

Development and Validation of a Standardized Tool for Prioritization of Information Sources

Holy Akwar¹, Harold Kloeze², and Shamir Mukhi¹

1. Canadian Network for Public Health Intelligence, National Microbiology Laboratory, Winnipeg, Manitoba
2. Canadian Food Inspection Agency, Owen Sound, Ontario

Abstract

Purpose: To validate the utility and effectiveness of a standardized tool for prioritization of information sources for early detection of diseases.

Methods: The tool was developed with input from diverse public health experts garnered through survey. Ten raters used the tool to evaluate ten information sources and reliability among raters was computed. The Proc mixed procedure with random effect statement and SAS Macros were used to compute multiple raters' Fleiss Kappa agreement and Kendall's Coefficient of Concordance.

Results: Ten disparate information sources evaluated obtained the following composite scores: ProMed 91%; WAHID 90%; Eurosurv 87%; MediSys 85%; SciDaily 84%; EurekaI 83%; CSHB 78%; GermTrax 75%; Google 74%; and CBC 70%. A Fleiss Kappa agreement of 50.7% was obtained for ten information sources and 72.5% for a sub-set of five sources rated, which is substantial agreement validating the utility and effectiveness of the tool.

Conclusion: This study validated the utility and effectiveness of a standardized criteria tool developed to prioritize information sources. The new tool was used to identify five information sources suited for use by the KIWI system in the CEZD-IIR project to improve surveillance of infectious diseases. The tool can be generalized to situations when prioritization of numerous information sources is necessary.

Key Words: Emerging Pathogens, Zoonotic Diseases, Early Warning, Intelligence, Response

Correspondence: holy.akwar@phac-aspc.gc.ca

DOI: 10.5210/ojphi.v8i2.6720

Copyright ©2016 the author(s)

This is an Open Access article. Authors own copyright of their articles appearing in the Online Journal of Public Health Informatics. Readers may copy articles without permission of the copyright owner(s), as long as the author and OJPHI are acknowledged in the copy and the copy is used for educational, not-for-profit purposes

Introduction

Emerging and re-emerging infectious diseases of humans are caused by pathogens, 75% of which originate from animals or their products [1]. Recent epidemics of these diseases (e.g., Ebola, novel influenzas) have served as a reminder of the capacity of these diseases to occur unexpectedly in new locations and to affect new species [2]. Surveillance of these pathogens is

essential to protect human and animal health and to avoid potential economic disruption due to trade barriers or restrictions [3]. Unfortunately, current disease surveillance has been ineffective and/or untimely in alerting officials to emerging zoonotic diseases [4]. The capacity to detect these diseases in a timely manner will improve potential interventions and effective management of health and socio-economic risks posed by them, and improve research capacity to investigate the biologic, socio-economic, ecologic and anthropogenic factors responsible for emergence and re-emergence of these diseases [5]. Early warning systems that harness media sources leveraging low-cost technologies and engaging public health professionals are rare but critical to underpin the ability to respond to threats posed by these emerging and zoonotic pathogens [6].

The Centre for Emerging and Zoonotic Disease Integrated Intelligence and Response (CEZD-IIR) is a Canadian Safety and Security Program (CSSP) [7] funded technology demonstration project. It is designed to leverage the Knowledge Integration using Web-based Intelligence (KIWI) and Collaboration Centre technologies within the secured Canadian Network for Public Health Intelligence (CNPHI) [8] informatics platform to detect aberrations and increase early warnings signal detection in response to emerging and zoonotic disease threats of public health significance in Canada.

To detect these threats in a timely manner, real time monitoring of numerous information sources is required. However, it is impractical to stream all the media sources in the world to identify relevant and significant alerts. Hence, a standardized prioritization tool was designed to screen the plethora of information sources available globally to identify the most relevant sources to serve as feeds to the KIWI system. The evaluation tool consists of independent criteria developed to assess an information sources' ability to contribute to early detection of aberrations and/or public health threats of emerging nature. The evaluation tool is designed to remain "ever-green" in order to cope with the evolving nature of emerging diseases and their multiple facets (animal, human, and environmental interconnections with multiple stakeholders).

Epidemiological techniques are well documented for use to evaluate and validate new tools by measuring agreement beyond chance among two or more raters (evaluators). The Kappa statistic is considered the method of choice to compute inter-rater agreement using nominal and categorical data [9,10]. The Fleiss Kappa is used to assess agreement among more than two raters [11,12]. Many methods exist to compute various measures of Fleiss Kappa depending on the nature of the data. Some of these include: a SAS macros called %MAGREE and %INTRACC; Proc FREQ, MEANS, PRINT; Proc Mixed procedures etc [11,13].

This study evaluated ten disparate information sources using a newly developed standardized criteria tool to demonstrate the validity and effectiveness of the tool to select emerging and zoonotic disease information sources. The specific objectives were to validate the standardized criteria tool used for prioritizing information sources that could serve as feeds to the KIWI system for zoonotic and emerging diseases and then to identify five information sources to be used for the pilot phase of CEZD-IIR project.

Methods

Steps to develop standardized criteria were based on guidelines for the development of assessment tools produced by the Australian National Quality Council and their ranking

methodology [14]. The criteria developed for evaluating and prioritizing information sources was based on the following assumptions: The system was being built mainly for emerging and zoonotic diseases and as such all the criteria were evaluated against this group of diseases; the system was being built to meet the needs of a broad-based stakeholder that are multi-disciplinary in nature. These stakeholders included Canada's Health Portfolio (Health Canada, Public Health Agency of Canada, and Canadian Food Inspection Agency), Agriculture and Agri-Foods Canada, provincial and territorial governments (Public Health), livestock producers in Canada, and frontline clinicians (physicians and veterinarians).

Data collection and manipulation

Standardized criteria were identified and developed through expert consultations and online survey of 59 multi-disciplinary experts drawn from academia, government, and industry organizations within Canada as follows: Provincial Government (35.6%); Federal Government (25.4%); Industry (18.6%); Academic Institutions (13.6%); Independent Consultants (1.7%); and "Other" (3.4%). The survey allowed for a comprehensive assessment of stakeholder views and opinions regarding specific evaluation criteria. From the responses received and analyzed, ten criteria were identified as follows:

1. *Timeliness*: The frequency and lag (the delay between the event occurrence and reporting) with which the information is produced and disseminated to those who need to know.
2. *Usefulness*: The purpose of the information to the stakeholders. That is, how well the information supports the work of the stakeholders.
3. *Authoritativeness (Reliability/Accuracy)*: The extent to which a piece of information represents the true value of the event being described or measured.
4. *Relevance*: Alignment of the information content with the scope of the system. Areas covered by information are well within the scope of the system.
5. *Representativeness*: Information represents the target population under investigation or consideration, enabling generalizability of inference from the information. (Includes both demographic and geographic representativeness)
6. *Credibility*: Perceived or real credibility of the institution or source of the information.
7. *Validity*: There is evidence that the information has been processed and authenticated.
8. *Quality*: Information from the source is consistent and complete every time.
9. *Linguistic Style*: The clarity and complexity of the language by which the information source presents its materials to its audience.
10. *Uniqueness (Rare and Specific)*: Pertinent information that is highly specialized to a specific pathogen/disease or population or area of health and cannot be obtained elsewhere. This criterion was not evaluated using the tool. It has been included but as a binary function only in order to allow for the inclusion of those sources that present information that is valuable and unique but not available in other ways.

Each criterion was then given a relative weight (Table 1) representing its relative importance to enabling early detection of emerging and zoonotic diseases that were elicited from the evaluators

through the survey and consultation processes. These criteria were utilized in a second survey which was administered to the internal project technical team and dedicated advisors. A total of 11 evaluators (raters) were each asked to evaluate the same ten information sources by utilization of the standardized criteria tool and guidelines were provided to navigate the web-links to access the content of each information source. Raters could complete the exercise in one sitting or save their answers and resume the evaluation at another time. For each of the evaluation criteria as described above the raters were asked to select either the option that best described the source being evaluated OR to select all appropriate choices as per the survey instructions for the specific criterion being evaluated.

Each information source was evaluated against each of the criteria; a total score for the source was computed by adding the points, adjusted by the weighting factor obtained for each criterion. The total score for an information source was equal to the sum of scores from criteria 1 to 9. If a rater determined that a source did not possess a criterion being assessed, the score for that criterion was entered as 0. Once all the information sources had been evaluated using the proposed tool, the sources were ranked by total score from highest to lowest and the top five selected for potential use for the pilot phase of the project.

Statistical Analysis

All data were coded and analyzed in SAS version 9.2 (SAS Institute Inc., Cary, NC, USA) analytical package. The independence between raters was ensured through the survey administration. One of the 11 raters evaluated only three information sources and was not included in this analysis, therefore descriptive statistics were performed on the ten raters and as well as the ten information sources.

Statistical inter-rater agreement on results of the evaluation of information sources was computed using Kappa Analysis. The traditional Kappa or Cohen Kappa is designed to measure agreement only between two raters [12]. In order to measure agreement among more than two raters as was the case for this study, the Fleiss Kappa and Kendall's Coefficient of Concordance were used [11,15]. All scores given by the raters were on a continuous scale and for the purpose of performing Kappa analysis, these continuous scale scores were transformed into an ordinal categorical scale. It is not uncommon to compute Kappa, weighted Kappa (using ordinal data) and intraclass-correlation (using continuous data) in the same study when the data are collected using a continuous scale [14]. Hence, the rater's scores were grouped into ordinal categories reflecting the total score they gave to each information source. The maximum points that any information source could obtain was 1438 due to the various weighting factors for the criteria (Table 1) and this number was divided equally to create five score categories as follows:

- Category 1 = 0 - 299 points (Lowest points category)
- Category 2 = 300 - 599 points
- Category 3 = 600 - 899 points
- Category 4 = 900 - 1199 points
- Category 5 = 1200 - 1438 points (Highest points category)

Different statistical methods were used to calculate multiple raters' agreement (Fleiss Kappa values). The ordinal generalized linear mixed model approach [16] using the Proc Mixed

Procedure with Restricted Maximum Likelihood option and random effect statement for both raters and sources were used in computing the Fleiss Kappa values. In order to optimize the Fleiss Kappa value, a backwards elimination strategy was used. This involved starting with all the information sources, computing Fleiss Kappa agreement, then deleting one at a time, the information source that improved the agreement the most by being deleted, and repeating this process until no further improvement in agreement was possible (referred to as backward elimination strategy).

Furthermore, the SAS Macro called “%MAGREE” was invoked to compute multiple raters’ agreement for ordinal categorical data (Kendall's Coefficient of Concordance) and the “%INTRACC” macro invoked for continuous data. The Kendall's Coefficient of Concordance was computed because it takes into consideration the extent of disagreement among raters.

Results

A total of 59 experts from a wide range of organizations/institutions participated as described to develop the standardized weighted criteria (Table 1). The weighting reflects the relative importance of each of the criteria to evaluate information sources that enable early detection, prevention and control of emerging and zoonotic diseases.

Table 1: A Standardized Weighted Criteria Evaluation Tool Developed through a Survey of Nation-wide Stakeholders

Criteria	Relative Weight*
Usefulness	19.6
Timeliness	19.5
Reliability	18.7
Relevance	15.6
Credibility	15.3
Validity	14.5
Representativeness	14.0
Quality	13.9
Linguistics	12.7
Uniqueness	Exempted [^]
Total Score	143.8**

*Weights reflect the average score given by evaluators to each criterion

[^] An information source will be exempted from evaluation and directly used as a feed to the system if it is unique and highly specialized in the type of information it produces relevant to this system that cannot be found elsewhere.

** This is the maximum scores that anyone information source can obtain when assessed using this tool.

Ten raters then used the criteria identified in Table 1 to evaluate ten information sources. Scoring varied amongst the raters and the overall maximum score (averaged over all 10 information sources) given by a Rater was 1274.6 out of possible 1438, with a minimum of 1122.8, mean of 1174.4 (Std. Deviation of 55.4) and median of 1154.6.

The information sources were ranked from the most relevant (highest score) to the least relevant (Table 2). ProMed had the highest score (91%) and the Canadian Broadcasting Corporation (CBC, a publically supported radio and television broadcaster) had the lowest (70%). The mean score for all the sources was 1174.4 (Std. Deviation of 103.7) and the median score 1198.3.

Table 2: Information Sources Evaluated by the Project Technical Team and Technical Advisory Group Using the Criteria Tool

Information Source		Score (max. possible by 10 raters =1438)	% Score	Rank
Name	Abbreviation			
Promed	ProMed	1313.28	91.33	1
World Animal Health Information System	WAHID	1289.20	89.65	2
Eurosurveillance	Eurosurv	1253.78	87.19	3
MediSys	MediSys	1224.82	85.18	4
Science Daily	SciDaily	1208.04	84.01	5
Eurek Alert	EurekAl	1188.48	82.65	6
Canadian Swine Health Board	CSHB	1120.06	77.89	7
GermTrax	GermTrax	1082.70	75.29	8
Google	Google	1064.40	74.02	9
Canadian Broadcasting Corporation	CBC	999.20	69.49	10

Statistical Agreement Analysis

There was a noticeable difference in the variability of the ratings depending on the information sources, as an example, all Raters (100%) scored Promed very highly such that all scores from each Rater entered category 5 (Table 3) whereas Raters were widely dispersed or split regarding scores they gave CSHB.

The Proc Mixed procedure with random effect statement - Estimation Method Restricted Maximum Likelihood (REML) produced an overall Fleiss Kappa of 0.41554 or 42.0% Agreement for all 10 Raters.

Table 3: The Distribution of Raters by Categories by Information Sources (%)

Information Sources	*Score Categories			
		Three	Four	Five
Name	Abbreviation			
Promed	ProMed	0	0	100
World Animal Health Information System	WAHID	0	10	90
Canadian Swine Health Board	CSHB	10	50	40
Eurosurveillance	Eurosurv	0	10	90
Eurek Alert	EurekAl	0	30	70
Canadian Broadcasting Corporation	CBC	20	80	0
GermTrax	GermTrax	0	70	30
MediSys	MediSys	0	30	70
Science Daily	SciDaily	0	40	60
Google	Google	10	80	10

* No Raters gave scores low enough to enter categories 1 and 2, these categories are excluded from the table above.

Fleiss Kappa Analysis

Number of information sources evaluated = 10

Number of evaluators (multiple raters) = 10

Null Hypothesis: Kappa = 0 (indicates agreement equal to chance)

The Fleiss Kappa value was then optimized by systematically reducing the number of information sources. Utilizing the backwards elimination strategy to eliminate information sources (without replacement) with the most variation in scores (Table 3) and re-computing Fleiss Kappa produced the following results:

Removing CSHB from information sources: Overall Fleiss Kappa = 0.481 = 48.1% Agreement

Removing SciDaily from information sources: Overall Fleiss Kappa = 0.534 = 53.4% Agreement

Removing GermTrax from information sources: Overall Fleiss Kappa = 0.581 = 58.1% Agreement

Removing MediSys from information sources: Overall Fleiss Kappa = 0.647 = 64.7% Agreement

Removing EurekAl from information sources: Overall Fleiss Kappa = 0.725 = 72.5% Agreement

The SAS Macro %MAGREE produced a Kendall's Coefficient of Concordance of 0.504 or 50.4% for the categorized ordinal data and the %INTRACC Macro produced an interrater reliability of 0.507 or 50.7% for the continuous data, for all 10 Raters. These results are essentially equal; the minor difference could be the result of loss of information when data are categorized.

Discussion

The final Fleiss Kappa analysis performed using results from the ten evaluators (raters) was on the following five information sources after excluding the rest:

1. Promed;
2. World Animal Health Information System;
3. Eurosurveillance;
4. Canadian Broadcasting Corporation (CBC News); and
5. Google

Table 4: Guidance for Interpretation of Fleiss Kappa Agreement adapted from Landis and Koch [17]

Kappa value	% Agreement	Degree of agreement
≤ 0	0	Poor
0 - 0.2	0 - 20	Slight
0.2 - 0.4	20 - 40	Fair
0.4 - 0.6	40 - 60	Moderate
0.6 - 0.8	60 - 80	Substantial
0.8 - 1	80 - 100	Almost perfect

The five information sources mentioned above were evaluated by 10 professionally diverse raters using this new tool and achieved a Fleiss Kappa agreement of 72.5%. A Fleiss Kappa agreement of 72.5%, constitutes substantial agreement based on the reference Table 4 for kappa agreement interpretation [17]. This result indicates that this tool can be used in screening future information sources to be used by the KIWI system. Kappa is designed to estimate agreement between two raters and to have ten raters of different expertise use this standardized criteria tool to assess these information sources and still obtain a substantial agreement indicates the effectiveness of the tool to prioritize information sources. As the number of information sources evaluated increases, the value of Kappa decreases as there is more room for disagreement with more information sources [18]. The results of this study are consistent with another study that evaluated the scoring accuracy and rater reliability of a new tool (Work-ability Support Scale) and observed a Fleiss Kappa of 79% agreement [19]. It should be underscored that achieving a substantial agreement assessing the above five information sources doesn't indicate how relevant these sources are to zoonotic and emerging diseases but rather how effective the Criteria Tool is for assessment of information sources. However, three valuable information sources made the top five based on both relevance to zoonotic and emerging diseases and their Fleiss Kappa

agreement and these were Promed, World Animal Health Information System and Eurosurveillance.

Overall, this study obtained a Fleiss Kappa value of 42% for all ten information sources assessed, which is moderate agreement and consistent with, but higher than another study that validated a grading system for lateral nasal wall insufficiency that obtained a Fleiss Kappa of 17%, in which they recommended training on how scores are applied [20]. Training of raters on how to use the assessment tool improves the agreement generally and is recommended for this new tool as well. The Fleiss Kappa of 42% and Kendall's value of 50% obtained in this study are consistent with findings of other studies: that assessed agreement between thoracic surgeons rating frail behavior of patient and obtained a Fleiss Kappa of 47% and Kendall's value of 85% [21]; and another that assessed the reliability of an injury scoring system for horses and obtained a Fleiss Kappa of 66% and Kendall's value of 88% [22]. The Kendall's value represents a robust measurement of agreement compared to the value generated by traditional Kappa. This is because the Kappa method does not adequately represent the agreement level and does not take into consideration the extent of disagreement. That is, Kappa does not differentiate a disagreement of 1 and 5 versus a disagreement of 4 and 5, whereas Kendall's approach considers the magnitude of each disagreement (the disagreement between 1 and 5 is greater than the disagreement between 4 and 5).

With respect to the actual scores of how pertinent the source was for zoonotic and emerging disease detection there was a reasonable diversity of ratings. Of the ten information sources assessed, some scored quite high, for example Promed, while other scored much lower. An example of the latter was the Canadian Swine Health Board (CSHB). This might be related to the raters' lack of in-depth knowledge about CSHB which was a relatively new system being implemented in Canada. For a vast majority of the information sources, the scores were mainly in categories 4 and 5. In all, five of the ten information sources (CSHB, SciDaily, GermTrax, MediSys, and EurekAI) initially included in the overall Fleiss Kappa agreement, had scores that varied widely from raters. These five sources were either new and/or uncommon to most raters. Computing a Fleiss Kappa for these relatively new and/or unknown sources produced an agreement of 4.8%, which is slight agreement and significantly lower compared to the 72.5% Substantial Agreement obtained from the other five sources that were well known to the raters regardless of their relevance to the zoonotic and emerging diseases.

Even though this study did not focus on the competence of the raters in assessing information sources, it is important to note that the rater's knowledge of the information source being assessed is critical to an effective utilization of the tool. Other scientific studies indicate that it is not uncommon to observe rater variation. For example, a study that used orthopedic specialists to classify ankle fractures observed Kappa values varying from 20 to 64% [23]. Poor multiple raters' agreement (Fleiss Kappa 21%) was also reported among generalist physicians assessing the clinical significance of drug-drug interactions [24] and a study that assessed multiple raters' variation in scoring radiological discrepancies observed a Fleiss Kappa of 17% [20]. As there was significant variation observed when raters used the tool to evaluate information sources that they were not familiar with, it is recommended that future raters be trained to use the tool as well as to improve their knowledge of the various information sources under evaluation. This is particularly important given the plethora of information sources available globally and the need to streamline these sources to optimize the performance of early warning surveillance systems.

Limitations

There were some limitations in this study. Firstly, the raters were not randomly selected from a pool of expert evaluators and therefore may not properly reflect the performance of future raters if new information sources are evaluated with the same tool. Secondly, the diverse professional background (federal & provincial/territorial governments, academia, and industry) of these raters might have contributed to some of the variation seen in the agreement. However, different experts were used to reflect and account for the joint multi-stakeholder nature of this project and tool being developed. Despite the limitations mentioned above, the level of agreement obtained indicates that an effective standardized criteria tool for prioritizing information sources has been successfully developed, engaging broad-based Canadian stakeholders.

Conclusion

An effective standardized criteria tool for prioritizing information sources has been successfully developed, engaging broad-based Canadian stakeholders. Five information sources evaluated by 10 professionally diverse raters using this tool obtained a Fleiss Kappa agreement of 72.5%, which is substantial agreement. This study validates the utility and effectiveness of the standardized criteria tool for prioritizing information sources. The new tool was used to select five information sources suited for use by the KIWI system in the CEZD-IIR project to improve surveillance of infectious diseases. The use of the tool can be generalized to situations where prioritization of numerous information sources is needed to allow selection or rationalization of these sources. Overall, results of this study indicate that this tool can be used in screening future information sources. This is particularly important given the plethora of information sources available globally and the need to streamline these to optimize the performance of surveillance systems.

Acknowledgements

We thank members of the Technical Advisory Group of the CEDZ-IIR project for their participation in the evaluation of the information sources using the newly developed tool. We also thank two Master of Public Health candidates from University of Guelph, Zana Dukadzinac and Shahnaz Sharifzadeh for their contributions with survey preparations and data collation. We would like to acknowledge that the CEDZ-IIR project is supported by funding from the Canadian Safety and Security Program (CSSP), managed through Defence Research and Development Canada (DRDC) Centre for Security Science (CSS) and hosted by the Canadian Food Inspection Agency (CFIA) providing logistics and management for the project and thank the project manager, Mr. Harry Gardiner for his support.

References

1. Merianos A. 2007. Surveillance and response to disease emergence. *Curr Top Microbiol Immunol.* 315, 477-509. [PubMed http://dx.doi.org/10.1007/978-3-540-70962-6_19](http://dx.doi.org/10.1007/978-3-540-70962-6_19)
2. Meslin FX, Stohr K, Heymann D. 2000. Public health implications of emerging zoonoses. *Rev Sci Tech.* 19, 310-17. [PubMed http://dx.doi.org/10.20506/rst.19.1.1214](http://dx.doi.org/10.20506/rst.19.1.1214)

3. Blancou J, Chomel BB, Belotto A, et al. 2005. Emerging or re-emerging bacterial zoonoses: factors of emergence, surveillance and control [doi]. *Vet Res.* 36, 507-22. doi:<http://dx.doi.org/10.1051/vetres:2005008>. [PubMed](#)
4. Institute of Medicine (U.S.). Committee on Achieving Sustainable Global Capacity for Surveillance and Response to Emerging Diseases of Zoonotic Origin, Keusch G. Sustaining global surveillance and response to emerging zoonotic diseases. Washington, D.C.: National Academies Press 2009:312.
5. Rweyemamu M, Kambarage D, Karimuribo E, et al. 2013. Development of a One Health National Capacity in Africa: the Southern African Centre for Infectious Disease Surveillance (SACIDS) One Health Virtual Centre Model [doi]. *Curr Top Microbiol Immunol.* 366, 73-91. doi:http://dx.doi.org/10.1007/978-3-662-45791-7_244. [PubMed](#)
6. Bisson IA, Ssebide BJ, Marra PP. 2015. Early detection of emerging zoonotic diseases with animal morbidity and mortality monitoring [doi]. *EcoHealth.* 12, 98-103. doi:<http://dx.doi.org/10.1007/s10393-014-0988-x>. [PubMed](#)
7. Canadian Safety and Security Program (CSSP). Available at: <http://www.drdc-rddc.gc.ca/en/dynamic-article.page?doc=canadian-safety-and-security-program/hzvlql9b>, 2016.
8. Mukhi SN, Aramini J, Kabani A. 2007. Contributing to communicable disease intelligence management in Canada. *Can J Infect Dis Med Microbiol.* 18(6), 353-56. [PubMed](#) <http://dx.doi.org/10.1155/2007/386481>
9. Cyr L, Francis K. 1992. Measures of clinical agreement for nominal and categorical data: the kappa coefficient. *Comput Biol Med.* 22, 239-46. [PubMed](#) [http://dx.doi.org/10.1016/0010-4825\(92\)90063-S](http://dx.doi.org/10.1016/0010-4825(92)90063-S)
10. Gwet KL. 2008. Computing inter-rater reliability and its variance in the presence of high agreement [doi]. *Br J Math Stat Psychol.* 61, 29-48. doi:<http://dx.doi.org/10.1348/000711006X126600>. [PubMed](#)
- 11 Gwet K. Computing Inter-rater Reliability With SAS System. *Series: Statistical Methods For Inter-rater Reliability Assessment* 2002:3.
12. Cohen J. 1960. A Coefficient of Agreement for Nominal scales. *Educ Psychol Meas.* 20, 37. <http://dx.doi.org/10.1177/001316446002000104>
- 13 Wongpakaran N, Wongpakaran T, Wedding D, et al. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol* 2013;13:61,2288-13-61 doi:10.1186/1471-2288-13-61 [doi].
14. Mandrekar JN. 2011. Measures of interrater agreement [doi]. *J Thorac Oncol.* 6, 6-7. doi:<http://dx.doi.org/10.1097/JTO.0b013e318200f983>. [PubMed](#)

15. Fleiss J. 1971. Measuring nominal scale agreement among many raters. *Psychol Bull.* 76, 378-82. doi:<http://dx.doi.org/10.1037/h0031619>.
16. Nelson KP, Edwards D. 2015. Measures of agreement between many raters for ordinal classifications [doi]. *Stat Med.* 34, 3116-32. doi:<http://dx.doi.org/10.1002/sim.6546>. [PubMed](#)
17. Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. *Biometrics.* 33, 159-74. [PubMed](#) <http://dx.doi.org/10.2307/2529310>
18. Kundel HL, Polansky M. 2003. Measurement of observer agreement [doi]. *Radiology.* 228, 303-08. doi:<http://dx.doi.org/10.1148/radiol.2282011860>. [PubMed](#)
19. Turner-Stokes L, Fadyl J, Rose H, et al. 2014. The Work-ability Support Scale: evaluation of scoring accuracy and rater reliability [doi]. *J Occup Rehabil.* 24, 511-24. doi:<http://dx.doi.org/10.1007/s10926-013-9486-1>. [PubMed](#)
20. Mucci B, Murray H, Downie A, et al. 2013. Interrater variation in scoring radiological discrepancies [doi]. *Br J Radiol.* 86, 20130245. doi:<http://dx.doi.org/10.1259/bjr.20130245>. [PubMed](#)
21. Ferguson MK, Thompson K, Huisingh-Scheetz M, et al. 2014. Thoracic surgeons' perception of frail behavior in videos of standardized patients [doi]. *PLoS One.* 9, e98654. doi:<http://dx.doi.org/10.1371/journal.pone.0098654>. [PubMed](#)
22. Mejdell CM, Jorgensen GH, Rehn T, et al. Reliability of an injury scoring system for horses. *Acta Vet Scand* 2010;52:68,0147-52-68 doi:10.1186/1751-0147-52-68 [doi].
23. Thomsen NO, Olsen LH, Nielsen ST. 2002. Kappa statistics in the assessment of observer variation: the significance of multiple observers classifying ankle fractures [doi]. *J Orthop Sci.* 7, 163-66. doi:<http://dx.doi.org/10.1007/s007760200028>. [PubMed](#)
24. Strasberg H, Chan A, Sklar S. Interrater agreement among physicians on the clinical significance of drug-drug interactions. AMIA Annu Symp; 2013