James W. Patterson. Automating the Human Factors Analysis and Classification System (HFACS): An Initial Investigation Based on Error Reports in Radiation Oncology. A Master's Paper for the M.S. in I.S. degree. April, 2020. 52 pages. Advisors: Lukasz Mazur and Yue (Ray) Wang

This study constitutes an evaluation of two datasets of error reports from the University of North Carolina School of Medicine's Department of Radiation Oncology. These errors were reported in accord with the Human Factors Analysis and Classification System (HFACS), using HFACS Level 1 and HFACS Level 2 codes.

Keywords from an initial dataset of 58 reports, a list of HFACS theory keywords, and a list of related thesaurus words were used to develop a dictionary of signal words. These words are related to the HFACS Level 1 code for "Condition of Operator" and the HFACS Level 2 code for "Inattention-Distraction." These words were evaluated for relevance to the "Inattention-Distraction" category, both at face value, and in the context of "Inattention-Distraction" reports from the initial dataset of 58 reports. The signal words were then evaluated a second time in the context of a second dataset of 3459 reports, for confirmation of contextual relevance. The findings suggest that while more data is needed for future research, findings could lead to the development of an automated and more user-friendly HFACS reporting system at UNC School of Medicine's Department of Radiation Oncology.

Headings:

Electronic health records

Human-computer interaction

Medical informatics

Medical records

Online databases

AUTOMATING THE HUMAN FACTORS ANALYSIS AND CLASSIFICATION SYSTEM (HFACS): AN INITIAL INVESTIGATION BASED ON ERROR REPORTS IN RADIATION ONCOLOGY

by
James W. Patterson

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2020

Approved by

____

Lukasz Mazur and Yue (Ray) Wang

Table Of Contents

INTRODUCTION

Errors are an inevitable aspect of human life; humans can err regardless of the industries or organizations in which they work. In high risk industries, the consequences of human error may be not only detrimental, but also fatal. It is for this reason that managers in high risk industries, like aviation, rail, nuclear power, and healthcare industries, are preoccupied with reducing human error. There are different views on the primary contributing factors to errors, but it is generally agreed that while individuals are responsible, the circumstances that unfold around people ultimately cause the assessments and actions that lead to errors. On an external level, issues such as legislation, regulation, and market activity can influence organizational policies and procedures that contribute to errors. On a more internal level, factors such as reasoning, judgment, ambiguity, and time management can contribute to errors.

The healthcare industry is one in which the alleviation of micro-level errors is a frequent focus of managers. Errors in healthcare are often described as "events," and "adverse" or "non-adverse" depending on whether or not patient safety is compromised. Human error can be evaluated both in terms of clinicians, and in terms of tools and reporting systems. From a clinician perspective, findings suggest that some employees are less likely to report incidents than others, and that the culture of reporting tends to be "in-house" and less reliant on directives. Barriers such as time, unsatisfactory processes, knowledge deficiency, cultural norms, inadequate feedback, and beliefs about risk influence which incidents are or are not reported.

Some tools are designed to track incidents in hospitals and identify their causes. HFACS (The Human Factors Analysis and Classification System) is one such tool, and it is the primary tool used by University of North Carolina (UNC) Health and the UNC School of Medicine's Department of Radiation Oncology. HFACS is used for analysis in both adverse and non-adverse events. The HFACS system consists of two hierarchical sets of codes that are used to identify the root cause of errors.

The research aims for this paper are:

1. To develop a methodology to extract signal words within error reports from Radiation Oncology.
2. To develop a dictionary of signal words for human error categories that correlate with HFACS codes.
3. To evaluate whether the corresponding dictionary of words can be used to determine if text documents belong to a given category or not.

The long-term aims of this paper are:

1. To implement a more automated HFACS system in Radiation Oncology, that will a) be easier for clinicians to use, and b) more uniformly and comprehensively describe errors that occur.
2. That this new system will lead to fewer errors and an increase in positive patient safety outcomes.

Literature Review

A literature review was used to organize relevant past research for this paper, and this also serves as a background for the primary research of this paper. The literature review begins with a general overview of human error, as it has been addressed in a variety of industries and organizational settings.

Human error in a variety of industries and organizational settings

In addition to healthcare, human error has been examined within the process control, nuclear power, chemical and petrochemical industries (Kirwan, 2017). In these industries, emphasis for error identification is typically placed on risk assessment, task analysis, error identification, quantification, and representation of errors in analysis (Kirwan, 2017). This is supplemented by error reduction analysis, quality assurance, and documentation (Kirwan, 2017).

Some researchers consider human error to be the starting point of investigation rather than the conclusion (Dekker, 2017). They argue that the circumstances that unfold around people cause their assessments and actions to change accordingly, and that reverse engineering human error can clarify this (Dekker, 2017). In this vein, Wallace et al. argue for the creation of reliable taxonomies in these industries (Wallace et al., 2006).

Despite the call for creation of a uniform and comprehensive taxonomy for reporting errors, the literature suggests that many industries have not approached the issue of human error through the use of such a taxonomy. HFACS is an example of a taxonomy that considers human error at the bottom level. It has been applied across a number of industries, and serves as a focal point of this paper.

The Human Factors Analysis and Classification (HFACS) system

HFACS consists of two sets of codes, as shown in Figure 1 below. The first set of codes provide a general description of the conditions under which errors occur (in red as shown below) - this includes information about the root causes of error. The second and third set of codes provides a more nuanced description of contributing factors to the error in question (in tan and white as shown below).

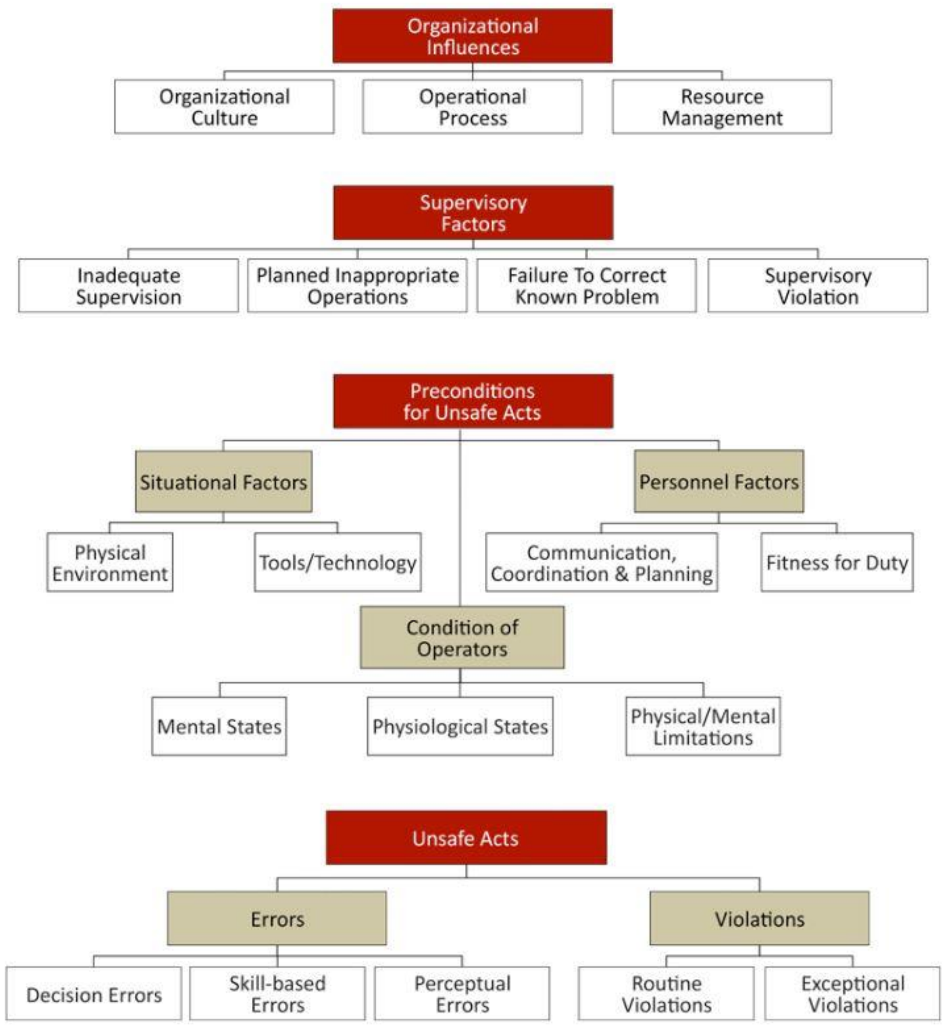


Figure 1 - HFACS Level 1 and Level 2 Codes

HFACS was originally developed to study the root cause of errors in the aviation industry, but its methodology can be used to help any organization perform root cause analysis - to determine "why incidents occur" as opposed to "who is responsible" (Diller et al., 2013). This can highlight key relationships between errors at the operational level, and organizational inadequacies at adjacent and higher levels (Diller et al., 2013). This might suggest that active failures are promoted by latent conditions in organizations. (Li et al., 2006).

In terms of non-healthcare industries that have used HFACS, the Australian Defence Force (ADF) has used HFACS to classify factors that contribute to flight incidents (Olsen et al., 2010). HFACS has been used to perform root cause analysis on helicopter maintenance errors (Rashid et al., 2010). In the mining industry, HFACS data has revealed that skill-based errors are the most common types of unsafe acts (Patterson et al., 2010). In civil and military aviation, HFACS data has been used to identify a framework around which new investigative methods can be designed and existing accident databases can be restructured (Shappell et al., 2000). In the shipping industry, HFACS has been modified to integrate FAHP (Fuzzy Analytical Hierarchy Process) (Celik et al., 2009). FAHP improves the HFACS framework by providing an analytical foundation and grouping decision-making abilities (Celik et al., 2009).

In the context of Radiation Oncology, Mosaly et al. have conducted research on the agreement and lack thereof between novices and experts analyzing HFACS data and codes (Mosaly et al., 2013). This suggests that modifications to the existing HFACS system can be used to not only increase this agreement, but also to enhance usability on the part of clinicians and uncover more of the behaviors and actions that contribute to errors (Mosaly et al., 2013).

The behaviors and actions that contribute to errors in healthcare settings are numerous and varied, and clinicians do not always contextualize errors that occur in terms of HFACS categories. Therefore, it is important to understand errors not only in terms of HFACS categories, but also in terms of other categorical factors.

Human error in healthcare settings

Epstein et al. find that internal basic skills, new formats that assess clinical reasoning, expert judgment, management of ambiguity, professionalism, time management, learning strategies, and teamwork promise a multidimensional assessment of factors that contribute to errors (Epstein et al., 2002). Donaldson et al. argue that external factors such as legislation, regulation, and market activity influence the quality of care and handling of mistakes (Donaldson et al., 2000). They also argue that internal improvements can be made in leadership, data collection and analysis, and system effectiveness (Donaldson et al., 2000).

Reason examines the person and system approaches to solving human fallibility. He finds that individuals in high reliability organizations recognize that human variability is a negative force to harness in averting errors (Reason, 2000). These individuals work hard to control this variability, which will case errors, and they are constantly preoccupied with the possibility of failure (Reason, 2000).

As with errors themselves, the behaviors and actions that contribute to handling and reporting of errors in healthcare settings are numerous and varied. Clinicians do not always describe errors that occur in terms of HFACS categories. Therefore, it is important to understand not only HFACS categories, but also additional factors that contribute to handling and reporting of errors.

How clinicians handle human error in health care settings

Kingston et al. find that nurses report more habitually than physicians, due to a work culture that provides directives, protocols, and the notion of security (Kingston et al., 2004). They also determine that the medical culture of physicians is less transparent, favors dealing with incidents "in house," and is less reliant on directives (Kingston et al., 2004). Evans et al. likewise determine that physicians report incidents less frequently than nurses, even though both groups believe they should report most incidents (Evans et. al, 2006).

Waring finds that physicians are concerned about managers and non-physicians engaging in regulation of medical quality through use of incident data (Waring, 2005). He alludes to the role of blame in inhibiting reporting, as well as the prevailing notion among physicians that errors are inevitable and potentially unavoidable, the prevalence of anti-bureaucratic sentiments, and the rejection of perceived excessive administrative duties (Waring, 2005).

General barriers to reporting include time, unsatisfactory processes, knowledge deficiency, cultural norms, inadequate feedback, beliefs about risk, and perception of the process (Kingston et al., 2004). To improve reporting, Evans et al. suggest that clarification should be established of what incidents should be reported (Evans et. al, 2006). Additionally, Evans et al. suggest that the process of reporting needs to be simplified, and that feedback should be given to reporters (Evans et. al, 2006).

Hutchinson et al. find a link between reporting rates, hospital characteristics, and other safety and quality datasets (Hutchinson et al., 2009). They determine that incident reporting rates from acute hospitals increase with time, from connection to the national reporting system (Hutchinson et al., 2009). They also find that higher reporting rates are associated with a more positive safety culture (Hutchinson et al., 2009).

One long-term aim of this study is with regard to the development of a more automated HFACS system, which will allow clinicians to more easily and comprehensively report errors and their contributing factors in Radiation Oncology. In tandem with the goal of increasing the detail and uniformity of reports with an automated HFACS system, there is an interest in understanding the contents of reports themselves. Reports should ideally be easily associable with relevant HFACS codes, and an approach is needed to evaluate the strength of association. Such an approach would naturally need to consider the language that clinicians use in their reports, and would need to consider keywords in particular. Text categorization is one such approach.

Feature selection in text categorization

One primary approach in our study involves extracting category-specific words, noting the number of times that a word occurs within documents of a certain HFACS category can be calculated. Word count is a heuristic method that can be used to evaluate the strength of association between a given word and a document category, and this approach is related to a broader methodology for selecting category-specific words. This methodology is known as feature selection for text categorization. In their study on this topic, Yang and Pedersen evaluate five methods of feature selection. These are known as document frequency (DF), information gain (IG), mutual information (MI), χ^2 (the chi-square test - CHI), and term strength (TS) (Yang et al., 1997).

DF is used to measure the number of documents in which a term occurs, while IG is used to number of bits of information obtained for category prediction by knowing the presence or absence of a term in a document. MI considers the two-way contingency of a term and a category, which involves the number of documents that contain both the term and the category, the number of documents that only contain the term but not the category, the number of documents that contain only the category but not the term, and the number of documents that contain neither. CHI measures the lack of independence between a term and a category, and can be compared to the χ^2 distribution with one degree of freedom to judge extremeness (Yang et al., 1997).

Yang and Pedersen find IG and CHI to be the most effective approaches in their experiments. Using IG thresholding, the authors find that removal of up to 98% of unique terms can actually yield an improved classification accuracy. They also find that DF thresholding, the simplest method with lowest cost in computation, can be used instead of IG or CHI when computation of measures is too expensive. IG and CHI are most effective in aggressive term removal (98% of unique terms), without producing a loss in categorization accuracy. (Yang et al., 1997).

Forman further examines the effect of multiple feature selection methods on a benchmark of 229 text classification problem instances gathered from Reuters, the Text Retrieval Conference (TREC), a data subset of MEDLINE from Oregon Health Sciences University (OHSUMED), and other resources. The results are measured according to accuracy, f-measure, precision, and recall for evaluation. A feature selection metric known as Bi-Normal Separation (BNS) outperformed the others by a substantial margin using these metrics of evaluation, except for precision, for which IG yielded the best result most often. Forman also determines that IG and CHI work poorly together due to correlated failures. BNS is consistently a member of optimal pairs of metrics for each of the four performance goals (Forman 2003).

While this study will not incorporate text classification methodology, the approaches outlined in the research above do provide a direction for future research, and for a methodology that advances beyond this study's use word count and margin calculations. Additionally, while text categorization provides an approach for measuring the association between words and categories, this approach does not account for issues related to word association beyond face value. In certain cases, two words may not appear to be related to each other at face value, or to a particular category. Yet, in such cases, an association between such a pair of words may actually exist. Semantic relatedness is an approach that is designed to address this issue.

Semantic relatedness as measured by topological similarity

In their research, Siblini and Kosseim measure the semantic relatedness between words as a means of developing related metrics for natural language processing applications. They explore the use of a lexicon (a vocabulary/dictionary) in conjunction with a number of semantic relation types. These types include weights and word definitions, in order to calculate semantic similarity between words using a new vector modeling approach. The weights use all 26 relations available in WordNet, in addition to information found in glosses. The information mined from WordNet led to the creation of a semantic network of 265,269 concepts, which are connected through a total of

1,919,329 relations. Each edge was assigned a weight according to a different category, and the relations and glosses were categorized into seven categories (Siblini and Kosseim 2013).

The authors calculated semantic relatedness between two words as the lowest cost path between the two words in a semantic network. Because the semantic network is directed, the maximum weight among both directions that link the two words is taken. Siblini and Kosseim argue that future work includes performing additional experiments to find the best values for certain parameters and class weights, which, as of the article's publication, have been set empirically over several small experiments. A more formal training is also needed to find the best combination of these parameters, and semantic information needs to be categorized into more than one category. Computing lexicon based semantic similarity must be similarly be addressed (Siblini and Kosseim 2013).

METHODOLOGY

The methodology for this paper constitutes the approaches used to address the three primary research aims. The methodology consists of two phases, the first of which is outlined below in Figure 2.

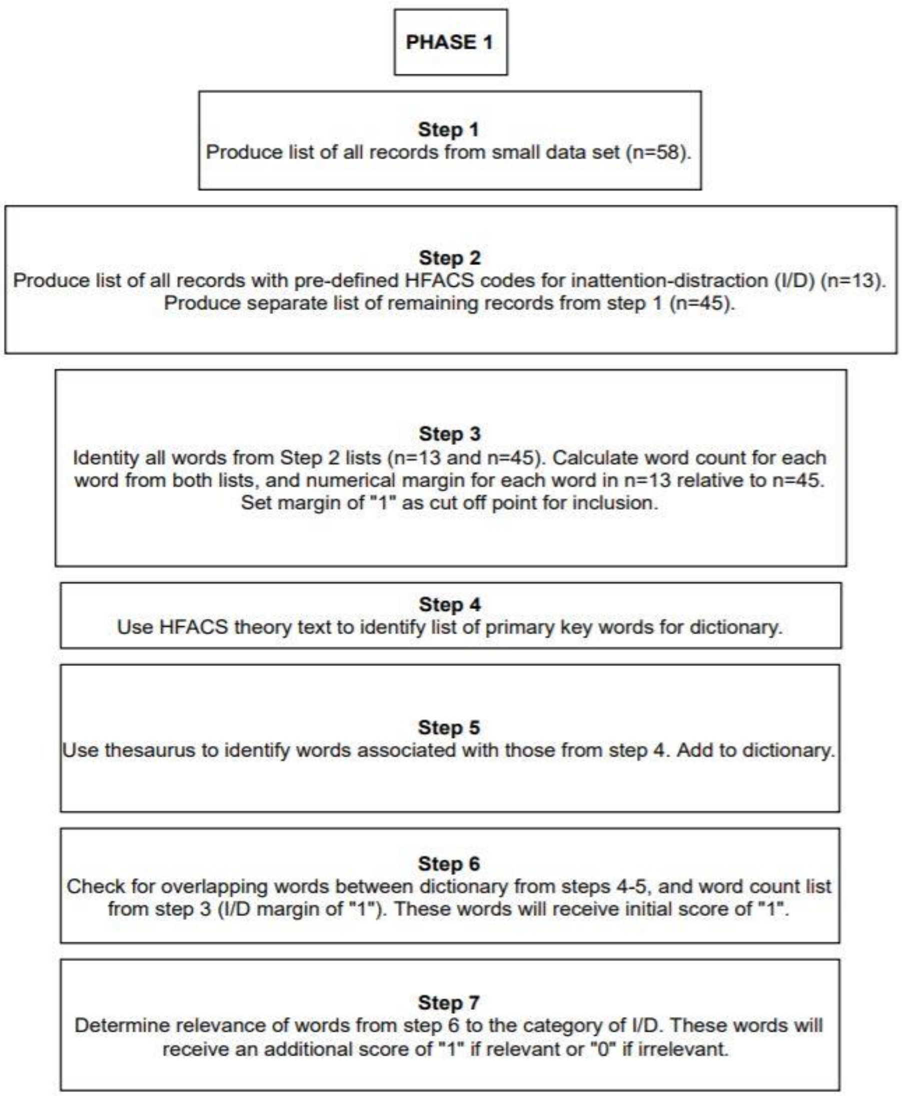


Figure 2 - Phase 1 Methodology

This project began when two Excel spreadsheets were distributed. The first contained a dataset of 58 records/reports from UNC Radiation Oncology. These reports were pre-selected because they contain information pertaining to Stereotactic body radiation therapy (SBRT), which is one of the most common procedures in Radiation Oncology. The second spreadsheet contains the corresponding report numbers (known as "GC" or "good catch" numbers), in addition to each report's associated HFACS Level 1 and HFACS Level 2 codes. Of particular interest is the Level 1 code "Condition of Operator," for reasons that will be explained below.

Step 1: Produce list of all records from small data set (n=58).

The first step was to copy and paste all 58 reports from the first spreadsheet into a new spreadsheet. All non-alphanumeric characters were removed, and comma separation values were added where spaces now appeared. This step was completed so that the spreadsheet could be fed into a program that would separate each word for a word count See Appendix 1.

Step 2: Produce list of all records with pre-defined HFACS codes for inattention-distraction (I/D) (n=13). Produce separate list of remaining records from step 1 (n=45).

After identifying the corresponding Level 1 and Level 2 codes for each report, it was determined that the overwhelming majority of reports correspond to the Level 1 code for "Condition of the Operator." Additionally, out of the 58 reports, most (13) correspond to the Level 2 code of "Inattention-Distraction." Because "Inattention-Distraction" is the most common Level 2 code in the reports, there is particular interest in keywords related to this category. Next, the 58 reports were separated into two new separate spreadsheets.

The first spreadsheet contains all 13 reports with a Level 2 code for "Inattention-Distraction" [see Appendix 2], and the second spreadsheet contains all of the remaining 45 reports [see Appendix 3]. By determining which words from these reports are "Inattention-Distraction"-related, and which are not, the hope was to develop a dictionary of words that could be used to predict whether or not an error is due to inattention and/or distraction.

Step 3: Identity all words from Step 2 lists (n=13 and n=45). Calculate word count for each word from both lists, and numerical margin for each word in n=13 relative to n=45. Set margin of "1" as cut off point for inclusion.

After completing a LinkedIn learning lesson, a Python program was subsequently developed to evaluate the two spreadsheets shown in Appendixes 2 and 3 [see Appendix 4]. In order to do so, the 'pandas' dataframe package was used to read both files as visual tables. The resulting .csv file exported from Python was then converted to .xlsx, and the "Inattention-Distraction" word count margin was calculated relative to the non-"Inattention-Distraction" word count. It was collectively determined that a margin of "1" would be used to determine which words are more related to "Inattention-Distraction" relative to the other HFACS codes [see Appendix 5].

Step 4: Use HFACS theory text to identify list of primary key words for dictionary.

There was some deliberation regarding which sources would be used to develop the primary list of words for the dictionary, because the methodology for choosing such a dictionary of words needed to simultaneously be based on HFACS theory, and to be as comprehensive as possible. Wiegmann and Shappell's original HFACS theory text was ultimately chosen for the development of the primary list of words. The belief is that basing the primary list on the original theoretical text produces an initial list of words most closely related to HFACS in Radiation Oncology. Despite the fact that HFACS was originally developed as a means of analyzing flight errors, it is now used in a wide variety of industries, as the literature review demonstrates.

In choosing which words to include, noun words related to the HFACS Level 1 code of "Condition of the Operator" and its related HFACS Level 2 codes - "Adverse Mental States," "Adverse Physiological States," and "Physical/Mental Limitations" - were included. This was chosen because many of the words listed in HFACS theory are either adjectives, or are specific to flight errors and therefore irrelevant in a Radiation Oncology setting. The adjectives are not only undescriptive of "Inattention- Distraction," but their selection in step 4 also had the potential to lead to additional challenges in steps 5-7, due to the inclusion of additional thesaurus words that would also be irrelevant to "Inattention-Distraction" [see Appendix 6].

It is worth mentioning that "Inattention-Distraction" is not an official HFACS Level 2 code as part of the original HFACS model, but was created internally within UNC Radiation Oncology as a means of more accurately describing the primary cause of operator errors in this setting [see Appendix 7].

Step 5: Use thesaurus to identify words associated with those from step 4. Add to dictionary.

After the primary list of words were gathered from the HFACS theory text, a thesaurus was used to identify all synonym words. "Thesaurus.com" was randomly selected for this task. No words were intentionally excluded from this step, due to the fact that word relevance was to be determined in step 7 [see Appendix 8].

Step 6: Check for overlapping words between dictionary from steps 4-5, and word count list from step 3 (I/D margin of "1"). These words will receive an initial score of "1".

The next step of the project was to check for instances of words that overlap between Appendix 8 and Appendix 2. Each of the Appendix 8 words were searched within Appendix 2, and notation was made of a) the overlapping word; b) the good catch number; and c) the "Inattention-Distraction" margin for the word according to Appendix 5 [see Appendix 9].

Step 7: Determine relevance of words from step 6 to the category of I/D. These words will receive an additional score of "1" if relevant or "0" if irrelevant.

As mentioned earlier, during step 5 no words were excluded from the thesaurus because the intention was to focus on word relevance during step 7. Based solely on the thesaurus selection method, all of these words should be related to "Inattention-Distraction" at face value, by virtue of the fact that they are related to words from the primary list of words. However, upon closer examination, it became clear that many of these words are not related to "Inattention-Distraction" at face value.

The word "attention" is related to "Inattention-Distraction," because it is an antonym of the word "inattention." Likewise, the word "confusion" is related to "Inattention-Distraction," because "confusion" can be either a symptom or a cause of "Inattention-Distraction". However, words like "natural" and "seen" have no obvious connection to "Inattention-Distraction" at face value. As such, 12 of the words from Appendix 9 are not specific to "Inattention-Distraction" at face value.

The deemed relevance of only 2 words from Appendix 9 limited our sample size significantly for examining contextual relevance in both datasets. I gathered with my advisors to determine how we could include additional words for testing of contextual relevance. The ultimate decision was made to review Appendix 5 for words that also appear to be related to "Inattention-Distraction" at face value, and that feature a margin ≥ 1. This led to the inclusion of three additional words, which were selected for testing of contextual relevance. The words are "aware," "forgot," and "missed." Like "confusion," each of these three words can be either a symptom or a cause of "Inattention-Distraction".

After selecting the five contextually relevant words, and the twelve non-contextually-relevant words from Appendix 9 and Appendix 5, all of the records from Appendix 2 in which these words appear were analyzed. This was done in order to determine the rationale for relevancy or irrelevancy of the words in context, regardless of whether or not the words appear to be relevant/irrelevant at face value [see Appendix 10].

During a discussion with my advisors, in which the decision was made to scan Appendix 5 for additional words to test against Appendix 2, a natural language processing (NLP) experiment was mentioned, which allows a user to determine the relationships between two words in accord with a theory known as "semantic relatedness." The study provided by Siblini and Kosseim, which is covered in the literature review, is also accompanied by a webpage at Olesk that allows a user to compare any two words using the methodology that the authors establish. This tool was used to test all combinations of the 14 words from Appendix 9, and the 3 words from Appendix 5 (17 words total) [see Appendix 11].

The methodology for this paper constitutes the approaches used to address the three primary research aims. The methodology consists of two phases, the second of which is outlined below in Figure 3.

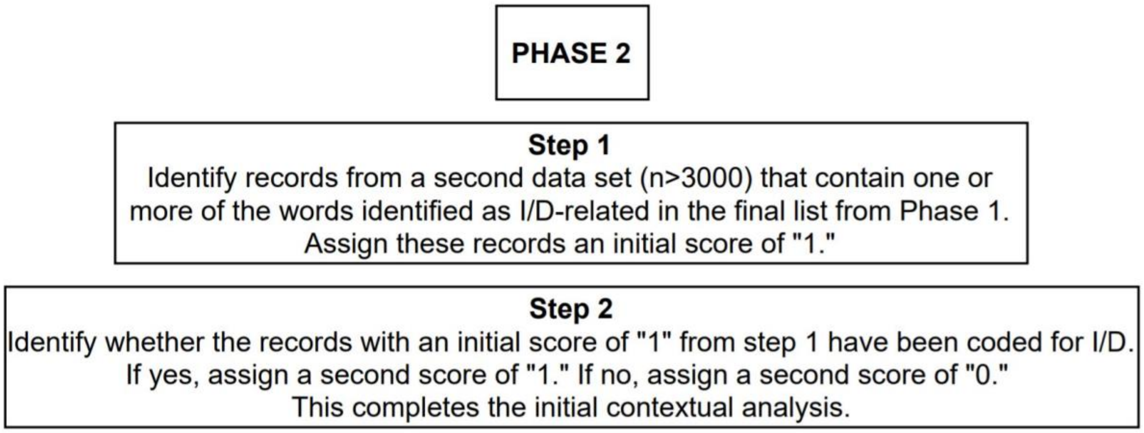


Figure 3 - Phase 2 Methodology

The next phase of this project involved the use of a second dataset of 3459 records from UNC Radiation Oncology. The purpose behind using this second dataset was to verify the contextual relevance of the 17 words selected for testing against Appendix 2 - first by determining which records are "Inattention-Distraction"-related, and then by testing the words themselves in the context of the 3459 records. These records contain information pertaining to SBRT, as is the case with the first dataset. This second set of records are also stored in a database known as ImprovementFlow, which is used by UNC Radiation Oncology to report errors.

Unlike the first dataset, these records are not accompanied by information related to HFACS Level 1 or Level 2 categories of error. This information is only available to employees of UNC Radiation Oncology. It was also not possible to export the records as an Excel spreadsheet. For this reason, each of the 17 words were manually searched in the records, and each record result was copied and pasted into an Excel spreadsheet [see Appendix 12b]. Note is made of a) each word; b) the good catch number of the associated record(s) in which the word appears; and c) the text of the record in which the word appears.

Step 1: Identify records from a second data set (n>3000) that contain one or more of the final identified I/D-related words from Phase 1. Assign these records an initial score of "1".

The first step was to identify which of the records from the second dataset contain one or more of the 5 words previously identified in Appendixes 9 and 5 as "Inattention-Distraction"-related. These corresponding records were assigned an initial score of "1." It is worth noting that 4 of the records containing the word "attention," and 1 record containing the word "aware," are verbatim duplicates (they feature different good catch numbers and dates, but identical text). The suspicion is that these duplicates are the result of some sort of human or software error. Perhaps the reporters copied and pasted text from prior good catches in cases where the error was identical, in order to save time and effort.

Step 2: Identify whether the records with an initial score of "1" from step 1 have been coded for I/D. If yes, assign a second score of "1." If no, assign score of "0". This completes the initial contextual analysis.

A resident from UNC Radiation Oncology noted which of the records labeled as related to "Inattention-Distraction" from step 1 (at face value) have been assigned an HFACS Level 2 code for "Inattention-Distraction." Of the five words labeled as "Inattention-Distraction," only 2 ("attention" and "forgot") receive a second score of "1" (i.e. were deemed to be contextually relevant to "Inattention-Distraction") in > 50% of records.

RESULTS

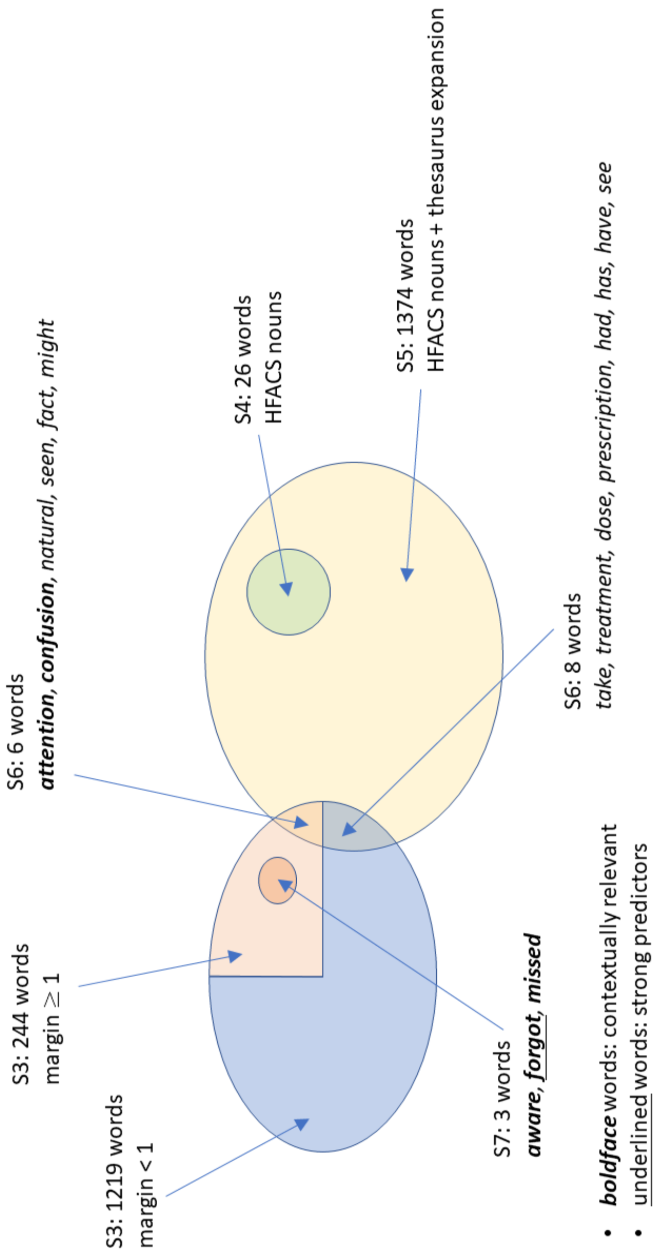


Figure 4 - Venn Diagram of Results

Figure 4 makes it clear that the Phase-1 methodology leverages two sources of evidence to select informative words: 1) empirical, annotated data about human errors (the circle on the left); and 2) theoretical descriptions of human errors, expanded by a linguistic resource (the circle on the right).

This section outlines the most relevant results of the study.

Step 3: Identity all words from Step 2 lists (n=13 and n=45). Calculate word count for each word from both lists, and numerical margin for each word in n=13 relative to n=45. Set margin of "1" as cut off point for inclusion.

The Python program created for this step: a) initializes the creation of a dictionary for Appendix 2; b) extracts each "unique" word from Appendix 2 and sets its count at "1"; c) increases the count for each "unique" word by "1" for each additional occurrence; d) initializes the creation of a dictionary for Appendix 3; e) extracts each "unique" word from Appendix 3 and sets the count at "1"; f) increases the count for each "unique" word by "1" for each additional occurrence; g) exports each unique word from Appendixes 2 and 3 into column "A" of the dataframe (only adds an Appendix 3 word in cases where the word did not previously appear in Appendix 2); h) calculates the total count per word from Appendix 2 in column "B"; i) calculates the total word count per word from Appendix 3 in column "C"; and j) exports the resulting spreadsheet as a .csv file (which is compatible with Excel, and can be converted to a .xlsx file) [see Appendix 4].

After converting the .csv file exported from Python to .xlsx, the "Inattention-Distraction" word count margin (13 records) was calculated relative to the non-"Inattention-Distraction" word count (45 records). The results from this were displayed in column "D", and all of the rows were sorted in alphabetical order by word in column "A" [see Appendix 5]. 244 of the resulting words have a count of ≥ "1", but only 12 words have a count = "2". 0 words have an "Inattention-Distraction" margin > "2".

Step 4: Use HFACS theory text to identify list of primary key words for dictionary.

"Inattention-Distraction" effectively matches the HFACS Level 2 codes and associated words below [see Appendix 6]. The 26 primary words selected for inclusion from HFACS theory are [see Appendix 8]:

- Adverse Mental States (13 words): "awareness," "complacency," "stress," "overconfidence," "vigilance," "saturation," "alertness," "drowsiness," "mental," "fatigue," "dysrhythmia," "attention," and "distraction."
- Adverse Physiological States (6 words, 1 word overlap): "illness," "hypoxia," "physical," ["fatigue" overlaps], "intoxication," "sickness," and "medications."
- Physical/Mental Limitations (7 words, 1 word overlap): "visual," "reaction," "overload," "experience," ["physical" overlaps], "capabilities," "aptitude," and "sensory" .

The primary list of words initially selected for inclusion, but then excluded because they are either adjectives/irrelevant descriptors (not bolded) or flight-related (bolded), are:

- "loss," "situational," "poor," "**flight**," "**get-home-it is**," "task," "circadian," "channelized," "medical," "**motion**," "sickness," "effects," "over-the-counter," "limitations," "insufficient," "time," "information," "inadequate," "complexity," "situation," "incompatible," "lack," "**fly**," and input."

Step 5: Use thesaurus to identify words associated with those from step 4. Add to dictionary.

A search of Thesaurus.com consisted of all 26 individual words, and all synonyms were included. This resulted in an extensive secondary list of 1374 words, which were combined with the 26 words from the primary list for a first-iteration dictionary of 1400 words [see Appendix 8].

Step 6: Check for overlapping words between dictionary from steps 4-5, and word count list from step 3 (I/D margin of "1"). These words will receive an initial score of "1".

The results of the search for overlapping words between Appendix 8 and Appendix 2 indicate an initial list of 14 unique signal words for the "Inattention-Distraction" category. There are 26 unique instances of a word occurring in overlap, and 39 instances of a word occurring in overlap total. Of the 14 unique signal words, 6 have an "Inattention-Distraction" margin of ≥ 1 in Appendix 5. These words are: "attention," "confusion," "natural," "seen," "fact," and "might." 8 of the 14 words have an I/D margin of < 1 in Appendix 5. These words are: "take," "treatment," "dose," "prescription," "had," "has," "have," and "see" [see Appendix 9].

Step 7: Determine relevance of words from step 6 to the category of I/D. These words will receive an additional score of "1" if relevant or "0" if irrelevant.

12 of the words from Appendix 9 are not specific to "Inattention-Distraction" at face value. The decision to review Appendix 5 for words that also appear to be related to "Inattention-Distraction" at face value, and that feature a margin ≥ 1, led to the inclusion of three additional words. These were selected for testing of contextual relevance. The words are "aware," "forgot," and "missed."

After selecting the five contextually relevant words ("attention," "confusion," "aware," "forgot," and "missed"), and the twelve non-contextually-relevant words ("fact," "might," "natural," "seen," "dose," "had," "has," "have," "prescription," "see," "take," and "treatment") from Appendix 9 and Appendix 5, all of the records from Appendix 2 in which these words appear were analyzed. The results of this analysis are 100% correlated with the relevance/ irrelevance of related words at face value [see Appendix 10]. In certain relevant cases, the reporter provides an explanation or suggestion in order to avoid repeating a given error moving forward.

The results of the semantic relatedness tool indicate that the words "inattention" and "distraction" are most closely related to "attention" (95% and 91%, respectively) and "confusion" (82% and 91%, respectively), as was determined from our own contextual analysis. "Inattention" is also "more related than not" (i.e. >50%) to the words "natural" (58%), "seen" (53%), "fact" (58%), "treatment" (54%), and "had/has/have" (all of which were calculated at 54%). However, this is not true of the word "distraction," which is not related to any words at greater than 50% except for "attention" and "confusion." Most of the remaining 14 dictionary words are more related to each other than not. The relationships between "aware," "forgot," "missed," and the 14 overlapping words were also tested. 30/42 (71.4%) of these relationships are >50%. The additional inclusion of three words from Appendix 5 in this experiment led to the provision of statistically significant results, since all three of these are more related than not in over 50% of pairs ("aware" in 11/14, "forgot in 9/14, and "missed" in 10/14) [see Appendix 11].

Phase 2 Results

Step 1: Identify records from a second data set (n>3000) that contain one or more of the words identified as I/D-related in the final list from Phase 1. Assign these records an initial score of "1".

In total, 2453 unique overlapping records from the second dataset are identified as containing one or more of the 17 words from Appendix 9 and Appendix 5. 4457 records are identified in total, constituting 2004 duplicate records [see Appendix 12b]. 375 records were assigned an initial score of "1" [see Appendix 12b].

Step 2: Identify whether the records with an initial score of "1" from step 1 have been coded for I/D. If yes, assign a second score of "1." If no, assign a second score of "0". This completes the initial contextual analysis.

Precision is used to measure the number of identified documents containing a given word that are relevant to "Inattention-Distraction" divided by the total number of identified documents that contain a given word. In total, the resident determined that 193/375 (51.47%) of the identified records are related to "Inattention-Distraction" (received an initial score of "1") [see Appendix 12b]. Of the five words labeled as "Inattention-Distraction," only 2 ("attention" and "forgot") receive a second score of "1" (i.e. were deemed to be contextually relevant to "Inattention-Distraction") in > 50% of records. "Attention" receives a score of 22/38 (58%), while "forgot" receives a score of 57/60 (95%). "Aware" receives a score of 32/84 (38%), "confusion" receives a score of 27/78 (35%), and "missed" receives a score of 55/115 (48%). It is worth noting that neither "attention" nor "forgot" appear in Appendix 6, and "forgot" does not appear in Appendix 9. While "attention" does appear in Appendix 9, its score of 58% suggests that it is not a strong predictor for "Inattention-Distraction" (it is almost a 50/50 chance). Only "forgot" appears to be a strong predictor for "Inattention-Distraction" at 95%.

CONCLUSION

As the methodology and results indicate, all three research aims outlined in the introduction for this paper were met:

1. **To develop a methodology to extract signal words within error reports from Radiation Oncology**

This was accomplished in Phase 1 - Step 3. A Python program was developed to determine the word count for documents in the "Inattention-Distraction" category (n=13), and the word count for all remaining documents (n=45). The numerical margin for each word in (n=13) relative to (n=45) was set at a cut-off point of "1" for inclusion.

2. **Develop a dictionary of signal words for human error categories that correlate with HFACS codes**

This was accomplished in Phase 1 - Steps 4, 5, and 6. The HFACS theory text was used to identify a list of primary key words for inclusion in a dictionary, and then a thesaurus was used to identify synonyms of these words and add them to the dictionary. Overlapping words between the dictionary and word count were identified. These words received an initial score of "1."

3. **Evaluate whether the corresponding dictionary of words can be used to determine if text documents belong to a given category or not.**

This was accomplished in Phase 1 - Step 7. It was deemed that only 2 words from the list of 14 could be used to determine if text documents belonged to the "Inattention-distraction" category or not. Three supplemental words were added from the word count, based on face value relevance to the "Inattention-Distraction" category, for a total of 5 words. All two-word combinations of the new list of 17 words were then evaluated for semantic relatedness.

The Venn diagram [Figure 4] makes it clear that the Phase-1 methodology leverages two sources of evidence to select informative words: 1) empirical, annotated data about human errors (the circle on the left); and 2) theoretical descriptions of human errors, expanded by a linguistic resource (the circle on the right). The hope was that words at the intersection of these two approaches would have the highest relevance, because their relevance is supported by two independent sources. Only one word, "forgot," was determined to be a predictor by the conclusion of Phase 2. "Forgot" does not appear at the intersection of these two approaches.

The final results of Phase 1 suggest that, while a thesaurus may be important in the development of a dictionary of words, this may not cover all relevant words, especially if only one thesaurus is used and the possibility exists that incorporating two or more thesauruses may lead to greater inclusion of relevant words in a dictionary. Additionally, it is possible that one or more thesauruses may even include words that are irrelevant. Even though only five words were selected as contextually relevant, the fact that one stems from Appendix 5 and one stems from Appendix 9 suggests that there is a need to consider both the actual words that appear in reports <u>and</u> HFACS theory when creating a controlled vocabulary for writing future reports.

It would also be helpful to have a larger data sample size for the first dataset (n=58), and even for the second dataset (n=3459). More records would theoretically increase the number of relevant words in Appendixes 5 and 9. In accord with more data, it would also be theoretically possible to test more categories than just "Inattention-Distraction." The primary interest was in "Inattention-Distraction" due to the fact that this is the most frequent category of occurrence in the dataset, but it is equally possible that more data would yield more relevant categories. With more time, it would also be helpful to test all (0, 0) words from Appendix 12a, to determine if any of these were contextually relevant in the second dataset (n=3459).

Feature selection for text classification (as mentioned in the literature review) may provide a future direction for research on the topic of contextual relevance. This may be used as a means of verifying and expanding upon the word count and margin methodology that has been developed, in combination with additional data. The goal of text classification expansion would be to train a text classifier algorithm (like chi-squared/CHI, document frequency/DF, or information gain/IG, as mentioned in the literature review) to predict categories in accord with feature selection, and then compare how these categories match to related categories in HFACS. This would require developing advanced knowledge of programming in a language like Python, which was used to build the word count in this project. It may also be worthwhile to further explore the methodology behind semantic relatedness. More specifically, it would be interesting to look at multiple approaches involved in building weighted semantic networks, and to see how these other approaches compare to the ones we explored in our own research.

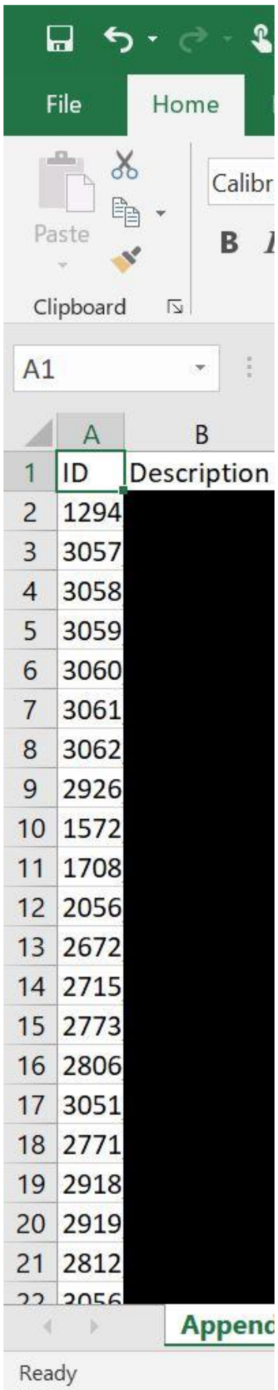
In terms of the long-term goals for this project, and for future research with additional data and sources, what this data does suggest is that there is a need to train users on how to write reports using a controlled vocabulary. This training will hopefully occur in accord with the development of an automated reporting system. Such a system would ideally lead to increases in the speed, efficiency, and ultimate accuracy with which reports are drafted in UNC-CH Radiation Oncology. Some of these goals may be accomplishable with ImprovementFlow software currently used in Radiation Oncology. Perhaps ImprovementFlow offers the ability to apply add-on software for ease of integration.

One example of add-on software would be a system that would auto-populate a list of words related to given HFACS Level 1 or Level 2 codes, from which the user can select the word(s) most related to the error that they are reporting. This would potentially accomplish three goals of significance: 1) it would create an effective "HFACS Level 3" of categories that will be more descriptive of a particular error. 2) This will be helpful for future researchers or coders, as they make sense of the content of errors, in the context of overarching HFACS Level 2 categories. Such a process may take away the "guess work" and perhaps some of the "stigma" associated with drafting error reports in which users may be reluctant to provide certain details that might otherwise be relevant in minimizing these errors moving forward. 3) This will also help the user to draft reports that more accurately and succinctly describe errors in question, and this is helpful both for drafting reports and for determining specific areas of improvement moving forward.

Limitations to this study include time constraints, sample generalizability, and implementation of text classification and feature selection. Likewise, a reader might expect a higher volume of reports and/or investigation of a greater number of methodologies for evaluation. The higher volume of reports was not pursued due to restrictions in securing an IRB. A greater number of methodological evaluations was not pursued due to aforementioned time constraints, in relation to identification and testing.

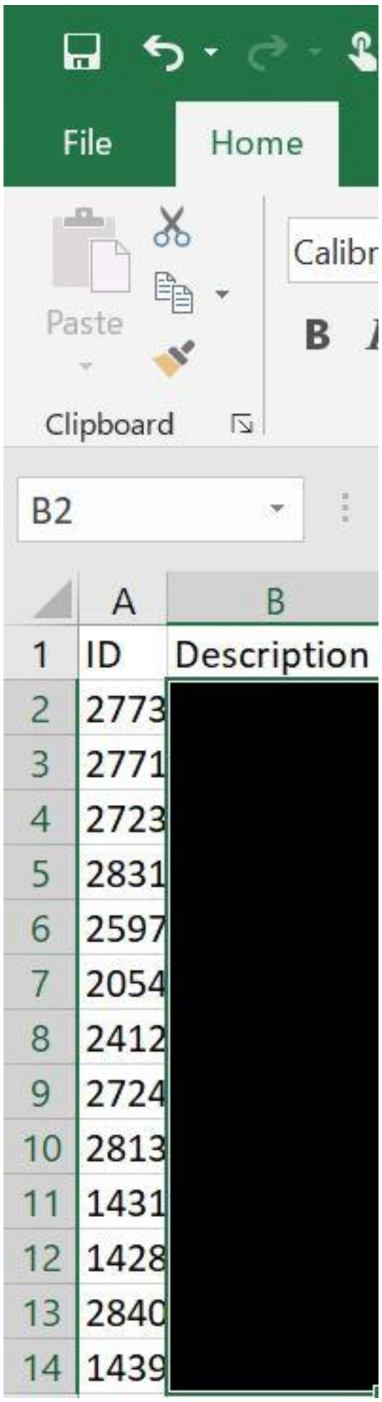
The results of these findings will be distributed to UNC Radiation Oncology for future research. As the number of error reports naturally increase over time, developing a more automated HFACS reporting system in UNC Radiation Oncology would ideally help researchers to more completely and consistently identify and describe the errors that occur. More detailed and uniform reports will hopefully also lead to a long-term reduction in the overall number of errors that occur in UNC Radiation Oncology.

APPENDIX 1

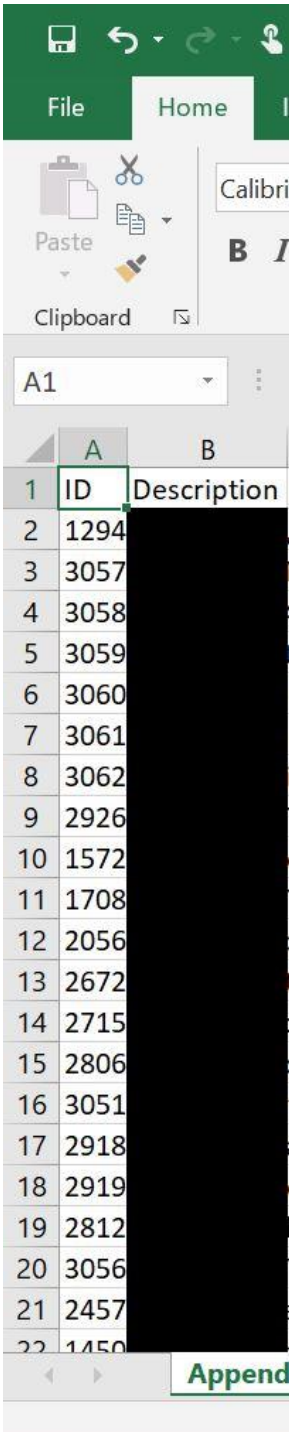


	A	B
1	ID	Description
2	1294	[redacted]
3	3057	
4	3058	
5	3059	
6	3060	
7	3061	
8	3062	
9	2926	
10	1572	
11	1708	
12	2056	
13	2672	
14	2715	
15	2773	
16	2806	
17	3051	
18	2771	
19	2918	
20	2919	
21	2812	
22	3056	

APPENDIX 2


File
Home
Paste
Clipboard
B2
A
B
1 ID
Description
2 2773
3 2771
4 2723
5 2831
6 2597
7 2054
8 2412
9 2724
10 2813
11 1431
12 1428
13 2840
14 1439
[redacted]

APPENDIX 3



	ID	Description
1	ID	Description
2	1294	[redacted]
3	3057	
4	3058	
5	3059	
6	3060	
7	3061	
8	3062	
9	2926	
10	1572	
11	1708	
12	2056	
13	2672	
14	2715	
15	2806	
16	3051	
17	2918	
18	2919	
19	2812	
20	3056	
21	2457	

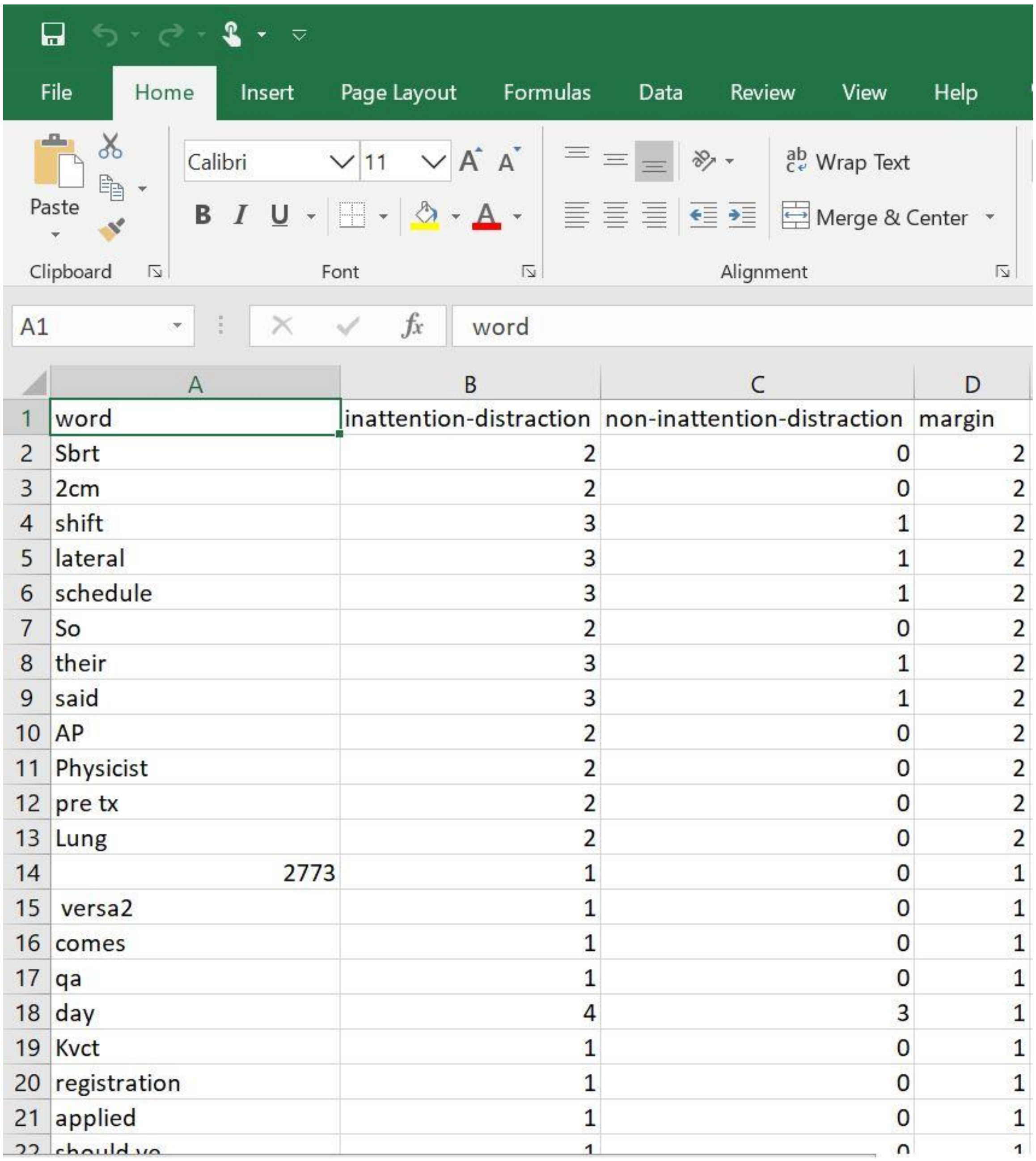
APPENDIX 4

Appendix 4.txt - Notepad

File Edit Format View Help

```
import pandas
dataframe_one = pandas.read_csv('Incident-data-idwithoutindex.csv')
dataframe_two = pandas.read_csv('Incident-data-noidwithoutindex.csv')
inattention_distraction = dataframe_one.iloc[:58, 1]
non_inattention_distraction = dataframe_two.iloc[:58, 1]

pos_dict = {}

for line in inattention_distraction:
    text = line.split(",")
    unique_words = {}
    for word in text:
        if word not in unique_words:
            unique_words[word] = 1
    for word in unique_words:
        if word in pos_dict:
            pos_dict[word]+=1
        else:
            pos_dict[word]=1

neg_dict = {}

for line in non_inattention_distraction:
    text = line.split(",")
    unique_words = {}
    for word in text:
        if word not in unique_words:
            unique_words[word] = 1
    for word in unique_words:
        if word in neg_dict:
            neg_dict[word]+=1
        else:
            neg_dict[word]=1

table = {}

for word, count in pos_dict.items():
```

APPENDIX 5



	A	B	C	D
1	word	inattention-distraction	non-inattention-distraction	margin
2	Sbrt	2	0	2
3	2cm	2	0	2
4	shift	3	1	2
5	lateral	3	1	2
6	schedule	3	1	2
7	So	2	0	2
8	their	3	1	2
9	said	3	1	2
10	AP	2	0	2
11	Physicist	2	0	2
12	pre tx	2	0	2
13	Lung	2	0	2
14	2773	1	0	1
15	versa2	1	0	1
16	comes	1	0	1
17	qa	1	0	1
18	day	4	3	1
19	Kvct	1	0	1
20	registration	1	0	1
21	applied	1	0	1
22	should ve	1	0	1

Appendix 6

Table 3.2 Selected examples of preconditions of unsafe acts

CONDITION OF OPERATOR

Adverse Mental States
- Loss of situational awareness
- Complacency
- Stress
- Overconfidence
- Poor flight vigilance
- Task saturation
- Alertness (drowsiness)
- Get-home-itis
- Mental fatigue
- Circadian dysrhythmia
- Channelized attention
- Distraction

Adverse Physiological States
- Medical illness
- Hypoxia
- Physical fatigue
- Intoxication
- Motion sickness
- Effects of OTC medications

Physical/Mental Limitations
- Visual limitations
- Insufficient reaction time
- Information overload
- Inadequate experience for complexity of situation
- Incompatible physical capabilities
- Lack of aptitude to fly
- Lack of sensory input

PERSONNEL FACTORS

Crew Resource Management
- Failed to conduct adequate brief
- Lack of teamwork
- Lack of assertiveness
- Poor communication/ coordination within & between aircraft, ATC, etc.
- Misinterpretation of traffic calls
- Failure of leadership

Personal Readiness
- Failure to adhere to crew rest requirements
- Inadequate training
- Self-medicating
- Overexertion while off duty
- Poor dietary practices
- Pattern of poor risk judgment

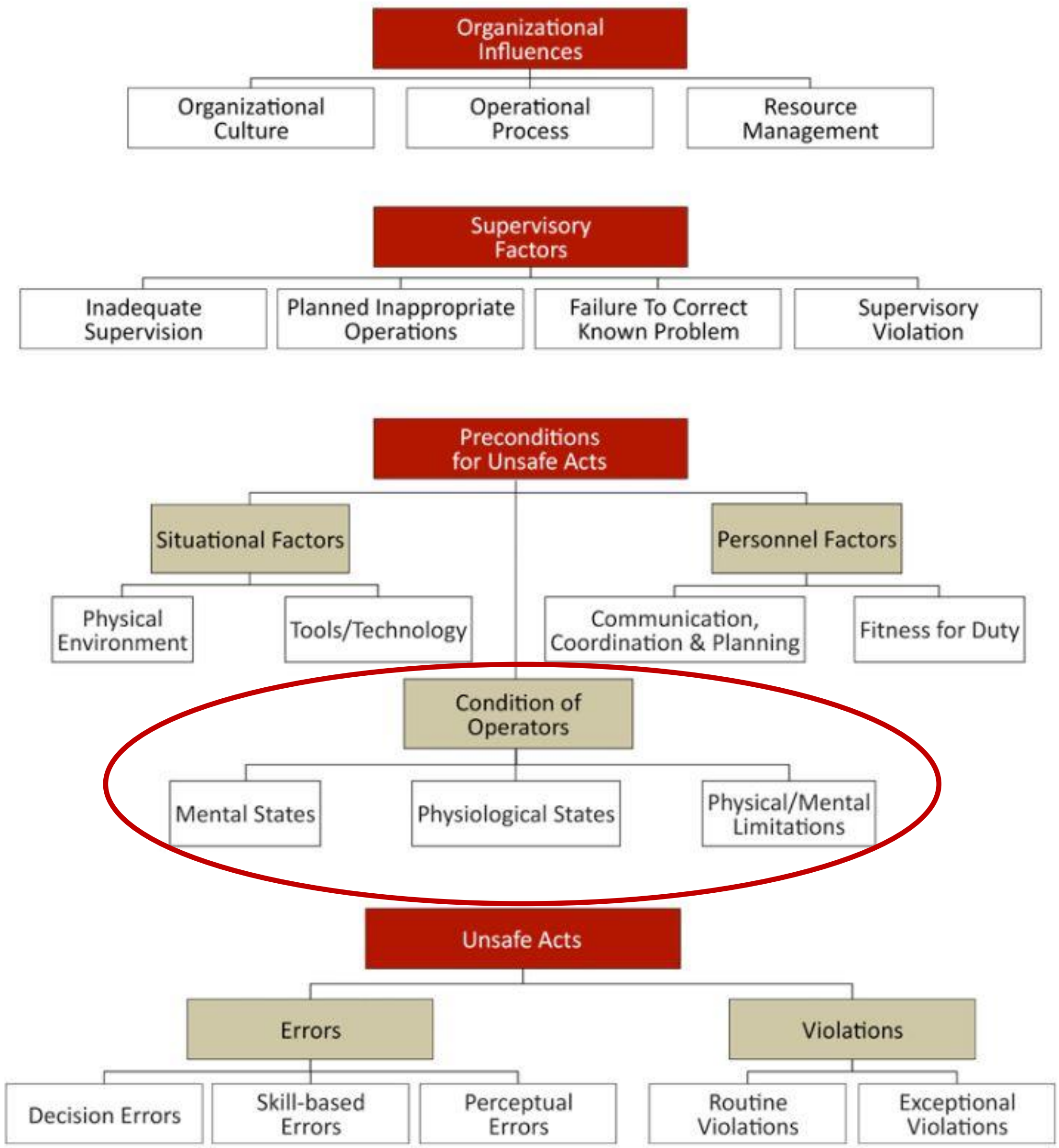
ENVIRONMENTAL FACTORS

Physical Environment
- Weather
- Altitude
- Terrain
- Lighting
- Vibration
- Toxins in the cockpit

Technological Environment
- Equipment/controls design
- Checklist layout
- Display/interface characteristics
- Automation

Note: This is not a complete listing.

APPENDIX 7


Organizational Influences
Organizational Culture
Operational Process
Resource Management
Supervisory Factors
Inadequate Supervision
Planned Inappropriate Operations
Failure To Correct Known Problem
Supervisory Violation
Preconditions for Unsafe Acts
Situational Factors
Personnel Factors
Physical Environment
Tools/Technology
Communication, Coordination & Planning
Fitness for Duty
Condition of Operators
Mental States
Physiological States
Physical/Mental Limitations
Unsafe Acts
Errors
Violations
Decision Errors
Skill-based Errors
Perceptual Errors
Routine Violations
Exceptional Violations

APPENDIX 8

Appendix 8.txt - Notepad

File Edit Format View Help

```
1400 words total/1372 not including key words (28 key words)
28 total key words/minus 2 overlapping key words (26 unique key words)
Adverse Mental States- (13 words) awareness, complacency, stress, overconfidence, vigilance, saturation, alertness, drowsiness, mental, fatigue,
dysrhythmia, attention, distraction
Adverse Physiological States- (6 words/plus 1 overlap) illness, hypoxia, physical, [fatigue], intoxication, sickness, medications
Physical/Mental Limitations- (7 words/plus 1 overlap) visual, reaction, overload, experience, [physical], capabilities, aptitude, sensory

55 words in GCs/53 not including key words (2 key words in GCs)

Key words in GCs- attention (Adverse Mental States), reaction (Physical/Mental Limitations)

Condition of Operator - Adverse Mental States

awareness
alertness
appreciation
attention - 1431 (1, 1)
consciousness
experience
information
perception
realization
recognition
understanding
acquaintance
acquaintanceship
aliveness
apprehension
attentiveness
cognizance
comprehension
discernment
enlightenment
familiarity
keenness
mindfulness
sensibilitv
```

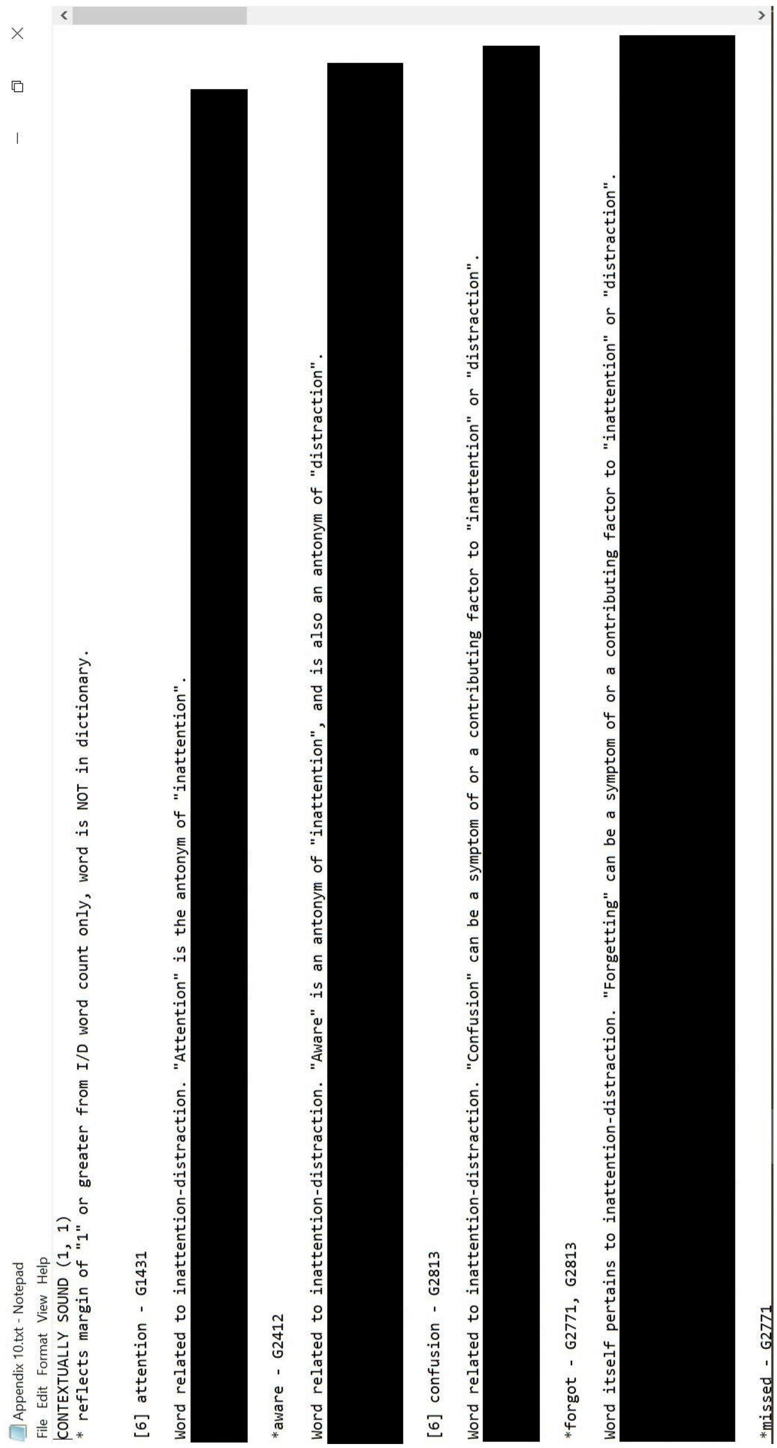
APPENDIX 9

Appendix 9.txt - Notepad

File Edit Format View Help

```
Count - 26 instances total (13 overlap) - 14 unique words - 6 words with I/D margin of "1"

6 - attention, confusion, natural, seen, fact, might
8 - take, treatment, dose, prescription, had, has, have, see

"*" before margin means duplicate
"*" after margin means margin of less than "1"

list number (1-37) - occurrence number - word - GC number - calculation (if multiple) - margin

1(16) - attention (1431) - margin 1
*2(355) - attention (1431)* - margin 1
3(610) - take (1428) - margin -1*
*4(662) - attention (1431)* - margin 1
*5(674) - treatment (2054) - (5,20) - combined margin -15*
*6(674) - treatment (2412)* - (5,20) - combined margin -15*
*7(674) - treatment (2597)* - (5,20) - combined margin -15*
*8(674) - treatment (2813)* - (5,20) - combined margin -15*
*9(674) - treatment (2840)* - (5,20) - combined margin -15*
*10(775) - confusion (2813) - margin 1
*11(824) - dose (2412) - (2,3) - combined margin -1*
*12(824) - dose (2723)* - (2,3) - combined margin -1*
*13(867) - natural (2054) - margin 1
14(982) - dose (2412)* - (2,3) - combined margin -1*
15(982) - dose (2723)* - (2,3) - combined margin -1*
*16(991) - prescription (1439) - (1,2) combined margin -1*
17(1006) - treatment (2054)* - (5,20) - combined margin -15*
18(1006) - treatment (2412)* - (5,20) - combined margin -15*
19(1006) - treatment (2597)* - (5,20) - combined margin -15*
20(1006) - treatment (2813)* - (5,20) - combined margin -15*
21(1006) - treatment (2840)* - (5,20) - combined margin -15*
*22(1044) - seen (2597) - margin 1
*23(1044) - seen (2723)* - margin 1
*24(1086) - take (1428) - margin -1*
25(1227) - fact (2771) - margin 1
*26(1233) - had (1431) - margin -17*
*27(1233) - had (2597)* - margin -17*
```

APPENDIX 10



Appendix 10.txt - Notepad

File Edit Format View Help

CONTEXTUALLY SOUND (1, 1)
* reflects margin of "1" or greater from I/D word count only, word is NOT in dictionary.

[6] attention - G1431

Word related to inattention-distraction. "Attention" is the antonym of "inattention".
[redacted]

*aware - G2412

Word related to inattention-distraction. "Aware" is an antonym of "inattention", and is also an antonym of "distraction".
[redacted]

[6] confusion - G2813

Word related to inattention-distraction. "Confusion" can be a symptom of or a contributing factor to "inattention" or "distraction".
[redacted]

*forgot - G2771, G2813

Word itself pertains to inattention-distraction. "Forgetting" can be a symptom of or a contributing factor to "inattention" or "distraction".
[redacted]

*missed - G2771

APPENDIX 11

Appendix 11.txt - Notepad

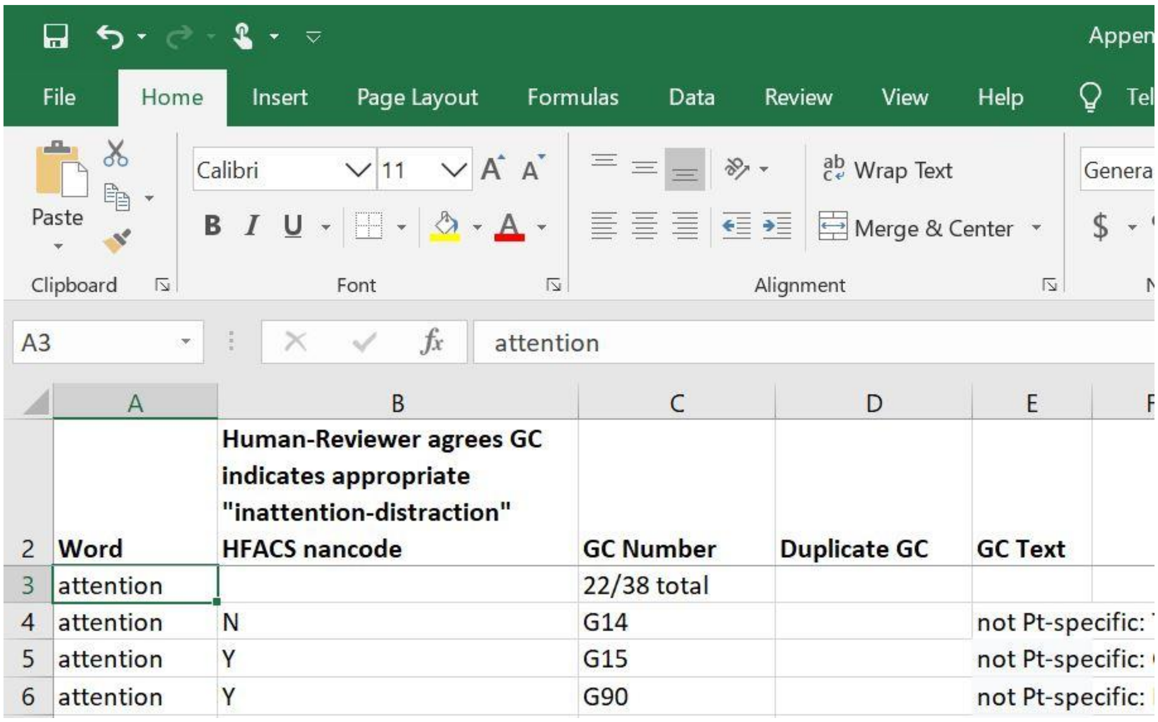
File Edit Format View Help

```
Semantic Relatedness:

Reference:     Reda Siblini and Leila Kosseim (2013). Using a Weighted Semantic Network for Lexical Semantic Relatedness. In Proceedings of Recent Advances
in Natural Language Processing (RANLP 2013), September, Hissar, Bulgaria.
02

6 - attention, confusion, natural, seen, fact, might
8 - take, treatment, dose, prescription, had, has, have, see
WC - aware, forgot, missed

inattention-attention:
96.00%
(inattention --- (sense [1.01 ]) ---> 105714175[inattention,] --- (antonym [0.25 ]) ---> 105710222[attending,attention,] <--- (reverse_sense [1.01 ]) ---
attention Totalcost =2.27)

inattention-confusion:
82.00%
(confusion --- (sense [1.05 ]) ---> 100073293[confusion,mix-up,] --- (hypernym [0.50 ]) ---> 100071785[error,fault,mistake,] --- (gloss [1.50 ]) --->
inattention Totalcost =3.05)

inattention-natural:
58.00%
(natural --- (sense [1.02 ]) ---> 301575285[natural,] --- (antonym [0.25 ]) ---> 301575756[artificial,unreal,] --- (also_see [0.25 ]) ---> 301120122
[counterfeit,imitative,] --- (also_see [0.25 ]) ---> 302470951[false,] --- (also_see [0.25 ]) ---> 300635278[wrong,incorrect,] --- (also_see [0.25 ]) --->
300023420[inaccurate,] --- (similar_to [0.25 ]) ---> 300023894[wrong,faulty,incorrect,] --- (derivationally [0.50 ]) ---> 100071785[error,fault,mistake,]
--- (gloss [1.50 ]) ---> inattention Totalcost =4.52)

inattention-seen:
53.00%
(see --- (sense [1.05 ]) ---> 200692380[regard,see,view,consider,reckon,] <--- (reverse_sense [1.01 ]) --- regard --- (sense [1.02 ]) ---> 105710673
[attentiveness,heed,paying_attention,regard,] --- (antonym [0.25 ]) ---> 105714322[heedlessness,inattentiveness,] --- (hypernym [0.50 ]) ---> 105714175
[inattention,] <--- (reverse_sense [1.01 ]) --- inattention Totalcost =4.84)
inattention-fact:
58.00%
(fact <--- (reverse_gloss [2.00 ]) --- 300021803[accurate,] --- (antonym [0.25 ]) ---> 300023420[inaccurate,] --- (similar_to [0.25 ]) ---> 300023894
[wrong,faulty,incorrect,] --- (derivationally [0.50 ]) ---> 100071785[error,fault,mistake,] --- (gloss [1.50 ]) ---> inattention Totalcost =4.5)
```

APPENDIX 12A

	A	B	C	D	E	F	G	H	I	J
16	Word	Good Catch Number	Apears in I/D word count?	I/D word count margin ? 1 ?	Appears in HFACS theory?	Appears in Thesaurus?	Appears in final dictionary?	Inattention-Distraction Record?	Word contextually related to Inattention-Distraction?	Final Calculation
17	attention	1431	Yes	Yes	Yes	Yes	Yes	Yes - 1	Yes - 1	(1, 1)
18	aware	2412	Yes	Yes	No	No	No	Yes - 1	Yes - 1	(1, 1)
19	confusion	2813	Yes	Yes	No	Yes	Yes	Yes - 1	Yes - 1	(1, 1)
20	forgot	2771	Yes	Yes	No	No	No	Yes - 1	Yes - 1	(1, 1)
21		2813	Yes	Yes	No	No	No	Yes - 1	Yes - 1	(1, 1)
22	missed	2771	Yes	Yes	No	No	No	Yes - 1	Yes - 1	(1, 1)
23	fact	2771	Yes	Yes	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
24	might	2723	Yes	Yes	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
25	natural	2054	Yes	Yes	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
26	seen	2597	Yes	Yes	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
27		2723	Yes	Yes	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
28	dose	2412	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
29		2723	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
30	had	1431	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
31		2597	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
32		2723	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
33		2724	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
34		2773	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
35		2840	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
36	has	2412	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
37	have	1439	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
38		2723	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
39	prescriptic	1439	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
40	see	2723	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
41	take	1428	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
42	treatment	2054	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
43		2412	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
44		2597	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
45		2813	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
46		2840	Yes	No	No	Yes	Yes	Yes - 1	No - 0	(1, 0)
47										

Appendix 12b



	A	B	C	D	E
2	**Word**	**Human-Reviewer agrees GC indicates appropriate "inattention-distraction" HFACS nancode**	**GC Number**	**Duplicate GC**	**GC Text**
3	attention		22/38 total		
4	attention	N	G14		not Pt-specific:
5	attention	Y	G15		not Pt-specific:
6	attention	Y	G90		not Pt-specific:

References

1. Celik M, Cebi S. Analytical HFACS for investigating human errors in shipping accidents. Accid Anal Prev. 2009;41:66-75.

2. Dekker, S. (2017). *The field guide to human error investigations*. Routledge.

3. Department of Health. (2000). An organisation with a memory. *Report of an Expert Group on Learning from Adverse Events in the NHS Chaired by the Chief Medical Officer*.

4. Diller T, Helmrich G, Dunning S, Cox S, Buchanan A, Shappell S. The human factors analysis classification system (HFACS) applied to health care. Am J Med Qual. 2013;29:181-190.

5. Donaldson, M.S., Corrigan, J.M., & Kohn, L.T. (Eds.). (2000). *To err is human: building a safer health system* (Vol. 6) National Academies Press.

6. Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Jama*, *287*(2), 226-235. [from Donaldson]

7. Evans, S. M., Berry, J. G., Smith, B. J., Esterman, A., Selim, P., O'shaughnessy, J., & DeWit, M. (2006). Attitudes and barriers to incident reporting: a collaborative hospital study. *BMJ Quality & Safety*, *15*(1), 39-43.

8. Ford EC, Los Santos LF, Pawlicki T, Sutlief S, Dunscombe P. The structure of incident learning systems for radiation oncology. Int J Radiat Oncol Biol Phys. 2013;86:11-12.

9. Ford EC, Fong de Los Santos L, Pawlicki T, Sutlief S, Dunscombe P. Consensus recommendations for incident learning database structures in radiation oncology. \ Med Phys. 2012;39:7272-7279.

10. Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of machine learning research*, *3*(Mar), 1289-1305.

11. Hutchinson A, Young TA, Cooper KL, et al. Trends in healthcare incident reporting and relationship to safety and quality data in acute hospitals: Results from the National Reporting and Learning System. Qual Saf Health Care. 2009;18:5-10.

12. Kingston, M. J., Evans, S. M., Smith, B. J., & Berry, J. G. (2004). Attitudes of doctors and nurses towards incident reporting: a qualitative analysis. *Medical Journal of Australia*, *181*(1), 36-39.

13. Kirwan, B. (2017). *A guide to practical human reliability assessment*. CRC press.

14. Li WC, Harris D. Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. Aviat Space Environ Med. 2006;77:1056-1061.

15. Mosaly, P. R., Mazur, L., Miller, S., Eblan, M., Falchook, A., Goldin, G., ... & Marks, L. B. (2015). Application of human factors analysis and classification system model to event analysis in radiation oncology. *Practical radiation oncology*, *5*(2), 113-119.

16. Mosaly PR, Mazur L, Burkhardt K, et al. Good catches: Program to improve radiation oncology process. Available at: http://www.hfes.org/web/HFESMeetings/HCSPresentations/hcs2013mosaly.pdf. Accessed June 2013.

17. Olsen NS, Shorrock ST. Evaluation of the HFACS-ADF safety classification system: Inter- coder consensus and intra-coder consistency. Accid Anal Prev. 2010;42:437-444.

18. Patterson JM, Shappell SA. Operator error and system deficiencies: Analysis of 508 mining incidents and accidents from Queensland, Australia using HFACS. Accid Anal Prev. 2010;42:1379-1385.

19. Rashid H, Place C, Braithwaite G. Helicopter maintenance error analysis: Beyond the third order of the HFACS-ME. Int J Ind Ergon.2010;40:636-647.

20. Reason J. Human error: Models and management. BMJ. 2000;320 (7237):768-770.

21. Siblini, R., & Kosseim, L. (2013, September). Using a weighted semantic network for lexical semantic relatedness. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013* (pp. 610-618).

22. Siblini, R., & Kosseim, L. (2013, September). CLaC Laboratory - NLP experiments - Words Semantic Relatedness. Retrieved from http://www.olesk.com/default.aspx#SemanticRelatedness

23. Shappell, S. A., & Wiegmann, D. A. (2000). The human factors analysis and classification system--HFACS.

24. Wallace B, Ross A. Beyond human error: Taxonomies and Safety Science. Boca Raton, FL: CRC Press. 2006.

25. Waring, J. J. (2005). Beyond blame: cultural barriers to medical incident reporting. *Social science & medicine, 60*(9), 1927-1935.

26. Wiegmann, D. A., & Shappell, S. A. (2017). *A human error approach to aviation accident analysis: The human factors analysis and classification system* (pp. 58).

Routledge.

27. Yang, Y., & Pedersen, J. O. (1997, July). A comparative study on feature selection in text categorization. In *Icml* (Vol. 97, No. 412-420, p. 35).