Relationships Between Central Quality Assurance Criteria for the Assessment of Statutory Health Insurance Patients for Further Development Using Medical Informatics Methods

Mareike Burmester[1,5], Paul-Ulrich Menz [2], Vera Ries [3], Klaus-Peter Thiele[3], Bernhard van Treeck[4] and Reinhard Schuster[1,5]

[1]*Medical Advisory Board of Statutory Health Insurance in Northern Germany (MD Nord), 23554 Lübeck, Germany*

[2]*Medical Advisory Board of Statutory Health Insurance in Westphalia-Lippe (MD Westfalen-Lippe), 48153 Münster, Germany*

[3]*Medical Advisory Service Institution of the Statutory Health Insurance in North Rhine (MD Nordrhein), 40212 Düsseldorf, Germany*

[4]*Federal Joint Committee (G-BA), 10587 Berlin, Germany*

[5]*Institute of Mathematics, University of Lübeck, 23562 Lübeck, Germany*

Keywords: Medical Quality Assurance, Statutory Health Insurance in Germany, Linear Optimization, Correlations and Partial Correlations.

Abstract: In recent years, a comprehensive system has been developed both in terms of content and IT for quality assurance in the assessment for the granting of benefits for persons with statutory health insurance by the Medical Advisory Boards, which has also been enshrined in German legislation. In addition to not insignificant formal criteria and criteria relating to specific assessment areas, four criteria relevant to the entire assessment spectrum are evaluated in detail. One- and two-dimensional criteria provide an overview as an introduction to the topic. Similar to the procedure in image processing, linear optimization methods are used to infer relevant intervals of the detailed parameters from row and column totals. Using correlations and partial correlations, the relationship between the central quality criteria is shown. Methods of spherical trigonometry are generalized. For each of the three sides of the quadrilateral of the four central criteria, it is of central importance that the partial correlations are greater or smaller than the correlations overall. This is determined by the modulus value, which in the application under consideration produces the same results on all sides of the tetrahedron under consideration.

1 INTRODUCTION

The Medical Advisory Board of Statutory Health Insurance (SHI) in Germany is responsible for conducting assessments of SHI-insured individuals in order to ascertain their medical and socio-medical requirements for benefit decisions by the statutory health and long-term care insurance funds. Additionally, the Medical Advisory Board provides advice to the aforementioned insurance funds on a case-by-case basis. The assessments of the Medical Advisory Board outside of long-term care insurance in the area of health insurance are predominantly carried out by doctors. They are impartial in their assessment and are only bound by their conscience; however, they are required to adhere to the legal framework and current medical knowledge to a significant extent, see (Gemeinsamer Bundesausschuss, 2022). The relationship between law and medicine is a key topic in the field of social medicine. The abstract legal claims of insured persons are concretised by multi-layered committees, resulting in binding national regulations. Medical findings, legislation and individual case decisions by courts in various instances resulting from insufficient concretisation set precedents for subsequent similar decisions, which may be cited as such in future instances. These developments and the further development of regulations by the committees as a framework for decision-making are naturally delayed to a lesser or greater extent. This has direct consequences for the ongoing necessary further training of all those involved as experts in this interdisciplinary context, see

Burmester, M., Menz, P.-U., Ries, V., Thiele, K.-P., van Treeck, B. and Schuster, R.
Relationships Between Central Quality Assurance Criteria for the Assessment of Statutory Health Insurance Patients for Further Development Using Medical Informatics Methods.
DOI: 10.5220/0013255600003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 712-719
ISBN: 978-989-758-731-3; ISSN: 2184-4305

(Altenstetter and Busse, 2005).

In 2016, Medical Advisory Board initiated a joint evaluation of existing quality assurance initiatives and the development of a comprehensive, standardised quality assurance process for all areas of assessment. This process was developed with the involvement of all regional Medical Advisory Boards and specialist competence units, which set the standards for assessment in each specialist area. The Medical Advisory Boards have trialled the new national quality assurance plan in three test runs throughout Germany, piloted it in the largest assessment area and then gradually implemented it in practice in all assessment areas. Finally, the national quality assurance plan was even taken up by the legislature, incorporated into the German Social Code and, following a broad national discussion among all stakeholders and interested parties, adopted in a guideline confirmed by the Federal Ministry of Health with sub-legal binding force. Quality assurance follows the generally recognised principles of the quality assurance cycle for continuous evaluation and further development, see (Gemeinsamer Bundesausschuss, 2023).

The assessment is carried out by the Medical Advisory Boards, which are essentially organised by federal state. Accordingly, an initial quality check is carried out in the regionally responsible Medical Advisory Board. For quality assurance purposes, the assessments are randomly drawn on a nationwide basis. The special feature of the newly implemented system is that a partial sample is sent to another Medical Advisory Boards for a second independent assessment. If there are noticeable differences between the self-assessment and external assessment, a further quality check is carried out by a third Medical Advisory Board.

The entire process of drawing up and providing expert opinions, along with all the stages of the audit described above, is conducted via a central server on which all the expert opinions specified for the audit are accessible in anonymised form, see (Medizinischer Dienst Bund, 2023), (Ries et al., 2020) and (Ries et al., 2023).

Should discrepancies remain following the third stage of the examinations, a discussion and consensus are reached on the specific case construction at nationwide consensus conferences. After consensus conferences were held for the first time in 2018 as part of the feasibility study, they are now established as a nationwide quality assurance instrument. They promote mutual learning and take place twice a year in each of the nine assessment areas, once in person and once online via video conference. Representatives of all regional Medical Advisory Boards take part in each consensus conference, as well as at least one representative of the assigned specialist competence unit. This makes it possible to promptly implement the need for further development in the regulations identified in the consensus conferences, as the competence units set the standards for the assessment.

A total of 20 criteria have been established for the evaluation of the assessment, which are standardised across all assessment areas. Furthermore, additional technical criteria are currently being harmonised with the general criteria. In addition to the aforementioned formal criteria, there are four criteria of central importance for standardised assessment processes in the regional Medical Advisory Boards, cf. (Gaertner and van Essen, 2024), (Gaertner and Gnatzy, 2011).

Three of these central criteria are operationalised as three test aspects for the accuracy of the expert opinion, namely criteria 14, 15 and 16. Triggered by limitations on the transfer of the case file due to data protection regulations that vary in strictness from region to region, the national quality assurance plan has created the new requirement for expert opinions with the fourth central criterion 9 that they must be convincing with a concise presentation of the characteristics of the individual case that are relevant to the assessment.

2 MATERIALS AND METHODS

The following analysis focuses on the medical fields (cf. (Anja Dippmann, 2024)):

- 100: Incapacity of work
- 200: Hospital care
- 400: New and unconventional examination and treatment methods / pharmacy
- 500: Prevention and rehabilitation
- 700: Medical supplies

In the year under review 2023, 16942 internal and 4953 external evaluations were carried out in these medical fields. In addition, only the central criteria

- Criterion 9: The report contains the information necessary to assess the facts of the case
- Criterion 14: The expert opinion takes into account current medical knowledge
- Criterion 15: The expert opinion takes into account the socio-medical requirements
- Criterion 16: The result of the expert opinion is plausible and understandable in the context of the facts presented

taken into account.

The results of the assessments of the expert opinions are presented on a three-level scale (green, yellow, red). The colour red means that the criterion to be assessed has not been met. Green means that the criterion has been fully met, while yellow indicates that there is room for improvement (cf. (Gerlach, 2001)).

The first step is to compare the internal and external ratings, differentiated by cause group. The frequency of occurrence of each scale value is analysed. This is followed by a specification where the distribution of criteria 14 and 16 is analysed separately.

The aim of consistent quality assurance across all assessment regions is to achieve the same assessments both within and between regions. The first step is to compare the internal and overall ratings. This results in a total of nine different combinations of ratings with the three possible scale values (green/green, green/yellow, ... red/yellow, red/red). This can also be visualised as a nine-box table, in the same way as a four-box table. The marginal sums can be calculated from the available values. In the present constellation, a total of six proportional values can be identified, with the sum of the vertical and horizontal marginal totals having the same value. This situation is typical of medical image processing, where the object is to be inferred from certain sums per viewing direction (see (Zeng, 2010), (Jan, 2005)). Compared to medical imaging, the dimension here is small, but it is suitable for demonstrating basic aspects.

The extent to which the actual combination deviates from the possible extreme values is of interest for standardised quality assurance. In the following, we will determine the interval in which the values in the nine-field table lie if the boundary values are specified. A linear optimisation problem is solved for this purpose. It is assumed that the modelling is carried out using linear optimisation and that monotonicity is given. In addition, the proportions of the joint assessment are assumed to be between 0% and 100%. The target value is defined as the interval in which either the largest valuation differences or identical valuations occur. If there are no values of common valuations in the practical application, but only the marginal totals, intervals for deviations can thus be determined. A separate analysis is also carried out for the individual criteria.

The comparison of internal and external assessments is not the only application that enables this approach. It is also possible to analyse the interrelationships between the criteria. In the following, medical field 100 for incapacity for work is analysed, whereby the assessments of criteria 14 (medical content) and 16 (comprehensibility of the results) are compared. Furthermore, an analysis is carried out for medical field 200 (hospital care).

Finally, the correlations, the partial correlations and the modulus value are analysed. For this purpose, the random variables X_1, X_2, X_3, X_4 are defined as the four central criteria of quality assurance. The correlation $\rho_{ij} = \rho_{ji}$ describes the relationship between two variables X_i and X_j, where $i, j \in \{1,2,3,4\}, i \neq j$ applies. The partial correlations are calculated according to the following formula:

$$\rho_{ij,k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{1-\rho_{ik}^2}\sqrt{1-\rho_{jk}^2}},$$

where $i, j, k \in \{1,2,3,4\}$ are pairwise different. The modulus value

$$m = \frac{1-\rho_{ij,k}^2}{1-\rho_{ij}^2}$$

is always identical for all three combinations of partial correlation and the corresponding correlation. If the modulus value $m > 1$, the following applies

$$\begin{aligned} & m && > 1 \\ \Leftrightarrow\ & \tfrac{1-\rho_{ij,k}^2}{1-\rho_{ij}^2} && > 1 \\ \Leftrightarrow\ & 1-\rho_{ij,k}^2 && > 1-\rho_{ij}^2 \\ \Leftrightarrow\ & \rho_{ij}^2 && > \rho_{ij,k}^2 \\ \Leftrightarrow\ & |\rho_{ij}| && > |\rho_{ij,k}|. \end{aligned}$$

The calculations described above are performed separately for the internal and overarching assessment. In addition, the internal valuation is analysed separately between the medical fields considered.

Furthermore, calculations in linear optimization with Mathematica from Wolfram Research are used.

3 RESULTS

Table 1 shows the differences between the internal and the overarching assessment of the expert opinions depending on the reason group for the four central criteria 9, 14, 15 and 16. Only those expert opinions that were subject to both an internal and an overarching assessment were taken into account.

Both the red and amber assessments show a more critical tendency in the overarching assessments. This may be due to the fact that the approach to the appraisal varies from region to region and that the structure of the appraisals is different and therefore less well known. Both are reasons for standardisation.

When assessing the medical requirements in accordance with criterion 14, the best results are shown

Table 1: The results of the internal and external assessment of all Medical Advisory Boards in relation to criteria 9 and 14-16 for the year 2023 are presented herewith. The values are expressed as a percentage (%).

Medical Field	Internal			External		
	green	yellow	red	green	yellow	red
100	84.5	11.5	4.0	77.9	16.8	5.3
200	95.3	3.7	1.0	93.5	4.4	2.2
400	89.8	8.3	1.9	85.3	10.0	4.7
500	82.9	12.3	4.8	82.6	12.6	4.8
700	93.7	4.6	1.7	88.6	6.8	4.6

Table 2: The results of the internal and external assessment of all Medical Advisory Boards in relation to criteria 14 for the year 2023. The values are expressed as a percentage (%).

Medical Field	Internal			External		
	green	yellow	red	green	yellow	red
100	97.4	1.8	0.8	95.6	3.8	0.6
200	99.4	0.5	0.1	99.1	0.7	0.2
400	89.6	9.0	1.4	82.8	13.4	3.8
500	95.8	3.1	1.0	93.9	4.0	2.1
700	97.5	1.5	1.0	95.1	2.5	2.4

Table 3: The results of the internal and external assessment of all Medical Advisory Boards in relation to criteria 16 for the year 2023. The values are expressed as a percentage (%).

Medical Field	Internal			External		
	green	yellow	red	green	yellow	red
100	75.0	19.5	5.4	68.3	26.8	5.0
200	91.6	6.8	1.6	88.6	8.4	3.0
400	92.3	5.4	2.4	85.7	9.1	5.2
500	78.1	17.3	4.6	78.7	16.4	4.9
700	88.7	9.4	1.9	82.3	11.5	6.2

even more clearly in Table 2 in medical field 200 'Hospital care'.

The plausibility of the results in criterion 16 in Table 3 is assessed more critically than in the previously analysed criterion 14. Here too, the results for hospital care show the best results.

Equal internal and overarching assessments are desirable for standardised quality assurance across all regions of the assessment. In this context, the question arises as to what extent the actual combination of criteria deviates from possible extreme values.

Linear optimisation as a processing method first requires the meaningful formulation of inequalities and target values, which are then compared with regard to their results. It seems sensible to maximise the number of equal evaluations. Alternatively, it would be possible to minimise the greatest differences. In a combination of both considerations, it would first be necessary to determine which weights would lead to a sensible solution.

Table 4: The percentage of identical or dissenting votes for the identical appraisals in the internal and external appraisals.

		Internal evaluation			Sum external
		green	yellow	red	
External evaluation	green	80.0	5.0	1.5	86.5
	yellow	7.2	1.6	0.5	9.3
	red	3.0	0.7	0.5	4.2
	Sum internal	90.2	7.3	2.5	100

Table 5: Results of the optimisation problem separately for the four main criteria. In the left-hand column are the actual values for the red/green deviation case, with the corresponding calculated intervals in which the value could lie with the specified limits. On the right, these values for the case where green/green, yellow/yellow and red/red were chosen.

Crit.	Maximum Difference		Consensus	
	Real	Interval	Real	Interval
9	5.9	[2.1 , 10.1]	77.2	[69.1 , 86.6]
14	2.2	[0.5 , 2.4]	91.9	[90.8 , 96.5]
15	4.8	[1.4 , 6.4]	82.2	[78.3 , 91.4]
16	4.8	[1.5 , 7.5]	77.0	[68.4 , 86.4]

Another option would be to analyse the equally problematic ratings. The assumption that a combined analysis would not result in equally poorly assessed appraisals is not based on empirical experience. Therefore, a minimum level can be assumed for other optimisation criteria. Alternatively, the real value with the minimum and maximum possible values without further target values can be considered for empirical determination.

With regard to the differentiation between internal and overarching audits with regard to the four defined criteria, the following can be stated:

With the exception of a small deviation, the values in each row and column are monotonically decreasing. For the linear optimisation problem with the further restriction that there are expert opinions that are rated red (criterion not fulfilled) both internally and across the board, an example of 0.4% with 0.5% in reality is assumed. This means that in reality there is complete agreement (green/green, yellow/yellow, red/red) for 82.1% of the expert opinions. Assuming that the minimum of common criteria assessed in red is not restricted in the expert opinion, the interval ranges from 76.6% to 90.2%. With the restriction, the minimum increases slightly to 77.8%.

For the largest valuation differences, which are present in 4.5% of cases in real terms, the interval

Table 6: Describes the percentages of how often green, yellow and red were selected in criteria 14 and 16 in medical group 100, and how often which combination of values occurred.

		Criteria 16			
		green	yellow	red	Sum
Criteria 14	green	73.6	19.5	4.3	97.4
	yellow	0.4	1.2	0.4	2.0
	red	0.1	0.1	0.4	0.6
	Sum	74.1	20.8	5.1	100

Table 7: Describes the percentages of how often green, yellow and red were selected in criteria 14 and 16 in medical group 200, and how often which combination of values occurred.

		Criteria 16			
		green	yellow	red	Sum
Criteria 14	green	90.6	6.8	1.8	99.2
	yellow	0.1	0.4	0.1	0.6
	red	0.0	0.0	0.2	0.2
	Sum	90.7	7.2	2.1	100

ranges from 1.4% to 6.6% without the additional restriction. With the additional restriction, the upper limit is reduced to 5.4%.

With regard to the individual criteria, the following result is obtained for the variant without additional restrictions (Table 5):

The comparison of the ratings between criteria 14 and 16 in the medical field 100 results in the following nine-field table (Table 6).

When looking at criteria 14 and 16, it is initially noticeable that there is a greater asymmetry in the evaluation matrix. In addition, the monotony in the rows and columns is more limited in the medical field 200 than in the medical field 100. This is related to the dependency between the criteria, which is considered next with the correlations.

In the medical field 100, the interval from 4.4% to 5.7% (i.e. outside the real value) lies at 4.3% maximum valuation differences with monotonicity assumption and without monotonicity assumption wider between 2.5% to 5.7%. The complete match is 75.3% in real terms with monotonicity assumption we underestimate the interval from 71.5% to 74.1%. Without the monotonicity assumption, on the other hand, a wider interval of 71.5% to 76.7% is achieved.

In the medical field 200, higher levels of agreement can again be observed (Table 7).

The largest valuation differences in real terms are 1.8%, with the intervals $[2.4\%, 6.9\%]$ exhibiting monotonicity and $[1.4\%, 2.4\%]$ lacking a monotonicity assumption. In real terms, there is complete agreement at 91.1% with the intervals $[89.8\%, 90.6\%]$ with monotonicity and $[89.8\%, 91.4\%]$ without a monotonicity assumption.

Finally, the measures of correlation, partial correlation and modular value of the four central criteria are analysed in more detail.

Table 8: Describes the correlation between two of the four central criteria. All evaluated reports were analysed internally for this purpose. The matrix is symmetrical due to the symmetry property of correlation.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.295	0.393	0.601
	14	0.295		0.371	0.348
	15	0.393	0.371		0.496
	16	0.601	0.348	0.496	

Table 9: The correlations between criteria 9, 14, 15, 16 for all externally assessed reports, regardless of the medical fields.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.290	0.444	0.623
	14	0.290		0.458	0.433
	15	0.444	0.458		0.458
	16	0.623	0.433	0.458	

Table 10: Describes the difference between internal and external ratings and the correlation between the four central criteria.

		External			
		9	14	15	16
Internal	9		0.252	0.345	0.529
	14	0.252		0.392	0.370
	15	0.345	0.392		0.490
	16	0.529	0.370	0.490	

In order to obtain an indication of how strongly the four central criteria correlate in the internal assessment of the expert opinions in the individual cause groups, the cause groups are analysed separately once again.

Table 11: Correlation of the criteria based on internal reports of the medical field 100. In this case, $n = 1576$ expert opinions were analysed.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.211	0.499	0.567
	14	0.211		0.203	0.261
	15	0.499	0.203		0.548
	16	0.567	0.261	0.548	

The highest correlation between the four criteria analysed can be seen in the expert opinions on medicinal products and on new and unconventional treatment methods (medical field 400), see Table 13, cf. (Schuster, 2022). The highest number of expert opin-

Table 12: Correlation of the criteria based on internal reports of the medical field 200. Here $n = 8040$ expert opinions were analysed.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.227	0.137	0.600
	14	0.227		0.238	0.292
	15	0.137	0.238		0.354
	16	0.600	0.292	0.354	

Table 13: Correlation of the criteria based on internal reports of the medical field 400. Here, $n = 1404$ expert opinions were the subject of analysis.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.513	0.525	0.568
	14	0.513		0.507	0.512
	15	0.525	0.507		0.542
	16	0.568	0.512	0.542	

Table 14: Correlation of the criteria based on internal reports of the medical field 500. Here, $n = 1375$ reports were analysed.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.366	0.487	0.575
	14	0.366		0.399	0.386
	15	0.487	0.399		0.595
	16	0.575	0.386	0.595	

Table 15: Correlation of the criteria based on internal reports of the medical field 700. In case of medical field 700, $n = 1380$ expert opinions were analysed.

		Second Criteria			
		9	14	15	16
First Criteria	9		0.334	0.377	0.639
	14	0.334		0.490	0.405
	15	0.377	0.490		0.458
	16	0.639	0.405	0.458	

ions on hospital care is found in medical field 200 in Table 12 and thus also in the total number of expert opinions (Table 8). It is also worth noting that the differences between internal (Table 8) and overall assessment (Table 9 and, moreover, the differences between these assessments vary only moderately. If internal and overall assessment were essentially the same, the difference in table 10 would not differ in this way. The smallest differences between the medical assessments (criterion 14) and the socio-medical assessments in the individual medical field analysed are for medicines (400) (Table 13) and medical aids (700) (Table 15).

The partial correlations between three of the four variables considered in quality assurance are analysed below. This can be viewed as three sides of a polyhedron. The modulus value described in the methods section indicates whether the correlations are larger or smaller than the partial correlations. As described in the methods section, this then applies to all permutations of the three variables. For the interpretation of the tetrahedron, it can be deduced that the modulus values are greater than 1 on all three sides, which means that the partial correlations are smaller than the correlations in terms of amount.

In a comparison between the internal and overarching tests, there are only moderate differences in the correlations, partial correlations and the module value for all combinations of criteria under consideration:

Table 16: The analysis of the correlation between the first and second of the four criteria shows that if the correlation via the third criterion is excluded, a value of the partial correlation close to zero can be assumed, which means that the two criteria considered can be regarded as independent. Since all values of the modulus value are greater than 1, all correlations are greater than the associated partial correlations.

Criteria			**int./**	**Partial Correlation**			**Modul**
1	**2**	**3**	**ext.**	$\rho_{12,3}$	$\rho_{13,2}$	$\rho_{23,1}$	**value**
9	14	15	int	0.175	0.290	0.319	1.062
9	14	15	ext	0.135	0.336	0.276	1.048
9	14	16	int	0.115	0.223	0.556	1.081
9	14	16	ext	0.071	0.288	0.485	1.062
9	15	16	int	0.136	0.354	0.509	1.160
9	15	16	ext	0.115	0.386	0.440	1.120
14	15	16	int	0.244	0.421	0.203	1.091
14	15	16	ext	0.261	0.403	0.222	1.102

A comparison of the criteria considered in the internal evaluation results in the following comparison for the individual cause groups:

According to the explanations in the methods section, all partial correlations are smaller in amount than the correlations. This says nothing about the sign, i.e. the change from positive to negative correlation or the reverse. In the present analysis, a change of sign occurs only once, namely in the medical field 200 (hospital care) in the correlation without the criterion 14 (medical content).

4 CONCLUSIONS

The four quality criteria under examination interact intensively, as evidenced by the fact that the partial correlations are consistently smaller than the correlations. It is frequently posited that the high correlation observed between two variables is a consequence of their dependence on a third variable. In this instance, all three variables are of significant content relevance when the triangular configuration of statistical corre-

Table 17: Analysis of the partial correlations considered in table 16 and the module value separately for the individual medical fields.

Criteria 1	Criteria 2	Criteria 3	Med field	Partial Correlation $\rho_{12,3}$	$\rho_{13,2}$	$\rho_{23,1}$	Modul value
9	14	15	100	0.129	0.115	0.477	1.029
9	14	15	200	0.202	0.214	0.087	1.011
9	14	15	400	0.336	0.326	0.358	1.204
9	14	15	500	0.215	0.272	0.400	1.102
9	14	15	700	0.185	0.417	0.260	1.087
9	14	16	100	0.079	0.175	0.542	1.040
9	14	16	200	0.068	0.200	0.573	1.050
9	14	16	400	0.314	0.313	0.415	1.223
9	14	16	500	0.191	0.230	0.505	1.113
9	14	16	700	0.107	0.265	0.584	1.113
9	15	16	100	0.274	0.371	0.405	1.232
9	15	16	200	-0.101	0.343	0.596	1.009
9	15	16	400	0.314	0.348	0.397	1.224
9	15	16	500	0.221	0.441	0.406	1.247
9	15	16	700	0.123	0.305	0.566	1.148
14	15	16	100	0.074	0.523	0.183	1.037
14	15	16	200	0.150	0.306	0.229	1.036
14	15	16	400	0.318	0.381	0.328	1.210
14	15	16	500	0.229	0.521	0.201	1.127
14	15	16	700	0.374	0.326	0.233	1.131

lations is taken into account. This remains the case when three edges of each of the four variables are considered as tetrahedrons. It is important to note that correlations and partial correlations are statistical variables that may not be able to explain a causal relationship. In the context under consideration, there are numerous substantive discussions between physicians with the participation of computer scientists in the consensus conferences described in the introduction. The extensive details of the discussion provided a basis for statistical examination. The result was that the substantive arguments put forward by doctors in the review of expert opinions were largely confirmed statistically. This alternating suggestion between practitioners and statisticians will contribute to the further development of quality assurance in the Medical Advisory Boards with high performance, cf. (Nüchtern et al., 2015), (Newhouse, 2002).

A further point of discussion in the further development of quality assurance in medical assessments is the relationship between cost and benefit. A reduction in the number of criteria to be assessed will result in a reduction in the effort required. On the one hand, the expert groups are engaged in a discussion regarding the identification of criteria that are similar in content and, therefore, amenable to a summary. This is particularly pertinent to criteria that are employed solely within individual event groups. The rationale for these criteria was based on the assumption that they could not be evaluated using the general criteria. However, the testing experience from the discussions in the consensus conferences has shown that this is certainly possible. This has resulted in a modification of the test criteria. This paper does not show to what extent the audit has led to a significant improvement in the assessment. While a modification and summarisation of the review criteria may appear to result in a deterioration, it has, in fact, led to an overall improvement.

In addition to the discussion of the proximity of test criteria in terms of content, the question also arises as to which criteria are viewed in the same way by the physicians in each case without content-related proximity. This concerns an association of judgement errors. This can also motivate a reduction in the number of endpoints.

This paper focussed on the fact that there are criteria whose further reduction would not make sense and would be clearly counterproductive for quality assurance overall.

REFERENCES

Altenstetter, C. and Busse, R. (2005). Health care reform in germany: patchwork change within established governance structures. *Journal of health politics, policy and law*, 30(1-2):121–142.

Anja Dippmann, *et al.*. (2024). Qualtitätssicherung der Begutachtung für die Gesetztliche Krankenversicherung QSKV - Jahresbericht 2024 - Datenjahr 2023. *Homepage MD Bund*.

Gaertner, T. and Gnatzy, W. (2011). Zum sachverständigenstatus im medizinischen dienst der krankenversicherung am beispiel des mdk hessen. *GuP*, 5:166–173.

Gaertner, T. and van Essen, J. (2024). Der Medizinische Dienst - eine sozialmedizinische Institution der Qualitätssicherung im Gesundheitssystem. *Das Gesundheitswesen*.

Gemeinsamer Bundesausschuss (2022). Beschluss des Gemeinsamen Bundesausschusses über die Veröffentlichung des Berichts des Medizinischen Dienstes Bund gemäß §16 Teil A MD-QK-RL: Bericht über die im Jahr 2021 durchgeführten Qualitätskontrollen. *Homepage Gemeinsamer Bundesausschuss*.

Gemeinsamer Bundesausschuss (2023). Richtlinie des Gemeinsamen Bundesausschusses nach §137 Absatz 3 SGB V zu Kontrolle des Medizinischen Dienstes nach §275a SGB V. *Homepage Gemeinsamer Bundessausschuss*.

Gerlach, F. M. (2001). *Qualitätsförderung in Praxis und Klinik: eine Chance für die Medizin; 24 Tabellen*. Thieme.

Jan, J. (2005). *Medical image processing, reconstruction and restoration: concepts and methods*. Crc press.

Medizinischer Dienst Bund (2023). Regelmäßige Berichterstattung der Medizinischen Dienste über ihre Tätigkeit und Personalausstattung. *Richtlinie des Medizinischen Dienstes Bund nach §283 Absatz 2 Satz 1 Nr. 8SGB V*.

Newhouse, J. P. (2002). Why is there a quality chasm? *Health affairs*, 21(4):13–25.

Nüchtern, E., Bahemann, A., Egdmann, W., van Essen, J., Gostomzyk, J., Hemmrich, K., Manegold, B., Müller, B., Robra, B., Röder, M., et al. (2015). Soziale Sicherheit braucht Sozialmedizin Selbstverständnis von ärztinnen und ärzten in der sozialmedizinischen Begutachtung und Beratung. *Das Gesundheitswesen*, 77(08/09):580–585.

Ries, V., Thiele, K.-P., Schuster, M., and Schuster, R. (2020). It-structures and algorithms for quality assurance in the health insurance medical advisory service institutions in germany. In *HEALTHINF*, pages 353–360.

Ries, V., Thiele, K.-P., van Treeck, B., Schroeer, S., Witt, C., and Schuster, R. (2023). It-structures and algorithms for quality assurance in the medical advisory service institutions in germany. step 2: To err is human. consensus-conferences. In *HEALTHINF*, pages 271–278.

Schuster, R. (2022). *Strukturelle Beratung der Krankenkassen durch den Medizinischen Dienst in der Pharmakotherapie. Die Verwendung epidemiologischer und gesundheitsökonomischer Analysen im Zeitalter von Digitalisierung und Bid Data*. Agor, K and Knieps, F and Hartweg, HR.

Zeng, G. L. (2010). *Medical image reconstruction*, volume 530. Springer.