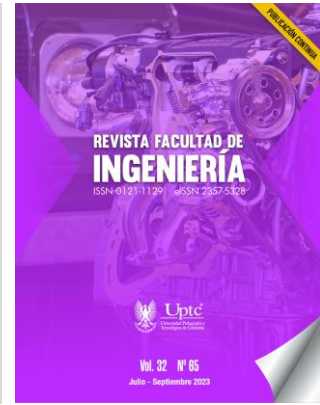


## Revista Facultad de Ingeniería

Journal Homepage: <https://revistas.uptc.edu.co/index.php/ingenieria>



# Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports

Daisy-Yisel Meneses-Lopez<sup>1</sup>

Martha-Eliana Mendoza-Becerra<sup>2</sup>

Salvador Garcia-Lopez<sup>3</sup>

**Received:** July 03, 2023    **Accepted:** September 13, 2023    **Published:** September 30, 2023

**Citation:** D.-Y. Meneses-Lopez, M.-E. Mendoza-Becerra, S. Garcia-Lopez, "Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports," *Revista Facultad de Ingeniería*, vol. 32, no. 65, e16314, 2023. <https://doi.org/10.19053/01211129.v32.n65.2023.16314>

<sup>1</sup> M. Sc. (c) Universidad del Cauca (Popayán-Cauca, Colombia). [dymeneses@unicauca.edu.co](mailto:dymeneses@unicauca.edu.co) ORCID: [0009-0001-2826-728X](https://orcid.org/0009-0001-2826-728X)

<sup>2</sup> Ph. D. Universidad del Cauca (Popayán-Cauca, Colombia). [mmendoza@unicauca.edu.co](mailto:mmendoza@unicauca.edu.co). ORCID: [0000-0003-4033-2934](https://orcid.org/0000-0003-4033-2934)

<sup>3</sup> Ph. D. Universidad de Granada (Granada, España). [salvagl@decsai.ugr.es](mailto:salvagl@decsai.ugr.es) ORCID: [0000-0003-4494-7565](https://orcid.org/0000-0003-4494-7565)



## **Abstract**

In the health sector, the reports on delivery of prescriptions and the assignment of medical appointments are generated by the Health Service Provider Institutions and delivered to the Health Service Promoting Entities. These reports usually have an incoherent structure; inconsistencies in the format; non-existent, incomplete, or non-standardized data. These problems affect data quality and hinder the reliability of the information. To address this, it is proposed to adapt Kahn's data quality categories, to these reports, considering that the health sector accepts them categories and contemplates not only the structure and domain of the data but also its completeness and plausibility (credibility). This research followed the methodology of Pratt's Iterative Research Pattern, studies related to the subject were observed, and the attributes of prescription delivery and appointment assignment were analyzed to understand the problem and its implications in detail. We then adapted the data quality categories proposed by Kahn, taking into account the problems identified in these reports. Subsequently, a group of health experts evaluated the proposed adaptation using the focus group technique. The results, according to their perception, showed that the prescription delivery report obtained 66.7% in the "Completely Agree" category and 33.3% in the "Agree" category; medical appointment assignment had 73.3% in "Completely Agree" and 26.7% in "Agree", according to the Likert scale. In conclusion, this research contributes to strengthening the data quality of these reports by providing guidelines to improve the reliability of the information.

**Keywords:** completeness; conformance; data quality; data quality categories; health; health regulatory reporting; medical appointment scheduling; medication delivery; plausibility.

## **Adaptación de categorías de calidad de datos de Kahn para reportes de entrega de medicamentos y asignación de citas médicas**

### **Resumen**

En el sector de la salud, los reportes de entrega de medicamentos y asignación de citas médicas son generados por las Instituciones Prestadoras de Servicios de

Salud y entregados a las Entidades Promotoras de Servicios de Salud. Estos reportes no suelen tener una estructura coherente, presentan inconsistencias en el formato, datos inexistentes, incompletos o no normalizados. Estos problemas afectan la calidad de estos y dificultan la confiabilidad de la información. Con el objetivo de abordar este problema, se propone adaptar las Categorías de Calidad de Datos de Kahn a estos reportes, teniendo en cuenta que estas son aceptadas por el sector salud y no solo contemplan la estructura y dominio del dato, sino también la completitud y plausibilidad (credibilidad) del mismo. Para llevar a cabo esta investigación se siguió la metodología del Patrón de Investigación Iterativa de Pratt, se observaron estudios relacionados con el tema y se analizaron los atributos de los reportes de entrega de medicamentos y asignación de citas médicas para comprender en detalle el problema y sus implicaciones. Luego, se adaptaron las categorías de calidad de datos propuestos por Kahn teniendo en cuenta los problemas identificados en estos reportes y, posteriormente, dicha adaptación fue evaluada por un grupo de expertos en el sector salud mediante la técnica de grupo focal. Los resultados, según la percepción de los expertos, demostraron que la adaptación realizada para el reporte de entrega de medicamentos obtuvo un 66.7% en la categoría “Completamente de Acuerdo” y 33.3% en “De Acuerdo”; para asignación de citas médicas un 73.3% en “Completamente de Acuerdo” y un 26.7% en “De Acuerdo” según la escala de Likert. En conclusión, esta investigación contribuye al fortalecimiento de la calidad de los datos de estos reportes en el sector salud y proporciona pautas para mejorar la confiabilidad de la información.

**Palabras clave:** asignación de citas médicas; calidad de datos; categorías de calidad de datos; completitud; conformidad; entrega de medicamentos; plausibilidad; reportes normativos en salud; salud.

## I. INTRODUCTION

The healthcare sector generates large amounts of data on a daily basis; however, the lack of data quality has become a problem for their analysis since low-quality information can lead to erroneous conclusions and decisions; instead, the use of accurate and reliable data will allow to make informed decisions and generate value in this sector [1].

In this context, Health Service Providing Institutions (IPS by its Spanish acronym) own large amounts of data, part of which is reported to Health Service Promoting Entities (EPS by its Spanish acronym) to follow-up the services provided. However, within these reports, in particular on the delivery of prescriptions and the assignment of medical appointments, there are frequent problems such as (i) attributes that lack a coherent structure and are inconsistent in format; (ii) incomplete, non-existent, or registered unique medication code with the Anatomical Therapeutic Classification (ATC) codes, which does not allow identifying the commercial presentation of the medication delivered to the affiliate; (iii) non-existent or incomplete diagnosis codes according to the International Classification of Diseases (CIE10 by its Spanish acronym); (iv) information on members who do not exist in the database or who are affiliated with other EPS; (v) quantities prescribed, delivered, days of treatment, dates of submission and delivery with errors and inconsistencies, e.g., delivered quantities greater than those formulated, delivery dates prior to the prescription [2]. Systematic mapping allowed us to find studies related to data quality. In [3], they propose a process that standardizes the data structure to detect and correct errors in the defined variables. In [4], they compare discrepancies between source and target data using three categories: completeness, consistency, and syntactic validity. Furthermore, in [5], they propose data cleaning to evaluate data, metadata, outliers, and duplicates; then, they detect anomalies and follow up inconsistencies to standardize the data. Kahn [6] proposes three categories of quality: *conformance*, which evaluates whether values comply with syntactic or structural constraints; *completeness*, which verifies the presence or absence of data at one or more points in time; and *plausibility*, which describes the credibility or veracity of values. These studies [4][5] have addressed data quality in the health sector, focusing on their

structure and domain and on the standardization and detection of anomalies to correct errors in the variables. However, they do not consider the *completeness* and *plausibility* categories proposed by Kahn, which are essential to validate the data. Furthermore, these Kahn categories have been used and accepted in the health sector in other countries [7][8][9].

Considering the importance of addressing Kahn's data quality categories, this research proposes to adapt these categories to the specific characteristics of prescription delivery and medical appointment assignment reports, seeking a higher quality of data that will allow EPS to perform analyses.

This research used the Iterative Research Pattern methodology, which Observation stage includes a systematic mapping based on [10][11]; and the structure and guidelines established by the Ministry of Health and Social Protection (MSPS by its Spanish acronym); for these, two reports were reviewed. In the *Problem Identification* stage, the problems concerning data quality were related to each attribute of the reports. In the *Solution Development* stage, the general and specific adaptation of Kahn's data quality subcategories to these reports was carried out. Finally, in the *Solution Testing* stage, the proposed adaptation to each subcategory was validated by a focus group with experts in the health sector.

This article presents the methodology used to adapt the two reports, describes this adaptation in detail and its evaluation by a focus group; finally, it presents the conclusions.

## **II. METHODOLOGY**

The research employed the Iterative Research Pattern (PII) methodology proposed by Pratt [12], which comprises four stages: observation, problem identification, solution development, and solution testing.

### **A. Observation**

**1) Systematic Review.** This review was conducted based on [10][11] to define the research questions, search terms, databases, and inclusion/exclusion criteria. The most relevant articles found in this review are presented below.

## Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports

In 2016, an optimization model for ETL was proposed [3], it contemplates: (i) *the Prerequisite phase*, which standardizes the data structure; (ii) *the Main phase*, which detects outliers and inconsistencies and records them in a table of variables; (iii) *the Alternative phase*, which stores the error history, manages the variables, and evaluates the process using two indicators (confidence and support). In 2018, an approach was proposed to validate ETL processes through balancing tests composed of five phases [4]: (i) *Defining generic properties* through completeness, consistency, and syntactic validity and checking for mismatches between data; (ii) *Identifying source-to-target mappings* through aggregation operations to join records; (iii) *Testing mappings* to verify matches between source and target record counts; (iv) *Approach evaluation* to detect record failures; and (v) *Automated mutation testing* to evaluate failures in the target table.

Later in 2019, a quality assurance (QA) process was proposed [7]; it focuses on Kahn's *Conformance* and *Completeness* categories, which are applied in all ETL stages starting with the *Completeness* category. Here, a count of source and transformed rows is made, followed by the *Relational Conformance* category, which checks that foreign key values match foreign sources. Finally, the *Value Conformance* category quantifies the amount of mapped and unmapped data. The same year, they proposed a data quality assessment through data cleansing [9], based on Kahn's data quality categories, applied in a data retention cycle: (i) evaluates the *Conformance* category of the data model at the table level and leaves the *Plausibility* and *Completeness* categories for later cycles; (ii) updates the data dictionary and data characterization; (iii) reports empirical data characterization; and (iv) discusses the results of the previous cycles and creates a plan for error mitigation.

In the same year, 2019, an automated framework for data cleansing was proposed [5] using a three-module architecture: (i) *Data evaluation* assesses raw data, extracting metadata and calculating descriptive statistics; (ii) *Data quality control* detects missing values, anomalies, and duplicates; and (iii) *Data standardization* ensures data matching with reference models through lexical and semantic comparisons.

Finally, in 2020, a data monitoring architecture is proposed to measure the quality of data in the ETL process [8] using Kahn's categories and employing diagnostic filters to generate error events, which are stored in the fact table linked to an audit dimension for their respective analysis.

The results obtained from the systematic mapping indicate that the health sector accepts Kahn's data quality categories [7],[8],[9]. This study involved the participation of approximately 100 professionals from different disciplines from the US and international networks and projects, who contributed to the development and revision of harmonized data quality terminology. Furthermore, this study involved more than 540 million patient records that support the robustness and relevance of these categories in the medical field. Also, in this systematic review, it can be identified that studies [3],[4],[5] have addressed data quality in the health sector, focusing on the structure and domain of the data, and on the standardization and detection of anomalies to correct errors in the variables. However, they do not consider the *Completeness* and *Plausibility* categories proposed by Kahn. These are important to validate the veracity of the data. In addition, it is noted that these studies were developed in the United States [7],[9], Germany [8], Italy [4], and Colombia [3]. The latter is applied in a case study of environmental data.

**2) Reports on Prescription Delivery and Medical Appointments.** We reviewed the consolidated reports generated by the IPS based on the structure established by the Ministry of Health and Social Protection (MSPS) through Resolution 1604 of 2013 [13] that defines the requirements for prescription delivery. Resolution 1552 of 2013 was also considered [14]; it sets the guidelines for medical appointment assignment reports. These resolutions establish the standards and procedures to be followed by IPS when generating reports intended for EPS to follow up and control prescription delivery and the assignment of medical appointments.

### **1. Identification of the problem**

Table 1 details the problems identified for each attribute of the two reports. The nomenclature used in the columns indicates: (V) Empty, a record with no data is found; (NN) Non-standardized, the data is not presented in a standard format; (E) Erroneous, the attribute value is incorrect or inaccurate; (IN1) Incomplete, it refers

**Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports**

to cases where information is missing or the data is incomplete; (IN2) Inconsistent, the attribute value does not match other data or is not consistent with general information; and (IN3) Inexistent, the data is not found.

**Table 1.** Problems identified for each attribute.

Attribute	V	NN	E	IN1	IN2	IN3
Pending quantity	X		X		X	
Prescribed quantity	X		X		X	
Delivered quantity	X		X		X	
Medication CUM code	X		X	X	X	X
Habilitation code	X		X	X	X	X
Concentration	X	X	X	X	X	X
Affiliate department	X	X	X	X		X
IPS department	X	X	X	X		X
Main diagnosis	X	X	X	X		X
Related diagnosis	X	X	X	X		X
Days of treatment	X		X		X	
Date of authorization	X		X	X	X	X
Delivery record date	X		X	X	X	X
Prescription date	X		X	X	X	X
Filing date of the medical prescription by the user	X		X	X	X	X
Date on which the user requests the appointment	X		X	X	X	X
Date on which the user requests the appointment to be assigned	X		X	X	X	X
Date of the appointment	X		X	X	X	X
Pharmaceutical form	X	X	X	X		X
Affiliate's municipality	X	X	X	X		X
IPS' municipality	X	X	X	X		X
Provider's NIT (without DV)	X		X	X	X	X
Name of the medication	X		X	X		X
Names and last names of affiliate	X		X	X		X
Identification number	X		X	X	X	X
Total number of appointments assigned	X		X		X	
Supplier's company name	X		X	X		X
Affiliated regime	X	X	X	X		X
Service requested	X	X	X	X	X	X
Document type	X	X	X	X		X
Form of administration	X	X	X	X		X
Genre	X	X	X	X		X
Affiliate's phone number	X					
Delivery record date	X		X	X	X	X
Type of medication delivery, at home or at point of delivery?	X	X				
Satisfactory delivery number	X		X	X	X	X



## B. Solution Development

First, a general adaptation of Khan's data quality categories and subcategories was performed and evaluated by implementing two strategies: (i) **Verification** to check whether the data values conform to established expectations and local knowledge; and (ii) **Validation** to check that data values are aligned with respect to external sources.

For this purpose, we started by defining each subcategory, identifying the most relevant characteristics; then, we determined the *Verification Criteria* that describe the adaptation made for each subcategory and the *Validation Criteria* that present the adaptation considering external sources, as shown in Table 2.

**Table 2.** General adaptation of data quality categories and subcategories.

Category	Subcategory	Attribute	Verification Criteria	Validation Criteria
Conformance	Value Conformance	Internal formatting restrictions	Data type, data format, length	NA
		Allowable values or ranges	Data domain	NA
	Relational Conformance	Complies with referential integrity restrictions	Foreign key	External databases
		Complies with the uniqueness constraint	Primary key	NA
		Complies with the nullity restrictions	Invalidity	NA
	Computational Conformance	The calculated values conform to computational or programming specifications	NA	NA
Completeness	Completeness	The absence of values at a single point in time is in accordance with local or common expectations	Attributes comply with the absence of the value	NA
Plausibility	Uniqueness Plausibility	Data values identifying a single object are not duplicated	Attributes that meet the condition	NA
	Atemporal Plausibility	It seeks to determine whether the values, distributions or densities of the observed data agree with local or "common" knowledge	Values and distributions	External databases
			Logical constraints between values	External databases
			Repeated measurement values	External databases

**Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports**

Category	Subcategory	Attribute	Verification Criteria	Validation Criteria
	Temporal plausibility	Observed or derived values conform to expected time properties	Observed or Derived Values	NA
			Sequences of values	NA
			Value density measures	NA

Once the verification and validation criteria had been identified, we proceeded to adapt each attribute of the prescription delivery and medical appointment assignment reports.

**C. Solution Testing**

The proposed adaptation was evaluated by a focus group composed of experts from the health sector with more than 20 years of experience (see Table 3). Before the focus group session, an invitation was sent via e-mail to the experts, attaching reading material with a description of the adaptation carried out. During the session, a presentation was made to explain the adaptation made to each subcategory of data quality in the two reports in detail. Then, the experts provided feedback and discussed the topic. At the end of the session, the experts were asked to complete a questionnaire (see Table 4) to evaluate the adaptation made for each of the reports.

**Table 3.** Health experts.

Expert ID	Occupation	Studies
E1	Vice president of health services, professor of postgraduate courses.	Surgeon, Specialist in Health Services Management, Specialist in Quality Management and Health Auditing.
E2	Expert advisor in health product development, Vice president of health.	Surgeon, Specialist in Health Services Management, Specialist in Quality Management and Health Auditing.
E3	Professional in National Health Services, General Practitioner	Physician and surgeon, Specialist in Hospital Administration, Specialist in Auditing and Quality Assurance with an emphasis in Epidemiology from EAN, Nation-wide Lecturer in the field of Health Administration, Advisor to Hospitals and E.S.E.S. in Cauca and Valle del Cauca.

**Table 4.** Questions in the expert evaluation questionnaire

Closed Questions						
ID	Do you consider the adaptation made to the data quality subcategories of the report to be adequate?	5	4	3	2	1
2	Value Conformance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Relational Conformance	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Completeness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Uniqueness Plausibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Atemporal Plausibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Temporary Plausibility	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Open Questions						
9	Regarding the report, do you consider that any criteria of the data quality subcategories should be added to the proposed adaptation? If so, which ones and why?					
10	Regarding the report, do you consider that any criteria of the data quality subcategories should be eliminated from the proposed adaptation? If so, which ones and why?					
11	Do you have any additional comments about adapting the proposed data quality categories?					
12	Do you have any additional suggestions about adapting the proposed data quality categories?					

Nomenclature: (5) Strongly Agree, (4), Agree, (3) Neither Agree Nor Disagree, (2), Disagree, (1) Strongly Disagree

### III. RESULTS

#### A. Adaptation of Data Quality Categories of the Medication Delivery Report

The adaptation of the *Value Conformance* and *Relational Conformance* subcategories to each attribute, along with the verification and validation assessment strategies, are shown in Table 5; the *Completeness* category in Table 6; the *Uniqueness Plausibility* subcategory in Table 7; the *Atemporal Plausibility* subcategory in Table 8; and the *Temporal Plausibility* subcategory in Table 9.

**Table 5.** Adaptation of Value Conformance and Relational Conformance - Prescription Delivery.

ID	Attribute	Value Conformance				Relational Conformance				
		Verification				Validation	Verification			Validation
		DT	LE	FO	DO		ON	PK	FK	
1	Document type	ST	2	NA	NA	NA	NN	NA	DBA	BDUA
2	Identification number	N	16	NA	NA	NA	NN	NA	DBA	BDUA
3	Names and last names	T	80	NA	NA	NA	NN	NA	NA	NA
4	Genre	ST	9	NA	F,M	NA	NN	NA	DBA	BDUA
5	Affiliated Regime	ST	12	NA	NA	NA	NN	NA	DBA	BDUA
6	Affiliate's Department	N	2	NA	NA	NA	NN	NA	DBA	BDUA
7	Affiliate's Municipality	N	5	NA	NA	NA	NN	NA	DBA	BDUA
8	Affiliate's phone number	N	20	NA	NA	NA	NU	NA	NA	NA

**Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports**

ID	Attribute	Value Conformance					Relational Conformance			
		Verification				Validation	Verification			Validation
		DT	LE	FO	DO		ON	PK	FK	
9	Medication CUM code	N	11	NA	NA	NA	NN	NA	NA	INV
10	Name of the medication	T	200	NA	NA	NA	NN	NA	NA	NA
11	Concentration	T	30	NA	E.g1	NA	NN	NA	NA	NA
12	Pharmaceutical form	T	180	NA	E.g2	NA	NN	NA	NA	NA
13	Form of administration	ST	40	NA	E.g3	NA	NN	NA	NA	NA
14	Days of treatment	N	5	NA	NA	NA	NN	NA	NA	NA
15	Prescribed quantity	N	5	NA	NA	NA	NN	NA	NA	NA
16	Delivered quantity	N	5	NA	NA	NA	NU	NA	NA	NA
17	Pending quantity	N	5	NA	NA	NA	NN	NA	NA	NA
18	Main diagnosis	ST	4	NA	NA	NA	NN	NA	CIE	CIE
19	Related diagnosis	ST	4	NA	NA	NA	NU	NA	CIE	CIE
20	Prescription date	D	10	FD	NA	NA	NN	NA	NA	NA
21	Authorization date	D	10	FD	NA	NA	NN	NA	NA	NA
22	Filing date of the medical prescription by the user	D	10	FD	NA	NA	NN	NA	NA	NA
23	Delivery record date	D	10	FD	NA	NA	NU	NA	NA	NA
24	Provider's NIT (without dv)	N	16	NA	NA	NA	NN	NA	DBI	REPS
25	Supplier's company name	T	100	NA	NA	NA	NN	NA	NA	NA
26	Type of prescription delivery, at home or at point of delivery?	I	1	NA	E.g4	NA	NU	NA	NA	NA
27	Satisfactory delivery number	T	50	NA	NA	NA	NU	NA	NA	NA

Nomenclature: (DT) Data Type, (LE) Length, (FO) Format, (DO) Domain, (ON) Obligatory nullability, (PK) Primary Key, (FK) Foreign Key, (ST) Short Text, (N) Numeric, (T) Text, (I) Integer, (D) Date, (NA) Not Applicable, (FD) DD-MM-YYYYY, (NU) Null, (NN) Not Null, (DBA) DB Affiliate EPS, (DBI) DB IPS Contracted EPS, (BDUA, by its Spanish acronym) DB Unique of Affiliates, (INV) DB National Institute for the Safety of Medicines and Food INVIMA, (CIE) DB Diseases CIE10, (REPS, by its Spanish acronym) DB Special Register of Health Service Providers, (E.g1) 200 mg, 300 mg,600mg/100000 i.u., (E.g. 2) Solution for injection, Tablet, (E.g. 3) Oral, Intramuscular, Infiltrative - local, (E.g. 4)1: Domiciliary, 2: Point of delivery, (M) Male, (F) Female.

**Table 6.** Adaptation of the Completeness category - Prescription delivery.

Attribute	Adaptation of Verification	Adaptation of Validation
Delivered quantity	If Delivered quantity is null or equal to 0, the Delivery Record Date must be null.	NA
Delivery record date	If Delivery Record Date is null, the Delivered quantity must be null or zero.	NA
Related diagnosis	If Related diagnosis is null, Main diagnosis must not be null.	NA

**Table 7.** Adaptation of the Uniqueness Plausibility subcategory - Prescription delivery.

Attributes	Adaptation of Verification	Adaptation of Validation
Document type		NA
Identification number		

Medication CUM code	There should be no duplicate records: Document type + Identification number + Medication cum code + Prescription date	
Prescription date		

**Table 8.** Adaptation of the Atemporal Plausibility subcategory - Prescription delivery.

Attributes	Adaptation of Verification	Adaptation of Validation
Document type	Document type + Identification number <b>matches</b> the Affiliates DB.	BDUA
Affiliated Regime	Affiliated Regime + Document type + Identification number <b>matches the</b> Affiliates DB.	BDUA
Affiliated Department and Municipality	Department of the affiliate + Municipality of the affiliate + Document type + Identification number <b>matches the</b> Affiliates database.	BDUA
Days of treatment	There should be no treatment days at 0, single dose medications are managed in one day.	NA
Delivered quantity	Delivered quantity + Pending quantity $\leq$ Prescribed quantity.	NA

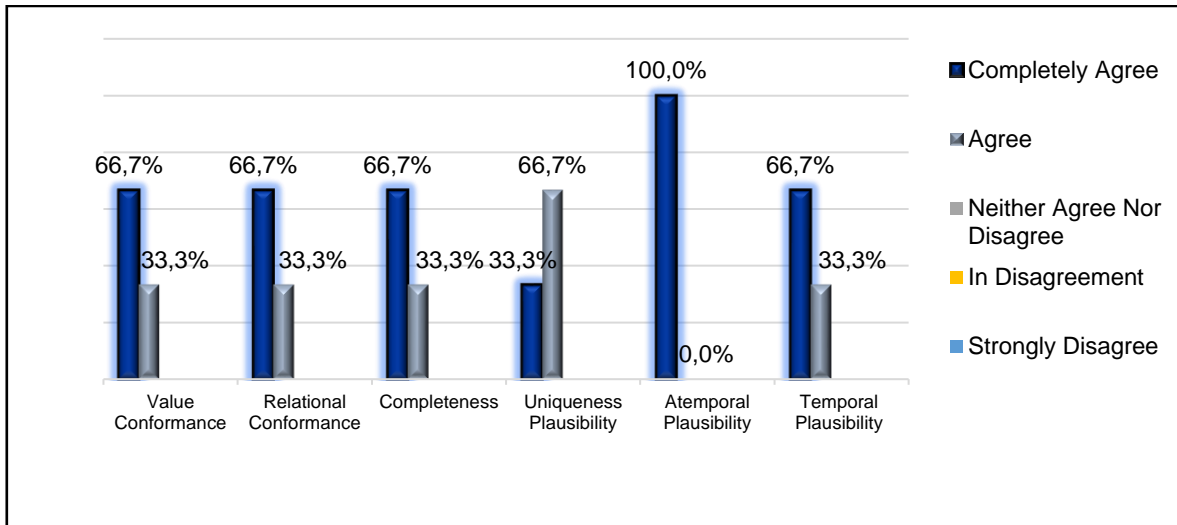
**Table 9.** Adaptation of the Temporal Plausibility subcategory - Prescription delivery.

Attributes	Adaptation of Verification	Adaptation of Validation
Delivery record date	Delivery record date $\geq$ Prescription date.	NA
Date of authorization	Date of authorization $\geq$ Prescription date.	NA
Delivered quantity ( <b>t</b> )	For two records at time <b>t</b> and <b>t-1</b> , which have the same: Document type + Identification number + Medication CUM code + Prescription date. It reviews: Pending quantity ( <b>t-1</b> ) - Delivered quantity ( <b>t</b> ) = Pending quantity ( <b>t</b> )	NA
(Prescription Date + Days of treatment) ( <b>t-1</b> )	For two records at time <b>t</b> and <b>t-1</b> , which have the same: Document type + ID number + Medication cum code + Days of treatment. It reviews: Prescription Date ( <b>t</b> ) > (Prescription Date + Days of treatment) ( <b>t-1</b> )	NA

### ***B. Validation of the Adaptation of the Prescription Delivery Report***

The results obtained from the expert focus group with respect to the adaptation of each subcategory are presented in Fig. 1, according to the Likert scale. The subcategory showing the highest adaptation is Atemporal Plausibility, with 100% in the "Completely Agree" category; followed by *Value Conformance*, *Relational Conformance*, *Completeness*, and *Temporal Plausibility* with 66.7% in "Strongly Agree" and 33.3% in "Agree". In addition, the *Uniqueness Plausibility* subcategory scored 66.7% in "Agree", and 33.3% in "Strongly Agree".

**Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports**



**Fig. 1.** Results of the focus group evaluation of the prescription delivery report.

Table 10 presents one of the experts' comments on the open-ended questions (see Table 4) with its respective improvement actions.

**Table 10.** Comments - Prescription delivery.

Expert Id	Comment	Action for improvement/ Justification
E3	Note that Resolution 521 was repealed at the end of the pandemic.	The attribute Group to which the affiliate belongs is excluded, Res.521 (Group 1,2,3,4, or none).

***C. Adaptation of Data Quality Categories of the Report on Assignment of Medical Appointments***

The adaptation of the *Value Conformance* and *Relational Conformance* subcategories to each attribute, along with the verification and validation assessment strategies, are shown in Table 11; the *Completeness* category in Table 12; the *Uniqueness Plausibility* subcategory in Table 13; the *Atemporal Plausibility* subcategory in Table 14; and *Temporal Plausibility* in Table 15.

**Table 11.** Adaptation of Value Conformance and Relational Conformance - Assignment of medical appointments.

ID	Attribute	Value Conformance					Relational Conformance			
		Verification				Validation	Verification			Validation
		DT	LE	FO	DO		ON	PK	FK	
1	Document type	ST	2	NA	NA	NA	NN	NA	BDA	BDUA
2	Identification Number	N	16	NA	NA	NA	NN	NA	BDA	BDUA
3	Names and last names	T	80	NA	NA	NA	NN	NA	NA	NA
4	Affiliated Regime	ST	12	NA	NA	NA	NN	NA	BDA	BDUA
5	Affiliate Department	N	2	NA	NA	NA	NN	NA	BDA	BDUA
6	Municipality of affiliate	N	5	NA	NA	NA	NN	NA	BDA	BDUA
7	Phone	N	30	NA	NA	NA	NN	NA	NA	NA
8	Date on which the user requests the appointment or surgery	D	10	FD	NA	NA	NN	NA	NA	NA
9	Date on which the user requests the appointment or surgery to be assigned to him/her	D	10	FD	NA	NA	NN	NA	NA	NA
10	Date for which appointment or surgery is assigned	D	10	FD	NA	NA	NN	NA	NA	NA
11	IPS Department	N	2	NA	NA	NA	NN	NA	DBI	REPS
12	IPS Municipality	N	5	NA	NA	NA	NN	NA	DBI	REPS
13	Provider's NIT (without dv)	N	12	NA	NA	NA	NN	NA	DBI	REPS
14	Qualification Code	N	12	NA	NA	NA	NN	NA	DBI	REPS
15	Supplier's company name	T	100	NA	NA	NA	NN	NA	NA	NA
16	Service requested	N	4	NA	NA	NA	NU	NA	DBI	REPS
17	Other specialty or Surgery	T	100	NA	NA	NA	NU	NA	NA	NA
18	Total number of appointments assigned	N	2	NA	NA	NA	NN	NA	NA	NA

Nomenclature: (DT) Data Type, (LE) Length, (FO) Format, (DO) Domain, (ON) Obligatory nullability, (PK) Primary Key, (FK) Foreign Key, (ST) Short Text, (N) Numeric, (T) Text, (I) Integer, (D) Date, (NA) Not Applicable, (FD) DD-MM-YYYY, (NU) Null, (NN) Not Null, (DBA) DB Affiliate EPS, (DBI) DB IPS Contracted EPS, (BDUA) DB Unique of Affiliates, (INV) DB National Institute for the Safety of Medicines and Food INVIMA, (CIE) DB Diseases CIE10, (REPS) DB Special Register of Health Service Providers

**Table 12.** Adaptation of the Completeness category - Assignment of medical appointments.

Attribute	Adaptation of Verification	Adaptation of Validation
Service requested	If the Service requested is null, Other Specialty or Surgery should not be null.	NA
Other specialty or Surgery	If the Other Specialty or Surgery is null, the Service requested should not be null.	NA

**Table 13.** Adaptation of the Uniqueness Plausibility subcategory - Assignment of medical appointments.

Attributes	Adaptation of Verification	Adaptation of Validation
Document type		NA

**Kahn's Data Quality Categories Adaptation for Prescription delivery and Medical Appointment Assignment Reports**

Attributes	Adaptation of Verification	Adaptation of Validation
Identification number	There should not be more than 2 duplicate records: Document type + Identification number + Date for which the appointment or surgery is assigned + Nit provider + Service Requested $\geq 2$	
Date for which the appointment or surgery is assigned		
Provider's NIT		
Service Requested		

**Table 14.** Adaptation of the Atemporal Plausibility subcategory - Assignment of medical appointments.

Attributes	Adaptation of Verification	Adaptation of Validation
Document type	Document type + Identification number <b>matches</b> the Affiliates DB.	Bdua
Affiliated Regime	Affiliated Regime + Document type + Identification number <b>matches the</b> Affiliates DB.	Bdua
Affiliate's Department and Municipality	Department of the affiliate + Municipality of the affiliate + Document type + Identification number <b>matches the</b> Affiliates database.	Bdua

**Table 15.** Adaptation of the Temporal Plausibility subcategory - Assignment of medical appointments.

Attributes	Adaptation Verification	Adaptation Validation
Date for which appointment or surgery is assigned	Date for which appointment or surgery is assigned $\geq$ Date on which the user requests the appointment or surgery.	NA
Date on which the user requests the appointment or surgery to be assigned to him/her	Date on which the user requests the appointment $\geq$ Date on which the user requests the appointment or surgery.	NA

***D. Validation of the Adaptation of the Report on Assignment of Medical Appointments***

The results obtained from the expert focus group regarding the adaptation of each subcategory are presented in

according to the Likert scale. The *Atemporal Plausibility* subcategory shows the highest adaptation, 100% in "Completely Agree"; *Relational Conformance*, *Completeness* and *Temporal Plausibility*, 66.7% in "Strongly Agree", and 33.3% in "Agree". In addition, the *Plausibility of Uniqueness* obtained 66.7% in "Agree" and 33.3% in "Completely Agree".



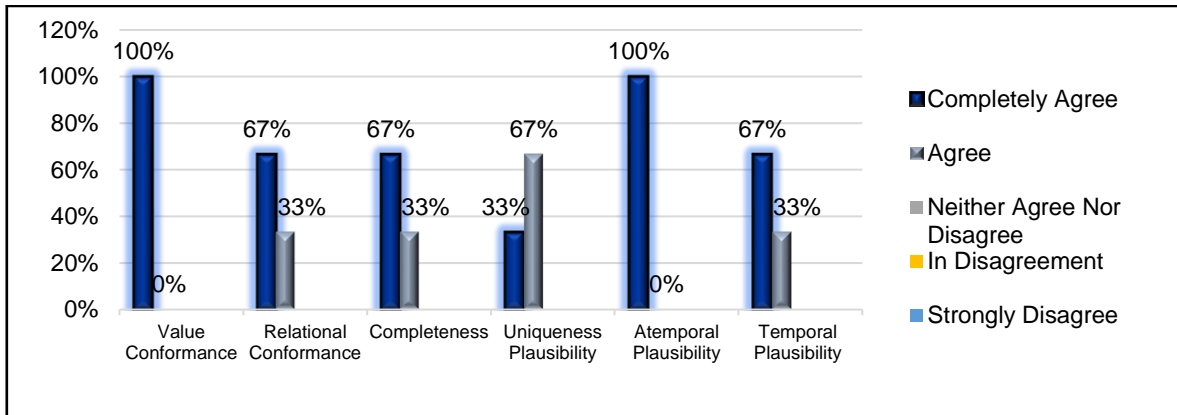


Fig. 2. Results of the focus group evaluation of the Assignment of medical appointments.

Table 16 presents one of the experts' comments on the open-ended questions (see Table 4) with their respective improvement actions.

Table 16. Comments - Assignment of medical appointments

Expert Id	Comment	Action for improvement/ Justification
E1	It is necessary to establish a specific codification for supra-specialties.	Its implementation depends on the IPS since there is no standardized classification by the MSPS.

#### IV. CONCLUSIONS

The adaptation of Kahn's data quality categories for the reports of the prescription delivery and assignment of medical appointments is proposed in this article improves data quality because the criteria defined in the *Conformance* category ensure that the types of data are correct; that the formats, lengths, and domains conform to internal restrictions; and that the relationships between specific attributes of the reports and external sources (BDUA, Invima, CIE10 and REPS) are correctly established. Regarding *Completeness*, it is guaranteed that the absence of data in an attribute is under MSPS regulations; and regarding *Plausibility*, it is ensured that the records are credible and that the quantities in the reports are logical and coherent, in addition to ensuring that each prescription delivery and medical appointment is correctly associated with the corresponding affiliate. In addition, the proposed adaptation of each attribute of the two reports could be applied by the EPS and contribute to obtaining more reliable reports and accurate indicator results, thus promoting more informed decisions in the health sector.

The results of the validation by the health sector experts' focus group indicate that the proposed adaptation meets the needs of data quality for the reports on the prescription delivery and assignment of medical appointments, considering that in all subcategories a percentage of 100% is obtained by adding up "Completely Agree" and "Agree"; which supports the importance of applying these data quality criteria. Considering the experts suggestions, it is proposed that this adaptation include parameters defined by the IPS for the supra-specialties, thus allowing greater control of the data quality (e.g., establishing a maximum number of appointments per supra-specialty). In addition, this adaptation should be incorporated into an ETL process to define a structured data quality flow at each stage, seeking to reduce errors and improve the overall quality of the information managed.

#### **FINANCING**

This research article is derived from the project "Data quality in the ETL process of prescription delivery and assignment of medical appointments for EPSs" funded by the Universidad del Cauca.

#### **AUTHORS' CONTRIBUTION**

**Daisy-Yisel Meneses-López:** Conceptualization, Investigation, Visualization, Writing-original draft.

**Martha-Eliana Mendoza-Becerra:** Conceptualization, Methodology, Supervision, Writing-review & editing.

**Salvador García-López:** Supervision, Writing-review & editing.

#### **ACKNOWLEDGMENTS**

To Universidad del Cauca and Universidad de Granada for their support for the development of this project.

#### **REFERENCES**

- [1] T. Dai, H. Hu, Y. Wan, Q. Chen, Y. Wang, "A data quality management and control framework and model for health decision support," in *12th International Conference on Fuzzy Systems and Knowledge Discovery*, Zhangjiajie, China, 2015, pp. 1792–1796. <https://doi.org/10.1109/FSKD.2015.7382218>

- [2] INVIMA, *ABC Seguridad en el uso de medicamentos*, 2014. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/CA/seguridad-en-la-utilizacion-de-medicamentos.pdf>
- [3] N. Duque, E. Hernández, Á. Pérez, A. Arroyave, D. Espinosa, "Modelo para el proceso de extracción, transformación y carga en bodegas de datos. Una aplicación con datos ambientales," *Ciencia e Ingeniería Neogranadina*, vol. 26, no. 2, pp. 95–109, May 2016. <https://doi.org/10.18359/rcin.1799>
- [4] I. Homayouni, H. Ghosh, S. Ray, "An Approach for Testing the Extract-Transform-Load Process in Data Warehouse Systems," *IEEE International Symposium on Software Reliability Engineering Workshops, Memphis, TN, USA*, pp. 158–161, Oct. 2018. <https://doi.org/10.1109/ISSREW.2018.000-6>
- [5] V. C. Pezoulas *et al.*, "Medical data quality assessment: On the development of an automated framework for medical data curation," *Computers in Biology and Medicine*, vol. 107, pp. 270–283, 2019. <https://doi.org/10.1016/j.compbiomed.2019.03.001>
- [6] H. G. Kahn, G. Michael, T. J. Callahan, J. Barnard, A. Bauck, J. Brown, B. N. Davidson, "Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data," *eGEMs*, vol. 4, no. 1, e1244, Sep. 2016. <https://doi.org/10.13063/2327-9214.1244>
- [7] K. E. Lynch *et al.*, "Incrementally Transforming Electronic Medical Records into the Observational Medical Outcomes Partnership Common Data Model: A Multidimensional Quality Assurance Approach," *Applied Clinical Informatics*, vol. 10, no. 5, pp. 794–803, 2019. <https://doi.org/10.1055/s-0039-1697598>
- [8] H. Spengler, I. Gatz, F. Kohlmayer, K. A. Kuhn, F. Prasser, "Improving Data Quality in Medical Research: A Monitoring Architecture for Clinical and Translational Data Warehouses," in *IEEE 33rd International Symposium on Computer-Based Medical Systems*, Rochester, USA, Jul. 2020, pp. 415–420. <https://doi.org/10.1109/CBMS49503.2020.00085>
- [9] L. G. Qualls *et al.*, "Evaluating Foundational Data Quality in the National Patient-Centered Clinical Research Network (PCORnet®)," *EGEMS (Wash DC)*, vol. 6, no. 1, e3, Apr. 2018. <https://doi.org/10.5334/egems.199>
- [10] M. Petersen, K. Feldt, R. Shahid, M. Mattsson, "Systematic mapping studies in software engineering," in *Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering, Italy*, pp. 68–77, 2008.
- [11] B. Kitchenham, P. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7-15, 2009. <https://doi.org/10.1016/j.infsof.2008.09.009>
- [12] K. Pratt, *Design Patterns for Research Methods: Iterative Field Research*, 2009. [http://www.kpratt.net/wp-content/uploads/2009/01/research\\_methods.pdf](http://www.kpratt.net/wp-content/uploads/2009/01/research_methods.pdf)
- [13] Ministerio de Salud y Protección Social, *Resolución 1604 de 2013*, 2013. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/resolucion-1604-de-2013.pdf>
- [14] Ministerio de Salud y Protección Social, *Resolución 1552 de 2013*, 2013. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/resolucion-1552-de-2013.pdf>