DT JJ NN</td>
<td>1 315</td>
<td>-**</td>
<td>1 220</td>
<td>**</td>
</tr>
<tr>
<td>NN DT NN IN</td>
<td>1 053</td>
<td>-**</td>
<td>1 058</td>
<td>**</td>
</tr>
<tr>
<td>IN DT NN IN DT NN</td>
<td>1 001</td>
<td>-**</td>
<td>1 001</td>
<td>**</td>
</tr>
</tbody>
</table>

*Prep + Det + Noun sing. or mass + Prep + Det + Noun sing. or mass*
<table>
<thead>
<tr>
<th>No</th>
<th>Filename</th>
<th>Solution 1 to 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CD000012</td>
<td>thus it is not possible to draw conclusions about the independent effects of the design of the birth</td>
</tr>
<tr>
<td>2</td>
<td>CD000013</td>
<td>slowing of the baby’s heart rate during labour. Infusing fluid into the uterus through a catheter</td>
</tr>
<tr>
<td>3</td>
<td>CD000029</td>
<td>did reduce stroke. Most strokes are caused by a sudden blockage of an artery in the brain (ca)</td>
</tr>
<tr>
<td>4</td>
<td>CD000030</td>
<td>treated with antipsychotic medication. This review highlights the limited evidence available for the use of this treatment in cap</td>
</tr>
<tr>
<td>5</td>
<td>CD000032</td>
<td>the following four findings. (1) Providing nutritional advice resulted in an increase in the mother’s pregnancy</td>
</tr>
<tr>
<td>6</td>
<td>CD000032</td>
<td>following four main findings. (1) Providing nutritional advice resulted in an increase in the mother’s pregnancy</td>
</tr>
<tr>
<td>7</td>
<td>CD000032</td>
<td>involving 1051 women showed no benefit for women and potential harm for the baby through an increase in the number of births</td>
</tr>
<tr>
<td>8</td>
<td>CD000032</td>
<td>) showed no benefit for women and potential harm for the baby through an increase in the number of births</td>
</tr>
<tr>
<td>9</td>
<td>CD000035_EN</td>
<td>The studies seemed to show a benefit in terms of the acidity of the baby’s blood</td>
</tr>
<tr>
<td>10</td>
<td>CD000038</td>
<td>machine. A CTG assesses the pattern of the baby’s heartbeats alongside the size of the mother’s condition</td>
</tr>
<tr>
<td>11</td>
<td>CD000039</td>
<td>stroke (a sudden brain attack either due to blockage or rupture of an artery in the brain)</td>
</tr>
<tr>
<td>12</td>
<td>CD000052</td>
<td>the lungs. The medication can be given by wet nebulisation or from an inhaler with a spacer device</td>
</tr>
<tr>
<td>13</td>
<td>CD000054_EN</td>
<td>all the trials assessed 'microbiological cure' (that is they looked for an eradication of the infection) and...</td>
</tr>
<tr>
<td>14</td>
<td>CD000064</td>
<td>blockage of an artery to a part of the brain causes swelling of that part of the brain. The...</td>
</tr>
<tr>
<td>15</td>
<td>CD000066</td>
<td>trials (involving 800 women) in this review. We found that the application of a skin preparation...</td>
</tr>
<tr>
<td>16</td>
<td>CD000066</td>
<td>during pregnancy. Only three trials (involving 461 women) looked at the severity of the stretch mark...</td>
</tr>
<tr>
<td>ID</td>
<td>Sentence</td>
<td>Context</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>6578</td>
<td>trial with 40 patients. There were no mortality or surgical complications</td>
<td>in either group in this trial</td>
</tr>
<tr>
<td>9933</td>
<td>including 212 participants reported deaths after surgery. There were no deaths</td>
<td>in either group in this trial</td>
</tr>
<tr>
<td>9943</td>
<td>be determined by chance so that the two groups are truly similar</td>
<td>in every respect except the treatment</td>
</tr>
<tr>
<td>9924</td>
<td>the use of broad searches. At least two people were involved</td>
<td>in every stage of the review</td>
</tr>
<tr>
<td>4925</td>
<td>more common in stented patients however the addition of antibiotic prophylaxis resulted</td>
<td>in no difference in the incidence</td>
</tr>
<tr>
<td>0015</td>
<td>adversely affects tasks such as reading, driving and face recognition.</td>
<td>In the absence of a cure</td>
</tr>
<tr>
<td>8816</td>
<td>timing regimens, but were unable to find any such trials.</td>
<td>In the absence of this evidence</td>
</tr>
<tr>
<td>7274</td>
<td>focus on the safety of the sedative techniques and involve anaesthesia personnel</td>
<td>in the administration of the sedation</td>
</tr>
<tr>
<td>87_EN</td>
<td>compared the effects of blocking the brachial plexus by injecting local anaesthetic</td>
<td>in the area below the collarbone</td>
</tr>
<tr>
<td>45_EN</td>
<td>of men aged between 65 and 79 years have an abdominal aneurysm</td>
<td>in the area of the aorta</td>
</tr>
<tr>
<td>6619</td>
<td>noisy breathing. These symptoms are the result of swelling that occurs</td>
<td>in the area of the windpipe</td>
</tr>
<tr>
<td>6927</td>
<td>(AMD) is a progressive disease of the macula (area</td>
<td>in the back of the eye</td>
</tr>
<tr>
<td>7920</td>
<td>new abnormal blood vessels. However, poor view of the retina</td>
<td>in the back of the eye</td>
</tr>
<tr>
<td>6126</td>
<td>vitreoretinopathy (PVR). Background The retina is the light-sensing tissue</td>
<td>in the back of the eye</td>
</tr>
<tr>
<td>4020</td>
<td>information about side effects and complications. Side effects may include pain</td>
<td>in the back of the knee</td>
</tr>
<tr>
<td>5620</td>
<td>with conventional amalgam fillings when placed directly into cavities in permanent teeth</td>
<td>in the back of the mouth</td>
</tr>
<tr>
<td>1093</td>
<td>methods. Meningococcal meningitis is a brain infection caused by bacteria living</td>
<td>in the back of the throat</td>
</tr>
<tr>
<td>8653</td>
<td>and what is febuxostat? Uric acid is a product normally present</td>
<td>in the blood as a result</td>
</tr>
<tr>
<td>44_EN</td>
<td>effects include high blood pressure, infection and an excess of glucose</td>
<td>in the blood of these preterm</td>
</tr>
</tbody>
</table>
might use special stitches to repair it.

This is called a 'repair'.

Some procedures can be performed arthroscopically (surgical instruments are inserted through a small incision or key hole and an endoscope to visualise the area and to guide the doctor is inserted through another incision), which can mean a shorter recovery time.

Review question This review, carried out by the Cochrane Oral Health Group, addressed the question of how effective tooth-colored (composite resin) fillings are compared with conventional amalgam fillings when placed directly into cavities in permanent teeth in the back of the mouth.

Background There is controversy over the best materials to use when restoring or filling holes caused by tooth decay in permanent teeth at the back of the mouth.

Amalgam fillings have been successfully used for over 150 years and are cost effective.

However, their use has declined over recent years partly because of the way they look and because of the perceived risk of mercury that is used in them.

Tooth-colored (composite resin) fillings are frequently used in the front teeth and also in permanent teeth at the back of the mouth.
Results
Salient features of PLS

- Explicitation of information-rich medical LSP
  - Lexicalized presentation of numerical comparisons
  - Unpacking of compact LSP
  - Defining terms & concepts

- Focus on user experience
  - Localisation of processes in the human body
  - Topicalisation of human participants
  - Generalisation & interpretation of results
  - Oral discourse features
Explicitation of numerical results (1/2)

- at least 50% pain relief → good levels of pain relief
- congestive heart failure: RR 5.11 → risk of heart toxicity is five times more likely
- overall survival: HR 0.66 → mortality is reduced by one-third
- caffeine used at doses of 100 mg or more → a dose equivalent to a mug of coffee
Explicitation of numerical results (2/2)

Abstract – Main results

(...) showed a significantly greater reduction in pain favouring topical capsaicin over placebo at one and two weeks (MD -23.80, 95% CI -44.81 to -2.79; NNT 3, 95% CI 2 to 47; MD -34.40, 95% CI -54.66 to -14.14; NNT 2, 95% CI 1.4 to 6 respectively).

Plain Language Summary

(...) patients also had better pain relief with capsaicin cream than for those given a placebo cream. On average, patients receiving the active treatment improved by 34 more points (out of 100) than the control group.
Unpacking of compact LSP (1/2)

- injection site reactions
  - skin reactions at the site of the injection

- abdominal aortic aneurysm (AAA)
  - an abdominal aneurysm in the area of the aorta

- near the bladder neck and proximal urethra
  - around the junction between the bladder and the urethra

- the external urethral sphincter muscle (EUS)
  - the muscle at the exit of the bladder that allows people to control their bladder emptying
Unpacking of compact LSP (2/2)

- Patients recovering from a stroke, patients recovering from surgery, elderly patients with a mix of conditions

- Duration of surgery

- Number needed to treat (NNT)

- Higher dietary levels of antioxidant vitamins

→ Patients who had a stroke, patients who had surgery, and elderly patients who had different types of conditions

→ How long each operation took to perform

→ The number of people that need to be treated

→ People who eat a diet rich in antioxidant vitamins
Multiple sclerosis (MS) is a chronic disease of the nervous system.

- Bleeding on the surface of the brain is called a subarachnoid haemorrhage.
- Pneumothorax (a condition when air escapes from the lung into the chest)
- Dilated blood vessels may appear near the surface of the skin (telangiectasia).

- Dry eye is a common disorder of the tear film
- Naltrexone is a long-acting drug (an opioid antagonist)
Localisation of processes in the human body

- In CTS there is increased pressure on a nerve (the median nerve) as it passes over the wrist towards the palm of the hand through the carpal tunnel.

- The injection can be placed in the shoulder joint or in locations around the shoulder such as the space under the acromion above the top of the humerus and the doctor can approach the patient from the front, side or back.
Abstract (ABS) – Main results

No separate safety data were available for patients with RA, however 44% of patients developed burning at the site of application and 2% withdrew because of this.

Plain Language Summary (PLS)

About 50% of patients who use capsaicin cream on their skin will develop this local burning but only 2 in 100 will stop treatment because of this.
Conclusions

- ABS: factual account of trial results
  → PLS: generalized conclusion interpreted from user’s perspective

- Regular tendencies of simplification in PLS:
  - Unpacking of compact & complex LSP (medical & statistics)
  - Reformulation of research processes and empirical findings towards more user-oriented terms
  - Topicalisation of human participants, etc.

- Next stage: adapt findings of textometric analysis into appropriate editorial guidelines on PLS for Cochrane authors
References


- McIlwain, Catherine (ed.) 2013. Standards for the reporting of Plain Language Summaries in new Cochrane Intervention Reviews. Available at: http://consumers.cochrane.org/PLEACS

