The International Review of Retail, Distribution and Consumer Research

ISSN: 0959-3969 (Print) 1466-4402 (Online) Journal homepage: www.tandfonline.com/journals/rirr20

On the measurement of inventory record inaccuracies

Christoph H. Glock, Aris A. Syntetos & Yacine Rekik

THE INTERNATIONAL REVIEW OF RETAIL, DISTRIBUTION AND CONSUMER RESEARCH
https://doi.org/10.1080/09593969.2025.2608729

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

On the measurement of inventory record inaccuracies

Christoph H. Glock [a], Aris A. Syntetos [b] and Yacine Rekik [c]

[a]Institute of Production and Supply Chain Management, Technical University of Darmstadt, Darmstadt, Germany; [b]PARC Institute of Manufacturing Logistics and Inventory, Cardiff Business School, Cardiff University, Cardiff, UK; [c]ODAI department, Emlyon Business School, Lyon, France

ABSTRACT

Inventory records are often inaccurate, and this is known to be the source of major cost and service inefficiencies in retailing. In today's e-commerce and omni-channel environments, customers increasingly expect real-time visibility of stock availability across locations, making inventory record accuracy a prerequisite for reliable services such as in-store pickup or home delivery. However, the practices of measuring and reporting errors in inventory records vary considerably across the sector. Further, and while the advantages of error-free stock records are apparent, retailers may tolerate minor discrepancies as inconsequential, blurring the line between 'accurate' and 'inaccurate' records. We set out, for the first time, to: i) understand which inventory record inaccuracy (IRI) measures have been proposed so far in the literature, ii) identify how inventory record inaccuracy measurement takes place in retail practice and how it is organizationally embedded, and iii) propose a set of measures to enable benchmarking and continuous improvement. To do so, we first review the pertinent literature using a systematic search and selection method, followed by interviews with 25 retail executives responsible for areas such as inventory loss and prevention. Results obtained from the literature review and the interview study were validated in a workshop involving 46 retail professionals. We find that retailers employ a wide variety of IRI metrics – with simple binary measures being the most common ones – and that there is little consensus on what constitutes an acceptable measure. We propose a set of desirable attributes for the error measures and offer several insights that should be valuable for any retail professional involved in inventory decisions making.

ARTICLE HISTORY
Received 11 May 2025
Accepted 18 December 2025

KEYWORDS
Retail; inventory control; inaccurate inventory records; inventory accuracy; error measurement

Background

Today, inventory management is almost exclusively fully computerised, and as such it relies upon accurate inventory records (i.e. correct information recorded in the inventory system about the quantity of a stock keeping unit (SKU) physically available). Often though (in fact too often) the system records are not reflective of the actual number of units available in stock. Such discrepancies between physical and system inventory are known as Inventory Record Inaccuracies (IRI) and affect a staggering 50%-70% of SKUs at

CONTACT Christoph H. Glock glock@pscm.tu-darmstadt.de

any point in time in retailing (e.g. DeHoratius and Raman 2008; Kang and Gershwin 2005; Rekik, Syntetos, and Glock 2019b). The consequences are dramatic, as positive IRI (i.e. when physical inventory is greater than system inventory) leads to excess inventory, and the opposite leads to stockouts. IRI causes disruptions to all retail environments, but it is especially problematic in buy-online-pick-up-in-store and store-based quick commerce scenarios. In situations where customers rely on quick deliveries, or where they reserve products online to pick them up in store, incorrect inventory records can cause far more damage than in regular retail environments due to higher customer expectations.

Removing IRI has been shown to introduce important cost efficiencies and a sales improvement potential of up to 11% (Rekik et al. 2025). For a comprehensive summary of the causes and implications of IRI, interested readers are referred to Rekik, Syntetos, and Glock (2019b). It follows that the very measurement and reporting of IRI (and the way we perform these tasks) is an important exercise. IRI measures can inform us of the root causes and the magnitude of the problem, guide decisions about intervention strategies (e.g. investment in anti-theft technologies or additional stock audits for problematic departments), and provide us with feedback on the effectiveness of those interventions and their improvements over time. Indeed, they are also used for many other purposes, such as cross-store comparisons and internal benchmarking within a retailer, determination of bonuses, etc. Given the importance of IRI measurement, one might expect it to have been studied extensively; however, that is not the case, and in fact ours is the very first contribution in this area, to systematically evaluate how earlier research – but also how the retail sector – addresses the IRI measurement problem. The only related work we are aware of is that of Shabani et al. (2021) that provided an overview of five IRI metrics used in earlier research, albeit without systematically searching the literature or investigating IRI measurement practices in industry.

Measuring and reporting IRI is not a straightforward exercise. Consider the case of a high-volume selling item (say cans of beans, which sell in the order of hundreds in a typical retail outlet) associated with some relatively insignificant discrepancy (call it five units for the purposes of our discussion). Is this a relevant error that should trigger corrective actions? Perhaps no retailer would (should) ever be interested in such minor discrepancy, which leads to the important question of what constitutes (in)accuracy in the first place and how to determine the boundary between accurate and inaccurate (how inaccurate something needs to be, to be treated as such for operational purposes).[1]

In addition, we argue that IRI measurement is further complicated by important requirements – or core criteria – that need to be met (see Discussion Section for a more detailed discussion of these criteria). Some of these requirements are context-independent. For example, error measures always need to be interpretable and linked to decisions making (though it is still the context that will determine that link). Others are (retail) context-specific and dictate scale independence (to be able to summarise results across SKUs) and symmetry (to be able to avoid asymmetric, over- and under-penalisation, unless intended to). We discuss these issues and propose a set of desirable attributes to facilitate error metric selection and operationalisation.

We employ a mixed-method approach, outlined in the Methodology Section, that consists of three steps. First, we review the pertinent literature using a systematic literature search and selection approach to provide a comprehensive overview of IRI metrics that have been used in prior research to-date. Secondly, we conduct semi-structured interviews with 25 retail executives responsible for such relevant areas as inventory loss and prevention. The results of the literature review and the interviews were validated in an expert workshop with 46 participants from the retail sector.

We find that despite a wide range of IRI metrics proposed in the literature, there is no clear consensus on which measures are most appropriate for specific operational purposes. Retailers, for their part, employ a variety of metrics – often favouring simple binary or range-based approaches – that are not always aligned with academic definitions. Our analysis highlights a critical gap between theory and practice, which we address by proposing a set of desirable metric attributes to support consistent, interpretable, and decision-relevant IRI measurement.

The remainder of our paper is organised as follows: The next section describes the methodology of our research. The third section presents the outcome of the literature review before we discuss our interview findings and resulting insights. The Discussion Section proposes a set of desirable attributes to facilitate errors metric selection and details operationalisation issues in relation to that and managerial implications. The final section concludes this manuscript and discusses natural next steps of research in this area.

Methodology

To address our research objectives, we adopted a three-part mixed-method approach. First, we conducted a systematic literature review of peer-reviewed academic sources to identify and classify existing metrics used to measure IRI. This was followed by a series of semi-structured interviews with 25 retail professionals responsible for areas such as inventory loss and prevention, enabling us to capture current industry practices and perspectives on IRI measurement. Finally, we conducted a validation workshop with 46 experts from the retail sector to refine and confirm our findings. The remainder of this section provides a detailed account of each methodological component, including the design, data collection procedures, and analytical steps.

Literature review

First, we employed a *systematic literature review* that enables us to provide a comprehensive coverage of IRI metrics that have been proposed in earlier pertinent research. We followed Cooper (2010) and Hochrein and Glock (2012) and applied the following steps to identify relevant works: first, we defined a set of keywords including 'inventory record inaccuracy', 'inventory record accuracy', 'inventory accuracy', 'inventory inaccuracy', and 'inventory discrepancy' that we used for searching the scholarly database Scopus. Papers featuring at least one of the keywords in their title, abstract or list of keywords, which were published in English and appeared in a peer-reviewed academic journal qualified for inclusion in the initial literature sample. The abstracts of all papers in the initial sample were read to verify their relevance, and irrelevant works were removed. All works that remained in the sample were then read and screened for the use of a metric[2] that quantifies IRI. Papers were excluded from further consideration if they:

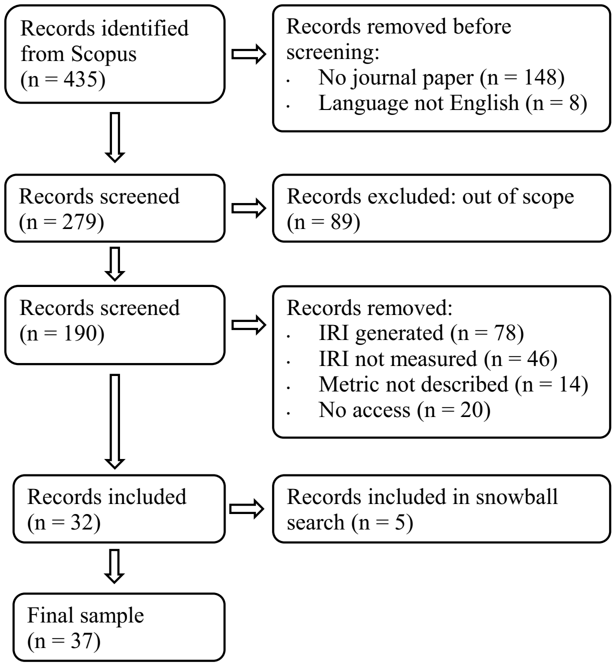

Figure 1. Illustration of the sample generation process.

- propose mathematical models that generate IRI, e.g. by assuming that an error causing discrepancies is normally distributed with mean μ and standard deviation σ, albeit without measuring IRI.
- discuss IRI, albeit without measuring it.
- measure IRI, albeit without providing a clear explanation of how the metric was calculated.
- could not be made available via the university libraries of the three authors of the work at hand and via requests sent to the authors of the papers in question.

In the last step, we applied a snowball search during which we checked the references of all sampled papers, as well as all papers citing one of the sampled papers, for relevance. The final sample consisted of 37 works, with the sample generation process following the PRISMA methodology (Page et al. 2021) illustrated in Figure 1.

Interview study

Following from the literature review, we conducted a series of *expert interviews* to gain insights into retailers' practices in IRI measurement; this was (partly at least) motivated by the scarcity of research on this topic. We opted for semi-structured interviews, in which the informants are asked a series of predetermined but open-ended questions, and where the interview may deviate from this set of questions based on the informants' answers (Grosse et al. 2016). Informants were recruited via the mailing list of the Efficient Consumer Response (ECR) Group, and we can therefore assume that our study relies on informed respondents. Subscribers of this mailing list were informed in a sequence of posts that the researchers were searching for retail experts working in areas such as loss protection, inventory control or inventory auditing to participate in an interview study on IRI metrics. The call was supplemented by directly approaching practitioners in the authors' professional networks. In adding informants to our interview study, we applied both the 'maximum variation principle' and the 'replication logic' (Anand, Gardner, and Morris 2007; Eisenhardt and Graebner 2007). Thus, we tried to include retailers with different characteristics in terms of location, retail segment (e.g. grocery, fashion) and business model (mainly brick-and-mortar vs. online) that may influence the way IRI is handled by the respective companies. In addition, we also made sure that some similar cases were included to be able to verify if an observation made at one retailer could be replicated at another. Informants were added to the study until saturation occurred where further increasing the sample size would likely not lead to new insights (Eisenhardt 1989). Table A1 in the Appendix gives an overview of the interview sample.

All interviews were conducted online via a video conference platform, typically involving two authors of this paper. The interviews ranged from 20 to 70 minutes each. With the informants' consent, all interviews were audio- and video-recorded and subsequently transcribed using Sonix.ai. Following the suggestions of Eisenhardt (1989) and Corbin and Strauss (1990), transcription and coding commenced while the interview study was still in progress. This approach allowed completed interviews to inform subsequent ones. Initial interview questions were formulated in advance but remained open-ended to avoid preconception and to foster a dialogue that encouraged spontaneous and detailed responses (Roulston 2010). The questions aimed at developing our understanding around i) the very conceptualisation of IRI (do organisations talk about accuracy or inaccuracy), ii) the evolution of IRI measurement over time; iii) the sources of relevant data and frequency of IRI calculation, iv) the level of measurement (how granular measurements are), and v) the purpose (say facilitating strategic versus operational decision making). The section discussing the outcome of the interviews is structured along the main sets of questions discussed above. During the interview conversation, follow-up and probing questions were used to gather more information on a relevant topic (see for related approaches Glock et al. 2017; Grosse et al. 2016). This interview method enhanced data validity and provided new insights that the interviewers might not have anticipated (Yin 2018).

Evaluation workshop

Finally, we conducted an evaluation workshop (see, e.g. Thoring, Mueller, and Badke-Schaub 2020) delivered jointly with the ECR Retail Loss Group in March 2025. The workshop involved 46 retail experts with a background in loss prevention, some of whom had participated in the earlier interviews. Results obtained both from the literature review and the interview studies were presented to and discussed with the experts to validate the concepts. That is, the presentation of our preliminary findings was followed by facilitated group discussions, wherein participants provided feedback and additional insights. We used this input to validate our concepts and, where necessary, minor adjustments were made.

This workshop served as a member validation exercise (cf. Thoring, Mueller, and Badke-Schaub 2020), ensuring the practical credibility of our findings and refining our interpretations based on collective expert feedback. By combining a literature review with primary qualitative data analysis and a validation step, our mixed-method approach provides a robust foundation for our analysis of IRI measurement practices.

Results of the literature review

Our systematic literature search identified 37 academic papers that measure IRI. Table A2 in the Appendix provides an overview of all papers contained in our sample and classifies them according to the type of study conducted, the type of metric used, whether the metric was reported in volume (units) or value, and whether the metric was reported in the interviews. We postpone the detailed discussion of the metrics identified in the literature review to the next section, where they are presented along with those identified during the interviews.

A total of 47 IRI metric/paper combinations were found in the sampled papers. Most works considered a single IRI metric, with only a few papers discussing two or three metrics. Table 1 shows the outcome of the literature review. Different types of metrics were discussed in the literature sample:

- Absolute error metrics that disregard the sign of the error (that is, a −2 and a + 2 error will both be treated as +2).
- Signed error metrics that track the direction of the error.
- Percentage (relative) error metrics that express the error (or absolute error) as a percentage of the inventory, where the latter can be either the physical or the system one. This is an issue of great importance and is revisited in the next section.
- Binary metrics that only differentiate between correct and incorrect stock records.

In some cases, the metrics were discussed and applied at the SKU level, while in other cases, researchers aggregated them over multiple SKUs. Binary metrics were found to be most popular, followed by absolute relative metrics. Interestingly, in only 9 out of 47 cases, the metrics enable checking the direction of the error.

In Figure 2 we report the number of IRI publications over time, in non-overlapping 3-year buckets. An interesting trend emerges; there was almost no interest at all in this area until the mid-1980's, a situation that changed in the 90's (with some contributions being offered in the literature) and even more so in the last 15 years. The increasing use of inventory and supply chain software/computerised systems in the 90's, dictated by the

Table 1. Overview of the types of IRI metrics discussed in the literature.

Metric type	#
Binary measurement (correct versus incorrect) expressed as a percentage of (in)correct records over the total number of records (SKUs)	17
Absolute percentage error: absolute error expressed as a percentage of the physical or system inventory	7
Binary measurement (correct versus incorrect) where an acceptable range (say within 1) determines accuracy	6
Signed error: straight difference between the physical and system inventory	5
Signed percentage error: signed error expressed as a percentage of the physical or system inventory	4
Absolute error: no interest in the sign of the error	4
Binary error: correct or incorrect	3
Relative error: ratio of the physical/system inventory	1
TOTAL	**47**

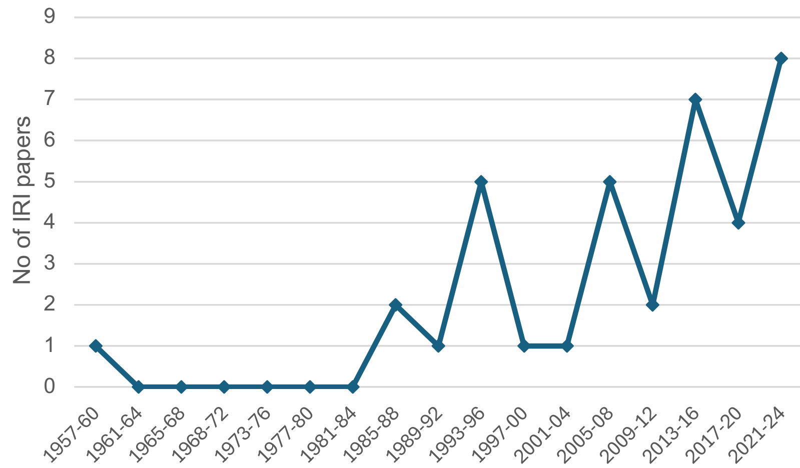

Figure 2. Timeline of IRI publications.

ever-increasing size of stock bases across all sectors, meant that IRI slowly but steadily attracted the attention of academics, still (we argue) not to an extent reflective of the importance and implications of the problem.

Returning to how IRI is measured, our review of the literature highlights that most studies rely on relatively simple, binary measures and often overlook whether system inventories exceed or fall short of actual stock. This lack of directional insight suggests a potential gap between current academic approaches and the more nuanced requirements of real-world inventory management. Consequently, we next turn to our interview study to explore how retailers measure IRI in practice and to determine whether similar patterns – and gaps – emerge.

Results of the interviews

IRI measurement

From the semi-structured interviews, we identified a number of metrics used to measure and report IRI. In the following, we formally present each metric in mathematical terms, offer some explanatory comments as to how the metrics work, and provide a numerical example to demonstrate their use. We also provide a short discussion on each metric to unpack its usefulness or critique some disadvantages. Where the metrics are supported by the academic literature, we provide relevant references. Finally, we discuss, where applicable, important variations of each metric, i.e. other forms it can take (commenting again, where applicable, on relevant references).

Let us introduce the notation used for the remainder of the article before we delve into the metrics:

t: time index – discrete time interval (say a day, week, month) which varies amongst retailers depending on the context

$x^{IS}_{k,t}$ inventory level displayed for SKU k in the information system at the end of time t (when we refer to SKUs, we refer to SKU/store combinations)

$x_{k,t}^{PH}$ inventory level of SKU k physically available in the store at time t

$q_{k,t}$ number of units of SKU k sold in the store within t

α_k^{IS} tolerance level for SKU k with reference to the inventory level in the information system, in percent

β_k^{IS} tolerance for SKU k in units

K number of SKUs available in the store (or that may reflect some other aggregating opportunity, say within a department, a category, or a family of products across which we wish to summarise performance).

Signed errors

Signed errors at the SKU level differentiate between positive and negative inventory discrepancies:

$$IRI_{k,t}^{sign} = x_{k,t}^{PH} - x_{k,t}^{IS}$$

SKU-level signed differences can be very informative, in that negative and positive discrepancies have different implications and, generally, different root causes too. The metric can be aggregated over a set of K SKUs (in the store, or some other desirable level of aggregation):

$$IRI_t^{sign-aggr} = \sum_{k=1}^{K} IRI_{k,t}^{sign} = \sum_{k=1}^{K} \left(x_{k,t}^{PH} - x_{k,t}^{IS} \right)$$

Note that positive and negative inventory discrepancies for different SKUs considered in this metric may cancel each other out. We illustrate how the metric works in an example: Assume for presentation purposes that we have only three SKUs, with $x_{1,t}^{PH} = 12$, $x_{2,t}^{PH} = 49$, $x_{3,t}^{PH} = 99$ and $x_{1,t}^{IS} = 10$, $x_{2,t}^{IS} = 50$, $x_{3,t}^{IS} = 100$. Then $IRI_t^{sign-aggr} = +2 - 1 - 1 = 0$. The example shows that even though none of the stock records are correct, at an aggregated level, the error is zero.

Signed differences were used by Best et al. (2022), DeHoratius et al. (2023), Ishfaq and Raja (2020a, 2020b), and Rekik, Syntetos, and Glock (2019a). In statistical terms, the signed error measure reports 'bias', i.e. a persistent positive or negative discrepancy between the physical and IS inventory levels, across SKUs. So, any number close to zero indicates that, across the board, physical and IS inventories are almost equal, which is not particularly helpful when it comes to action taken on a SKU-by-SKU basis. This would change if we measured it in value, rather than units. Then we would know if the value of the stock we have equals the value we expect to have based on our records; if some units are missing, but some extra ones (that we didn't expect to have) are present, then the net effect may be zero, which is important from an accounting point of view. Also, this is often used in retail companies to proxy their shrink value (if the number returned is negative).

The very way the signed difference is reported is also an important issue. We argue that it is more meaningful and intuitively appealing to consider the physical stock minus the system stock, as then a negative (-) sign implies we have less than we think we do, and a positive (+) implies more. We do appreciate that the way this measurement is carried out varies, and indeed in the academic literature and practice sometimes the signed difference is measured the other way around. Should retailers wish to compare the IRI across different product groups or stores, it is important to clearly establish what is subtracted from what to ensure that the results are indeed comparable.

Finally, there have been some cases reported in the literature (none of which was picked up in the interviews) where the signed difference is 'scaled' by dividing by either the system or physical inventory. Those could be described as *signed percentage errors*. Chaitien and Ramingwong (2024) reported errors as physical minus system inventory levels and divided that by the system inventory. Barratt, Kull, and Sodero (2018) reported signed differences in the same way but divided them by the physical inventory. We will argue later in this section that relative differences should be expressed in relation to physical stock rather than what the records show to have. Kull et al. (2013) reported errors as system minus physical inventory levels and divided that by the physical inventory. They recognised that the latter can sometimes be zero (as more generally this could also apply to system inventory of course, when it is considered in the denominator) in which case they set the error to zero, as otherwise the metric is not defined (we cannot divide by zero). Finally, Shabani et al. (2021) also considered the system minus the physical inventory but divided that by the number of units received (incoming deliveries). We return later to percentage metrics, and various constraints and advantages associated with them; for the time being, it is sufficient to say that expressing errors in relative terms is both intuitively appealing and facilitates summarisations across SKUs.

Binary errors

For a single SKU k, the binary error is calculated as follows:

$$IRI_{k,t}^{bin} = \begin{cases} 0 & if \left| x_{k,t}^{PH} - x_{k,t}^{IS} \right| = 0 \\ 1 & otherwise \end{cases}$$

This metric is 1 if a stock record is found to be inaccurate and 0 otherwise; it does not track the size or direction of the error. For a set of K SKUs, the error is expressed as a percentage, by computing the number of SKUs that have a (positive or negative) discrepancy and by dividing this value by the total number of SKUs considered:

$$IRI_t^{bin-aggr} = \frac{\sum_{k=1}^{K} IRI_{k,t}^{bin}}{K} \times 100$$

As an example, assume there are five SKUs. For three of the five SKUs, the system inventory matches the physical inventory, so $\left| x_{k,t}^{PH} - x_{k,t}^{IS} \right| = 0$ and thus $IRI_{k,t}^{bin} = 0$ for these SKUs. Assume for one SKU, there is one unit missing, and for a second SKU, there is one unit too much. Then for both products, we have $\left| x_{k,t}^{PH} - x_{k,t}^{IS} \right| > 0$ and consequently $IRI_{k,t}^{bin} = 1$. This gives $IRI_t^{bin-aggr} = \frac{0+0+0+1+1}{5} = \frac{2}{5} = 0.4$. So, 40% of the SKUs have an inaccurate stock record in this example, or vice versa, 60% of the SKU records are accurate.

The (relative) binary error has been used by Bertolini et al. (2015), Brooks and Wilson (1995), Chuang, Oliva, and Liu (2016, 2022), Collins et al. (2006), Gemechu et al. (2021), Graff (1987), Gumrukcu, Rosetti, and Buyurgan (2008), Gutesa, Jebena, and Kebede (2024), Ishfaq and Raja (2020a, b), Kang and Gershwin (2005), Meyer (1991), Miller (1997), Millet

(1994), Bernard (1985), Wang, Chen, and Xie (2010), and Wilson (1995). This measure is useful especially for high-level reporting of the percentage of items that are (in)accurate, and indeed it could be safely used for comparison purposes across various departments within a store, or stores of the same retailer, or across retailers, etc. Empirical research has also shown that the binary metric can be powerful for guiding managerial action. Rekik et al. (2025) conducted an experiment in which they investigated how removing inaccuracies from stock records ($IRI_{k,t}^{bin} = 1$) impacts sales, and found that correcting stock records can increase sales by 4 to 11%. For a targeted intervention, knowledge of the size and direction of the errors is crucial though. The binary metric does not convey this information.

Variants of the binary error metric define a permitted error for the stock record. There are two common versions – one expresses the permitted error as a percentage, and the other one in units.

If the error range is expressed as a percentage, then the metric is calculated as follows:

$$IRI_{k,t}^{range,\%} = \begin{cases} 0 & if \left| x_{k,t}^{PH} - x_{k,t}^{IS} \right| \leq \alpha_k^{IS} \cdot x_{k,t}^{IS} \\ 1 & otherwise \end{cases}$$

The permitted error for the stock records is defined here as a fraction of the inventory displayed in the system ($\alpha_k^{IS} \cdot x_{k,t}^{IS}$). If the error does not exceed this permitted tolerance, the stock record is considered correct ($IRI_{k,t}^{range,\%} = 0$) and otherwise incorrect ($IRI_{k,t}^{range,\%} = 1$).

As an example, assume the inventory system displays 10 units of SKU k ($x_{k,t}^{IS} = 10$), and the physical inventory is 11 units (case 1: $x_{k,t}^{PH} = 11$) or 9 units (case 2: $x_{k,t}^{PH} = 9$). The tolerance level is 10% ($\alpha_k^{IS} = 0.1$). In both cases, the stock record is considered accurate (case 1: $|11 - 10| = 1 \leq 0.1 \cdot 10 = 1$; case 2: $|9 - 10| = 1 \leq 0.1 \cdot 10 = 1$).

This error range was used by Miller (1997) and Rinehart (1960). Even though the use of the system inventory level $x_{k,t}^{IS}$ seems to be popular in calculating the tolerance level both in the literature and industry, we argue that the tolerance level should be determined based on $x_{k,t}^{PH}$, since we are interested in characterising the error based on what we have in stock, *not* what we think we have in stock!

The second variant expresses the permitted error for the stock record in units. In this case, the metric is calculated as follows:

$$IRI_{k,t}^{range,units} = \begin{cases} 0 & if \left| x_{k,t}^{PH} - x_{k,t}^{IS} \right| \leq \beta_k^{IS} \\ 1 & otherwise \end{cases}$$

Again, if the error does not exceed this permitted tolerance, the stock record is considered correct ($IRI_{k,t}^{range,units} = 0$) and otherwise incorrect ($IRI_{k,t}^{range,units} = 1$).

Consider an example where the inventory system displays 10 units of SKU k ($x_{k,t}^{IS} = 10$), the physical inventory is 11 units (case 1: $x_{k,t}^{PH} = 11$) or 9 units (case 2: $x_{k,t}^{PH} = 9$) and the tolerance is 2 units ($\beta_k^{IS} = 2$). In both cases, the stock record is considered accurate (case 1: $|11 - 10| = 1 \leq 2$; case 2: $|9 - 10| = 1 \leq 2$).

This error range was used by Hardgrave, Aloysius, and Goyal (2009), Kang and Gershwin (2005), and Sheppard and Brown (1993). An obvious concern raised with regards to this metric is the arbitrary determination of the permitted tolerance. The same applies to the error range expressed as a percentage, but much less so, since its relative nature makes it more 'trustworthy'. If we say that we always accept a 20% discrepancy, this representatively applies to a SKU with an average (physical or system) stock of say 10 units and one with 100 units, as in the first case we are looking at a 2 units-tolerance but in the second case a 20-units one. In some cases, say in DIY retail, where we keep a very wide range of items with a relatively small inventory for each (about 10 units), setting a unit-based tolerance, e.g. 1 unit, does make sense, as in the majority of cases this translates to something reasonable, in this case 10%. Alternatively, using some ABC-type classification to assign different tolerances to SKUs of different importance (nature) also makes sense and should facilitate some effective error management (Miller 1997; Paul 1985). But otherwise, (universal) unit-based tolerances are problematic and should be avoided.

Another form this metric can take is a value-based one. If the monetary difference between actual and system inventory is larger than some pre-defined cut-off value, then the record is deemed inaccurate; else, accurate (Rinehart 1960).

Absolute error measures

For a single SKU k, the absolute error is calculated as follows:

$$IRI_{k,t}^{abs} = \left| x_{k,t}^{PH} - x_{k,t}^{IS} \right|$$

For a set of K SKUs, the absolute error can be aggregated over all SKUs (say in a store):

$$IRI_{t}^{abs-aggr} = \sum_{k=1}^{K} \left| x_{k,t}^{PH} - x_{k,t}^{IS} \right|$$

This metric does not differentiate between positive and negative inventory discrepancies but just tracks the absolute difference between the physical and system inventory. When summing up the error over all SKUs, positive and negative discrepancies do not cancel each other out.

Consider an example with only three SKUs, with $x_{1,t}^{PH} = 12$, $x_{2,t}^{PH} = 49$, $x_{3,t}^{PH} = 99$ and $x_{1,t}^{IS} = 10$, $x_{2,t}^{IS} = 50$, $x_{3,t}^{IS} = 100$. Then $IRI_{t}^{abs-aggr} = 2 + 1 + 1 = 4$.

The absolute error measure was used by DeHoratius and Raman (2008), Goyal et al. (2016), Hardgrave, Aloysius, and Goyal (2013), and Raman, DeHoratius, and Ton (2001). The aggregation of the absolute error across SKUs returns something more meaningful than aggregating based on signed errors. However, this comes with an implicit assumption that the direction of the error is not important, further implying some symmetric loss function (i.e. the loss or negative consequences of positive errors are the same with those of negative ones).

Absolute percentage errors (APE)

At the SKU level, the metric is defined as follows:

$$IRI_{k,t}^{APE} = \frac{\left| x_{k,t}^{PH} - x_{k,t}^{IS} \right|}{x_{k,t}^{IS}} \quad 100$$

It first calculates the absolute size of the error, and then it divides it by the inventory level recorded in the system to obtain the absolute percentage error. Note that this metric is only defined for SKUs with a positive system inventory, i.e. $x_{k,t}^{IS} > 0$.

As an example, assume that the inventory system displays 4 units of SKU k ($x_{k,t}^{IS} = 4$), and the physical inventory is 6 units (case 1: $x_{k,t}^{PH} = 6$) or 2 units (case 2: $x_{k,t}^{PH} = 2$). In both cases, the error is two units (case 1: $|6 - 4| = 2$; case 2: $|2 - 4| = 2$), and the absolute percentage error is 0.5 or 50%.

This metric was used by Ernst, Guerrero, and Roshwalb (1993).

For an across-SKUs use of the metric, the Mean Absolute Percentage Error (MAPE) is defined as follows:

$$IRI_{t}^{MAPE} = \frac{\sum_{k=1}^{K} IRI_{k,t}^{APE}}{K} \times 100$$

This metric simply averages the APEs of a number of SKUs to calculate a statistic that captures collective rather than individual IRI. Its relative nature makes it an ideal candidate for reporting inaccuracies across many SKUs.

For presentation purposes, assume that we have only two SKUs, with $x_{1,t}^{PH} = 6$, $x_{2,t}^{PH} = 4$, and $x_{1,t}^{IS} = 4$, $x_{2,t}^{IS} = 2$. Then, $IRI_{1,t}^{APE} = 50\%$ and $IRI_{2,t}^{APE} = 50\%$. The $IRI_{t}^{MAPE} = \frac{50\% + 50\%}{2} = 50\%$.

There are four important points to make regarding both the APE and the MAPE.

- First, we argue that using the IS inventory level as the reference value (i.e. in the denominator) is logically flawed. Discrepancies (errors) should be established in relation to reality (i.e. what we actually have in stock) as opposed to expectations (i.e. what we think we have in stock).
- Second, and as discussed above, if the system inventory is zero then the ratio cannot be defined, and this calculation should be omitted. Or alternatively, and when we want to summarise things across SKUs, to obtain the Mean APE (MAPE), we may proceed as follows. Instead of getting the average of all absolute percentage errors, we can get the absolute difference of the sum of the physical inventory levels minus the sum of the system inventory levels, and we divide that by the sum of the system inventory levels. That would get us around the problem.
- Third, this measure suffers from a statistical drawback: it is asymmetric. Assume the inventory system displays 4 units of SKU k ($x_{k,t}^{IS} = 4$), and the physical inventory is 6 units ($x_{k,t}^{PH} = 6$). As discussed above, the absolute error then is two units ($|6 - 4| = 2$), and the absolute percentage error is $2/4 = 0.5$ or 50%. Let us now swap the values reported here and assume that the inventory system displays 6 units of SKU k ($x_{k,t}^{IS} = 6$), and the physical inventory is 4 units ($x_{k,t}^{PH} = 4$). In this case, the absolute error is still two units ($|4 - 6| = 2$), but the absolute percentage error is $2/6 \approx 0.33$ or 33%. That is, the APE over-penalizes positive than negative errors (see, e.g. Boylan and Syntetos 2021).
- We should also note that this metric can take other forms, notably (and as discussed above) having the physical inventory level in the denominator (Bruccoleri, Cannella,

and La Porta 2014; Destro et al. 2023). Or alternatively, some average (across time periods) inventory level may appear in the denominator. This may be either some physical inventory average (Fleisch and Tellkamp 2005) or system-related one (DeHoratious et al., 2023). Admittedly, these metric variations relate to simulation-based academic studies where reporting averages across time makes sense. And the same is true when reporting the unit of simulation time in the denominator (as in Chuang and Oliva 2015). These variations wouldn't necessarily reflect industry applications.

A related metric is the symmetric Absolute percentage error (sAPE). At the SKU level, it is defined as follows:

$$IRI_{k,t}^{sAPE} = \frac{\left| x_{k,t}^{PH} - x_{k,t}^{IS} \right|}{\frac{x_{k,t}^{PH} + x_{k,t}^{IS}}{2}} \times 100$$

It first calculates the absolute size of the error at the SKU level, which is then divided by the average of the physical inventory available in stock and the inventory level recorded in the system to obtain the symmetric absolute percentage error.

For example, assume we have two SKUs. For the first SKU, the inventory system displays 4 units ($x_{1,t}^{IS} = 4$), and the physical inventory is 6 units ($x_{1,t}^{PH} = 6$). For the second SKU, the inventory system displays 6 units ($x_{s,t}^{IS} = 6$), and the physical inventory is 4 units ($x_{2,t}^{PH} = 4$). Although the absolute error is two units in both cases, the APE would be, as previously discussed, 50% and 33% for the first and second SKU, respectively. However, the sAPE would be $\frac{2}{5} = 40\%$ in both cases.

Note that unlike the $IRI_{k,t}^{APE}$ metric that could not be defined for SKUs with zero system inventory, i.e. $x_{k,t}^{IS} = 0$, this metric is more flexible and can accommodate zero physical or system inventory levels. In addition, it also corrects for the asymmetry problem of APEs. Interestingly though, it introduces another type of asymmetry that is likely to be even more important than the one it corrects (Goodwin and Lawton 1999). Assume that the inventory system displays 4 units of SKU k ($x_{k,t}^{IS} = 4$), and the physical inventory is 6 units (case 1: $x_{k,t}^{PH} = 6$) or 2 units (case 2: $x_{k,t}^{PH} = 2$). In both cases, the error is two units (case 1: $|6 - 4| = 2$; case 2: $|2 - 4| = 2$), but the symmetric absolute percentage error would be $\frac{2}{5} = 40\%$ and $\frac{2}{3} = 66\%$, for the first and second case, respectively. So, this error metric over-penalises negative errors. Please note that in both cases the APE would be 50%, so it would not be affected at all by this asymmetry.

For summarisations across SKUs, the symmetric Mean Absolute percentage error (sMAPE) is defined as follows:

$$IRI_{t}^{sMAPE} = \frac{\sum_{k=1}^{K} IRI_{k,t}^{sAPE}}{K} \times 100$$

This metric simply averages the sAPEs of a number of SKUs to calculate a statistic that captures collective rather than individual IRI.

For example, assume we have two SKUs. For the first SKU, the inventory system displays 4 units ($x^{IS}_{1,t} = 4$), and the physical inventory is 6 units ($x^{PH}_{1,t} = 6$). For the second SKU, the inventory system displays 6 units ($x^{IS}_{s,t} = 6$), and the physical inventory is 4 units ($x^{PH}_{2,t} = 4$). The sAPE would be $\frac{2}{5} = 40\%$ in both cases. And the $IRI^{sMAPE}_{t} = \frac{40\%+40\%}{2} = 40\%$.

The fact that sMAPE uses both the physical and system inventory level in the denominator is rather unnatural, but its motivation was the correction of the statistical problem reported for the APEs rather than its intuitive appeal. However, and as numerically explained above, this measure suffers from a statistical (asymmetry) problem that is likely to be more important and detrimental than the one it corrects. We argue that this metric lacks both intuitive appeal and statistical rigour and we suggest it is not used in practice.

Shrink

Shrink calculates the signed difference between the inventory level displayed in the system and the physical inventory available for a particular SKU. Shrink emphasises loss and subtracts the physical inventory from the system inventory. If the difference is negative, it is omitted and this SKU is not considered in this calculation. But if it is positive, it is recorded and expressed as a percentage of the sales (in units) over some predefined time interval (say a week, or the time elapsed since the last stock count/reconciliation):

$$IRI^{shrink}_{k,t} = \frac{\left(x^{IS}_{k,t} - x^{PH}_{k,t}\right)^{+}}{q_{k,t}} \times 100$$

As an example, assume that the inventory system displays 10 units of SKU k ($x^{IS}_{k,t} = 10$), and the physical inventory is 11 units (case 1: $x^{PH}_{k,t} = 11$) or 9 units (case 2: $x^{PH}_{k,t} = 9$), with weekly sales (in units) for both cases being $q_{k,t} = 10$. In the first case, the error is $10 - 11 = -1$, implying that there is one unit more in stock than the inventory system thinks, and this is not taken any further, as we are only interested in shrink. In the second case, the error is $10 - 9 = 1$, which indicates that one unit got lost. And the $IRI^{shrink}_{k,t} = \frac{1}{10} = 10\%$.

This metric was used by DeHoratius and Raman (2008). There is a great debate in the practitioner community as to whether shrink and IRI are the same thing. We argue they are not, as the former refers explicitly to loss whereas the latter is inclusive of both possibilities (i.e. loss and extra stock). An alternative calculation we have come across implicitly assumes that all (or most) of the errors will be positive, thus ignoring the (according to those who advocate this theory: few) cases where the error is negative. In this case, the metric becomes:

$$IRI^{shrink}_{k,t} = \frac{\left(x^{IS}_{k,t} - x^{PH}_{k,t}\right)}{q_{k,t}} \times 100$$

Summarisation of the above across SKUs would indicate the shrink over a collection of items:

$$IRI_t^{shrink-aggr} = \frac{\sum_{k=1}^{K} \frac{\left(x_{k,t}^{IS} - x_{k,t}^{PH}\right)}{q_{k,t}}}{K} \times 100$$

We argue that extra stock is all too often the case, especially in situations/stores with large backrooms where stock is accumulated and unaccounted for. In these cases, the calculation above makes sense but only in as far as we report aggregate performance.

Before we close this section, another summary statistic that we have come across to report shrink in practice is the following:

$$IRI_t^{shrink-aggr} = \frac{\sum_{k=1}^{K} \left(x_{k,t}^{IS} - x_{k,t}^{PH}\right)}{\sum_{k=1}^{K} q_{k,t}} \times 100$$

In this case, the aggregate signed error is divided by the aggregate (across SKUs) sales to determine the shrink. Again, this is put forward on the assumption that we predominantly report SKUs with physical inventories being less than the system related ones, which, our experience tells us, is not the norm!

In summary, the interviews revealed a diverse set of IRI metrics, many of which align with those in the literature. Each has strengths and limitations, underscoring the need for careful selection of metrics – a topic we address in the last section of the paper. Beyond the metrics themselves, our interviews also uncovered several themes regarding the context and application of IRI measurement in retail organizations, which we detail next.

IRI Management

In addition to identifying and evaluating the types of metrics used to measure IRI, our interviews with retail professionals revealed several critical aspects of how these metrics are embedded within the broader context of retail operations. These insights provide a deeper understanding of the perception, evolution, practical application, and organizational implications of IRI measurement in industry.

Conceptualization of IRI: inaccuracy vs. accuracy

To explore how inventory record issues are conceptualized in practice, we conducted a content analysis of the interview transcripts to determine whether participants more frequently referred to *inventory record accuracy* (IRA) or *inventory record inaccuracy* (IRI). Although closely related, these framings carry distinct implications. References to *accuracy* typically signal aspirational or goal-oriented perspectives and focus on performance targets. In contrast, references to *inaccuracy* are more diagnostic or problem-oriented, often associated with operational failures or shrink-related issues. Understanding this framing is important, as it may reflect deeper cultural and organizational attitudes toward inventory control and management.

We performed a systematic keyword-based content analysis on the full set of 25 interview transcripts as described in Hsieh and Shannon (2005). First, we developed a keyword list for each category. For the *accuracy-related* framing, we searched for occurrences of the following terms and variants: 'accuracy,' 'accurate stock,' 'inventory accuracy,' 'stock integrity,' and 'correct stock records.' For the *inaccuracy-related* framing, we searched for the terms 'inaccuracy,' 'discrepancy,' 'error,' 'mismatch,' 'shrink,' and 'stock record issues.'

The transcripts were imported into a text analysis tool, where keyword searches were performed. Each match was reviewed in its immediate context to verify relevance and avoid false positives. Statements that were ambiguous or did not clearly indicate a preference toward either framing were excluded from the final tally. In total, 203 relevant references were identified across all transcripts. Each reference was coded as either IRI-related or IRA-related, based on the surrounding phrasing and intended meaning.

Out of the 203 coded instances:

- 126 references (62%) framed the issue in terms of *inaccuracy* (IRI). These typically emphasized problems such as inventory mismatches, shrink, or stock discrepancies, often linking them to audit outcomes, supply chain disruptions, or store-level errors.
- 69 references (34%) framed the issue in terms of *accuracy* (IRA). These included goals for maintaining accurate records, performance targets (e.g. achieving 95% accuracy), or aspirations for improvement in stock visibility and control.
- 8 references (4%) were categorized as *ambiguous or neutral*, where the language used (e.g. 'file correctness') did not clearly favour either framing and lacked sufficient context to assign definitively.

The inaccuracy-oriented framing was particularly common in interviews with participants from loss prevention, auditing, and operations roles. These individuals often discussed IRI as a barrier to performance, emphasizing the need to identify, report, and resolve errors. In contrast, accuracy-oriented framing appeared more frequently in discussions of training, system performance, and long-term strategic improvement initiatives.

This finding echoes how terminology can reflect organizational focus – a firm that talks of 'accuracy' might be target-driven, whereas one that speaks of 'inaccuracy' is issue-driven. Recognizing this dichotomy is important for benchmarking: an 'inventory accuracy of 95%' might sound equivalent to '5% inaccuracy rate,' but psychologically they emphasize different things.

Evolution of IRI measurement practices

A recurring theme among interviewees was that IRI measurement practices have evolved significantly over the past three to five years, with several respondents expecting continued changes in the near future. Key developments over the last years include:

- Growing awareness of IRI's relevance and direct impact on product availability.
- The shift from simple metrics (e.g. binary ones) to more nuanced (e.g. range-based) indicators.
- Implementation of improved IT systems that facilitate data access and analysis. This includes the use of cloud-based platforms to enable real-time evaluations and centralized reporting as well as the use of RFID to enable frequent stock audits.
- Changes in the procedural setup of IRI measurement workflows.

- Integration of external vendor data to enrich and validate IRI metrics.
- Increases in staffing for teams focused on inventory record accuracy.

While a minority of retailers reported no significant change or even reductions in staffing in pertinent areas, the overwhelming consensus was that the maturity and sophistication of IRI measurement have improved as a direct consequence of increased resources to deal with this issue.

Sources of data and frequency of IRI calculation

Retailers rely on a variety of stock audit processes to generate the data needed for calculating IRI metrics. Interviewees reported several distinct types of audits:

- Wall-to-wall stock audits, in which the entire store inventory is counted. This is typically performed once per year, often in a rolling fashion such that different stores are audited weekly. In cases where retailers are equipped with RFID technology, full-store audits can be performed more frequently.
- Full category audits, focused on specific product groups such as fresh or high-value items. These are generally conducted more frequently, e.g. on a monthly basis.
- Cycle counts, which involve counting selected SKUs or categories on a regular basis (e.g. weekly or monthly), with the goal of covering the entire assortment over time.
- Targeted audits, where SKUs suspected to have inaccurate stock records are selected for counting.
- Gap scans, typically carried out daily or weekly, that identify and investigate shelf stockouts.

Wall-to-wall audits are often performed by external service providers or central auditing teams, while the other types of audits are usually conducted by store personnel. In addition, some retailers use stock management systems that track inventory movements (e.g. sales, deliveries, internal transfers) in real time. These systems can serve as a continuous data source for IRI calculation and allow for on-demand metric updates.

In most cases, IRI metrics are calculated following new data input, such as after a cycle count. Some retailers reported that IRI calculation aligns with internal reporting schedules, such as when preparing financial or operational reports for senior management. In these cases, the IRI metrics are calculated using the most recent data.

Levels of IRI measurement

Another important insight relates to the granularity at which IRI is measured. Interviewees described a variety of measurement levels, from highly aggregated views to detailed SKU-level analysis. Several organizations use a hierarchical approach to IRI measurement, starting from the SKU level and rolling up to higher levels such as category, department, or even category group (the numbers in brackets link the quote to the interviewee in Table A1):

"We use a category structure for our products. When we are measuring IRI, we start with the SKU level. Then it goes up to a class, then to a sub-department, then to a department, then to a category, and then to a category group." [#3]

The level of analysis often depends on the intended use of the metric. For instance, one interviewee noted:

"For financial purposes, we use aggregates, but for replenishment purposes, we are interested in the SKU level. On an aggregated level, it could be that half of the articles have too much of stock and the other half too little, and they just balance each other out. For financial reasons, this is okay, but financial correction is not a good indication of how good the stock on article level is. For replenishment purposes, we measure always on article branch level." [#19]

Others emphasized a top-down approach, starting with aggregated metrics at the company or regional level and drilling down to more granular detail as needed:

"We start off at an aggregated level. You look at a store, store by geography, look at the company level. We look at a regional level, an area level, a store level. We can get down to SKU level detail as well. But generally, you would look at it first and foremost at a company level. And then you drill down into probably a category level." [#13]

Our results indicate that many retailers use a hierarchy of analysis: many start at the SKU level and aggregate upward for reporting, while others begin with an overall indicator and drill down to pinpoint problem areas. The choice often depends on the metric's purpose (as the next section shows). This flexibility in granularity underscores the multifunctionality of IRI metrics and the importance of tailoring them to the context in which they are used.

Purpose of IRI metrics

Interviewees cited a wide range of purposes for calculating IRI metrics, which clustered into three overarching areas.

The first area concerns *strategic and financial oversight*. Here, IRI metrics support company-level financial management by quantifying shrink, aligning book stock values with physical inventory, forecasting lost sales, and clarifying the extent of product cannibalization.

A second area relates to *operational management and process improvement*, where IRI metrics inform efforts to improve internal routines and day-to-day operations. Interviewees explained that such measures help identify the root causes of inaccuracies, support preventative actions, assess the effectiveness of store-level procedures, and highlight where further refinement is needed. In addition, they are essential for improving the accuracy of auto-replenishment systems and for prioritising or triggering targeted stock audits.

The third area involves *behavioural influence and accountability*. In this domain, IRI metrics act as tools for shaping behaviours across the organisation and its partners. They can encourage better inventory practices among suppliers and distribution centres, and they provide input for performance-based incentive systems – such as leaderboards or financial rewards for high-performing stores.

Overall, these findings show that IRI metrics extend well beyond a purely diagnostic role. They contribute to strategic decision-making, operational improvement, and behavioural alignment across the retail system.

Challenges and resource constraints

Despite advancements that have been made over recent years, several interviewees emphasized that analytical capacity remains a bottleneck. Even with increasingly rich data sources and reporting tools, many organizations lack the human resources necessary to fully exploit this information. This highlights an ongoing need to balance technological capability with analytical support in order to maximize the value derived from IRI metrics.

Together, these insights shed light on how IRI metrics are operationalized in retail settings and underscore the importance of continuous investment in systems, processes, and people to effectively manage inventory record accuracy.

Overall, the interviews revealed not only which metrics are used in practice but also how they are implemented and evolved within organizations. These insights, combined with the literature review, inform the development of a set of desirable attributes for evaluating IRI metrics, which we turn to next.

Discussion

Our study revealed substantial heterogeneity in how IRI is measured in both academic literature and retail practice. While numerous metrics are available – ranging from binary accuracy indicators to various forms of relative error – their properties, interpretability, and applicability differ widely. This variability poses challenges for internal benchmarking, cross-firm comparison, and the implementation of systematic improvement strategies. In response, we propose some guidelines for evaluating and selecting IRI metrics based on a set of desirable attributes. These criteria are designed to ensure that the chosen metrics are not only methodologically sound but also practically relevant and decision oriented.

Criteria for evaluating IRI metrics

We identify four core criteria that robust IRI metrics should satisfy:

- Scale independence: Inventory management typically involves aggregating results across hundreds or thousands of SKUs. Metrics must therefore be scale-independent – able to be aggregated or compared across SKUs of different volumes. Binary indicators of accuracy (e.g. percentage of accurate SKUs) and relative metrics such as MAPE meet this criterion, whereas raw unit-based metrics (e.g. total absolute error) do not, unless normalized.
- Symmetry: Many metrics either implicitly or explicitly prioritize one direction of error over another – often focusing on shrinkage (system inventory exceeds physical stock). However, positive IRI (physical stock exceeds system records) can be equally problematic, especially for stores with large backrooms (where inventory is stored but not visible). Symmetric metrics treat overstatements and understatements with equal weight, ensuring balanced diagnostics. Measures like the MAPE and sMAPE both suffer from different types of asymmetries.
- Interpretability: For metrics to be actionable, they must be easily understood by decision-makers. This includes clarity in what is being measured (e.g. magnitude vs. direction of error) and what constitutes good or poor performance. For instance,

Table 2. Horses for courses! Examples.

Use Case	Recommended Metrics
Replenishment Optimization	Signed Error
Financial Reporting	Shrink %, Aggregate Absolute Error (in value)
Store Performance Benchmarking	% Accurate SKUs
Operational Root Cause Analysis	Signed Error, Error Distribution per SKU[3]

a store manager is more likely to act on a 30% inaccuracy rate than a signed net error of −12, unless the latter is contextualized.

- Decision linkage: Metrics must support decision-making. This includes aligning with operational, financial, or strategic objectives. For example, a shrink percentage tied to sales volume may be relevant for financial reporting but inadequate for replenishment decisions, which depend on SKU-level accuracy. A good metric should clearly indicate what corrective action is needed, where, and to what extent.

These criteria are not mutually exclusive but should be viewed as a set of complementary principles. Few metrics meet all criteria fully, but reference to those criteria allows practitioners and researchers to make informed trade-offs based on context.

Practical implications and metric selection guidance

Our findings indicate that many retailers rely on binary or range-based metrics due to their simplicity and ease of communication. While these are valuable for high-level reporting, they often obscure the scale and direction of the problem, making them less useful for targeted interventions. Conversely, more advanced percentage-based metrics offer deeper insight but may be difficult to interpret or aggregate if not carefully applied.

Based on our findings, we recommend that practitioners adopt a 'composite approach' to IRI measurement – using at least two complementary metrics. For example, combining a binary accuracy rate (percentage of SKUs deemed accurate) with a mean absolute percentage error (MAPE) can provide both breadth and depth: the former enables cross-store comparisons, while the latter captures the severity of errors (albeit in an asymmetric way).

Furthermore, our findings highlight the importance of aligning metrics with their intended purpose. If the aim is to support replenishment systems, SKU-level directional errors (e.g. signed error or shrink percentage) are essential. For financial oversight, aggregated value-based metrics or shrink expressed in values may be more appropriate. Table 2 outlines common use cases and suggests suitable metrics.

Theoretical contributions and research implications

The proposed set of criteria/guidelines advances the literature by providing the first structured evaluation of IRI metrics grounded in both scholarly research and industry practice. It also serves as a diagnostic lens to critically assess the suitability of existing measures. Previous studies have tended to adopt metrics without detailed justification or consideration of their operational consequences. Our criteria offer a common language for researchers and practitioners to navigate metric selection more deliberately.

This work opens several avenues for future research. First, empirical studies could evaluate the impact of different IRI metrics on operational outcomes, such as availability, sales, or shrinkage. Second, design science approaches could be used to co-develop decision-support systems that incorporate multi-metric dashboards based on our results. Finally, as retail environments become increasingly data-driven, understanding how real-time technologies (e.g. RFID, IoT) influence the choice and performance of IRI metrics is an important next step.

Notes

1. Please note that our discussion emphasises inventory management and relevant considerations. For example, the five missing units do play a role for accounting and taxation purposes, no doubt. So, strictly speaking, our discussion is not about whether an error qualifies as such, but whether we should adopt any (cost-intensive) corrective actions in response to that. Please also note that actually all retailers do measure IRI by law, but they don't necessarily do it thinking about accuracy but rather because they want to know how much they have lost to theft and other causes over the course of the year versus what they thought they should have. This also ensures they have an accurate record of the actual value of the inventory they hold to satisfy the external auditors.
2. Please note that the words 'metric' and 'measure' are used interchangeably for the purposes of this article.
3. The value and relevance of examining the distribution of errors has not been considered thus far in the paper, but such a distribution (say a relative frequency diagram) is an obvious and natural visual descriptive summarisation of the errors to inform and guide root cause analysis.

Acknowledgments

We would like to thank Efficient Consumer Response (ECR) Retail Loss for supporting this research, Colin Peacock for his inputs to the conceptualization of this research, and the entire retail community for so enthusiastically contributing to our work. We would also like to thank Rogelio Oliva from Texas A&M for his constructive comments on an earlier version of the paper. Finally, we are grateful to the anonymous reviewers for their comments that helped to improve our paper and its presentation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Christoph H. Glock is a Professor of Production and Supply Chain Management at Technical University of Darmstadt, Germany. His research interests include inventory management, supply chain management, warehousing, retail operations management, and human factors in logistics and inventory systems. His work has appeared in various international journals, such as IISE Transactions, Decision Sciences, OR Spectrum, the Journal of Business Logistics or the European Journal of Operational Research.

Aris A. Syntetos is Professor of Decision Science at Cardiff Business School, where he researches inventory-forecasting, -optimisation, -management and -governance. He holds the DSV Chair of Manufacturing and Logistics, and is Director of the joint Cardiff University-DSV PARC Institute that resides in the University. Many of Aris' algorithms feature in best selling supply chain software packages. He is the Editor In Chief of the IMA Journal of Management Mathematics (IMAMAN) published by Oxford University Press.

Yacine Rekik is Professor of Operations & Supply Chain Management at emlyon business school. He holds a PhD in Industrial Engineering from École Centrale Paris and previously served as a Research Associate at the Distributed Information & Automation Laboratory of the University of Cambridge. His research focuses on RFID-enabled supply chain performance, data-driven supply chain optimization, and sustainable operations. He is also a research affiliate of the ECR Community, where he collaborates with leading European retailers on inventory record inaccuracy, its implications, detection, and mitigation.

ORCID

Christoph H. Glock http://orcid.org/0000-0001-6006-0070
Aris A. Syntetos http://orcid.org/0000-0003-4639-0756
Yacine Rekik http://orcid.org/0000-0002-2021-6680

Data availability statement

The participants of this study did not give written consent for their data to be shared publicly, so due to the sensitive nature of the research, supporting data is not available.

References

Anand, N., H. K. Gardner, and T. Morris. 2007. "Knowledge-Based Innovation: Emergence and Embedding of New Practice Areas in Management Consulting Firms." *Academy of Management Journal* 50 (2): 406–428. https://doi.org/10.5465/amj.2007.24634457

Barratt, M., T. J. Kull, and A. C. Sodero. 2018. "Inventory Record Inaccuracy Dynamics and the Role of Employees within Multi-Channel Distribution Center Inventory Systems." *Journal of Operations Management* 63 (1): 6–24. https://doi.org/10.1016/j.jom.2018.09.003

Bernard, P. A. 1985. "Cycle Counting – the Missing Link." *Production and Inventory Management Journal* 26 (4) :27–40.

Bertolini, M., E. Bottani, G. Romagnoli, and G. Vignali. 2015. "The Impact of RFID Technologies on Inventory Accuracy in the Apparel Retailing: Evidence from the Field." *International Journal of RF Technologies* 6 (4): 225–246. https://doi.org/10.3233/RFT-150069

Best, J., C. H. Glock, E. H. Grosse, Y. Rekik, and A. A. Syntetos. 2022. "On the Causes of Positive Inventory Discrepancies in Retail Stores." *International Journal of Physical Distribution & Logistics Management* 52 (5/6): 414–430. https://doi.org/10.1108/IJPDLM-07-2021-0291

Boylan, J. E., and A. A. Syntetos. 2021. *Intermittent Demand Forecasting: Context, Methods and Applications*. Hoboken, NJ ; Chichester, West Sussex: Wiley.

Brooks, R. B., and L. W. Wilson. 1995. "Mix It Up: Cycle Counting for Inventory Accuracy." *Materials Management in Health Care* 4 (7): 26–27.

Bruccoleri, M., S. Cannella, and G. La Porta. 2014. "Inventory Record Inaccuracy in Supply Chains: The Role of Workers' Behavior." *International Journal of Physical Distribution & Logistics Management* 44 (10): 796–819. https://doi.org/10.1108/IJPDLM-09-2013-0240

Chaitien, P., and S. Ramingwong. 2024. "Enhancing Warehouse Management Efficiency for Precast Concrete Product Business." *Science & Technology Asia* 29 (2): 19–31. https://doi.org/10.14456/scitechasia.2024.22.

Chuang, H. H.-C., and R. Oliva. 2015. "Inventory Record Inaccuracy: Causes and Labor Effects." *Journal of Operations Management* 39-40 (1): 63–78. https://doi.org/10.1016/j.jom.2015.07.006

Chuang, H. H.-C., R. Oliva, and S. Kumar 2022. "Group-Level Information Decay and Inventory Inspection: An Empirical-Analytical Approach." *Journal of Operations Management* 68: 130–152. 10.1002/joom.1167.

Chuang, H. H.-C., R. Oliva, and S. Liu. 2016. "On-Shelf Availability, Retail Performance, and External Audits: A Field Experiment." *Production and Operations Management* 25 (5): 935–951. https://doi.org/10.1111/poms.12519

Collins, T. R., M. D. Rosetti, H. L. Nachtmann, and J. Oldhamn. 2006. "The Use of Multi-Attribute Utility Theory to Determine the Overall Best-in-Class Performer in a Benchmarking Study." *Benchmarking: An International Journal* 13 (4): 431–446. https://doi.org/10.1108/14635770610676281

Cooper, H. M. 2010. *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*. 4th ed. Thousand Oaks: Sage Publications..

Corbin, J. M., and A. Strauss. 1990. "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria." *Qualitative Sociology* 13 (1): 3–21. https://doi.org/10.1007/BF00988593

DeHoratius, N., A. Holzapfel, H. Kuhn, A. J. Mersereau, and M. Sternbeck. 2023. "Evaluating Count Prioritization Procedures for Improving Inventory Accuracy in Retail Stores." *Manufacturing & Service Operations Management* 25 (1): 288–306. https://doi.org/10.1287/msom.2022.1119

DeHoratius, N., and A. Raman. 2008. "Inventory Record Inaccuracy: An Empirical Analysis." *Management Science* 54 (4): 627–641. https://doi.org/10.1287/mnsc.1070.0789

Destro, I. R., F. H. Staudt, K. Somensi, and C. Taboada. 2023. "The Impacts of Inventory Record Inaccuracy and Cycle Counting on Distribution Center Performance." *Production* 33:e20220077. https://doi.org/10.1590/0103-6513.20220077

Eisenhardt, K. M. 1989. "Building Theories from Case Study Research." *Academy of Management Review* 14 (4): 532–550. https://doi.org/10.2307/258557

Eisenhardt, K. M., and M. E. Graebner. 2007. "Theory Building from Cases: Opportunities and Challenges." *Academy of Management Journal* 50 (1): 25–32. https://doi.org/10.5465/amj.2007.24160888

Ernst, R., J.-L. Guerrero, and A. Roshwalb. 1993. "A Quality Control Approach for Monitoring Inventory Stock Levels." *Journal of the Operational Research Society* 44 (11): 1115–1127. https://doi.org/10.1057/jors.1993.184

Fleisch, E., and C. Tellkamp. 2005. "Inventory Inaccuracy and Supply Chain Performance: A Simulation Study of a Retail Supply Chain." *International Journal of Production Economics* 95 (3): 373–385. https://doi.org/10.1016/j.ijpe.2004.02.003

Gemechu, F., M. Ayalew, B. B. Tefera, T. Mulugeta, and A. Bekele. 2021. "Evaluating Inventory Management Practice of Antiretroviral Drugs in Public Health Facilities of Addis Ababa, Ethiopia." *HIV/AIDS - Research and Palliative Care* 13:1091–1099. https://doi.org/10.2147/HIV.S337479

Glock, C. H., E. H. Grosse, R. M. Elbert, and T. Franzke. 2017. "Maverick Picking: The Impact of Modifications in Work Schedules on Manual Order Picking Processes." *International Journal of Production Research* 55 (21): 6344–6360. https://doi.org/10.1080/00207543.2016.1252862

Goodwin, P., and R. Lawton. 1999. "On the Asymmetry of the Symmetric MAPE." *International Journal of Forecasting* 15 (4): 405–408. https://doi.org/10.1016/S0169-2070(99)00007-2

Goyal, S., B. C. Hardgrave, J. A. Aloysius, and N. DeHoratius. 2016. "The Effectiveness of RFID in Backroom and Sales Floor Inventory Management." *International Journal of Logistics Management* 27 (3): 795–815. https://doi.org/10.1108/IJLM-03-2015-0051

Graff, R. T. 1987. "The Limitations of Cycle Counting." *Production and Inventory Management Journal* 28 (4):39–43.

Grosse, E. H., S. M. Dixon, W. P. Neumann, and C. H. Glock. 2016. "Using Qualitative Interviewing to Examine Human Factors in Warehouse Order Picking: Technical Note." *International Journal of Logistics Systems and Management* 23 (4): 499–518. https://doi.org/10.1504/IJLSM.2016.075211

Gumrukcu, S., M. D. Rosetti, and N. Buyurgan. 2008. “Quantifying the Costs of Cycle Counting in a Two-Echelon Supply Chain with Multiple Items.” *International Journal of Production Economics* 116 (2): 263–274. https://doi.org/10.1016/j.ijpe.2008.09.006

Gutesa, A., T. Jebena, and O. Kebede. 2024. “Inventory Management Performance for Tracer Medicines in Public Health Facilities of Southwest Shewa Zone Oromia Region, Ethiopia: A Mixed Study.” *SAGE Open Medicine* 12:1–13. https://doi.org/10.1177/20503121241274041

Hardgrave, B. C., J. Aloysius, and S. Goyal. 2009. “Does RFID Improve Inventory Accuracy? A Preliminary Analysis.” *International Journal of RF Technologies: Research and Applications* 1 (1): 44–56. https://doi.org/10.1080/17545730802338333

Hardgrave, B. C., J. Aloysius, and S. Goyal. 2013. “RFID-Enabled Visibility and Retail Inventory Record Inaccuracy: Experiments in the Field.” *Production and Operations Management* 22 (4): 843–856. https://doi.org/10.1111/poms.12010

Hochrein, S., and C. H. Glock. 2012. “Systematic Literature Reviews in Purchasing and Supply Management Research: A Tertiary Study.” *International Journal of Integrated Supply Management* 7 (4): 215–245. https://doi.org/10.1504/IJISM.2012.052773

Hsieh, H.-F., and S. E. Shannon. 2005. “Three Approaches to Qualitative Content Analysis.” *Qualitative Health Research* 15 (9): 1277–1288. https://doi.org/10.1177/1049732305276687

Ishfaq, R., and U. Raja. 2020a. “Empirical Evaluation of IRI Mitigation Strategies in Retail Stores.” *Journal of the Operational Research Society* 71 (12): 1972–1985. https://doi.org/10.1080/01605682.2019.1640592

Ishfaq, R., and U. Raja. 2020b. “Effectiveness of Frequent Inventory Audits in Retail Stores: An Empirical Investigation.” *International Journal of Logistics Management* 31 (1): 21–41. https://doi.org/10.1108/IJLM-07-2018-0184

Kang, Y., and S. B. Gershwin. 2005. “Information Inaccuracy in Inventory Systems: Stock Loss and Stockout.” *IIE Transactions* 37 (9): 843–859. https://doi.org/10.1080/07408170590969861

Kull, T. J., M. Barratt, A. C. Sodero, and E. Rabinovich. 2013. “Investigating the Effects of Daily Inventory Record Inaccuracy in Multichannel Retailing.” *Journal of Business Logistics* 34 (3): 189–208. https://doi.org/10.1111/jbl.12019

Meyer, H. 1991. “Eight-Step Approach to Inventory Accuracy. A Case Study at Leviton Manufacturing Company.” *Production and Inventory Management Journal* 32 (2) :40–42.

Miller, G. J. 1997. “Inventory Accuracy in 60 Days!” *Hospital Materiel Management Quarterly* 19 (1) :12–25.

Millet, I. 1994. “A Novena to Saint Anthony, or How to Find Inventory by Not Looking.” *Interfaces* 24 (2): 69–75. https://doi.org/10.1287/inte.24.2.69

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, et al. 2021. “The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews.” *PLOS Medicine* 18 (3): e1003583. https://doi.org/10.1371/journal.pmed.1003583

Raman, A., N. DeHoratius, and Z. Ton. 2001. “Execution: The Missing Link in Retail Operations.” *California Management Review* 43 (3): 136–152. https://doi.org/10.2307/41166093

Rekik, Y., R. Oliva, A. A. Syntetos, and C. H. Glock. 2025. *Inventory Record Inaccuracy in Grocery Retailing: Impact of Promotions and Product Perishability, and Targeted Effect of Audits*. Available at arXiv.

Rekik, Y., A. A. Syntetos, and C. H. Glock. 2019a. “Modelling (And Learning From) Inventory Inaccuracies in E-Retailing/B2B Contexts.” *Decision Sciences* 50 (6): 1184–1223. https://doi.org/10.1111/deci.12367

Rekik, Y., A. A. Syntetos, and C. H. Glock. 2019b. “Inventory Inaccuracy in Retailing: Does It Matter?” https://ecr-shrink-group.com/page/inventory-record-accuracy-research .

Rinehart, R. F. 1960. “Effects and Causes of Discrepancies in Supply Operations.” *Operations Research* 8 (4): 543–564. https://doi.org/10.1287/opre.8.4.543

Roulston, K. 2010. “Considering Quality in Qualitative Interviewing.” *Qualitative Research* 10 (2): 199–228. https://doi.org/10.1177/1468794109356739

Shabani, A., G. Maroti, S. de Leeuw, and W. Dullaert. 2021. “Inventory Record Inaccuracy and Store-Level Performance.” *International Journal of Production Economics* 235:108111. https://doi.org/10.1016/j.ijpe.2021.108111

Sheppard, G. M., and K. A. Brown. 1993. “Predicting Inventory Record-Keeping Errors with Discriminant Analysis: A Field Experiment.” *International Journal of Production Economics* 32 (1): 39–51. https://doi.org/10.1016/0925-5273(93)90006-7

Thoring, K., R. M. Mueller, and P. Badke-Schaub. 2020. “Workshops as a Research Method: Guidelines for Designing and Evaluating Artifacts Through Workshops.” In *Proceedings of the Annual Hawaii International Conference on System Sciences*, 5036–5045.Hawaii..

Wang, H., S. Chen, and Y. Xie. 2010. “An RFID-Based Digital Warehouse Management System in the Tobacco Industry: A Case Study.” *International Journal of Production Research* 28 (9): 2513–2548. https://doi.org/10.1080/00207540903564918

Wilson, J. M. 1995. “Quality Control Methods in Cycle Counting for Record Accuracy Management.” *International Journal of Operations & Production Management* 15 (7): 27–39. https://doi.org/10.1108/01443579510090390

Yin, R. K. 2018. *Case Study Research and Applications: Design and Methods*. 6th ed. Thousand Oaks, California: SAGE Publications.

Appendix

Table A1. Overview of the interview sample.

#	Role of the interviewee	Retail sector	Location
1	Retired Retail Expert	N/A	North America
2	Head of Global Store Operations	Grocery	Europe
3	Commercial Manager Loss Prevention	Household hardware	Oceania
4	Head of Stock Operations	Grocery	Europe
5	Performance Manager	Grocery	Europe
6	Product Director	Online Grocery	Europe
7	Supply Chain Developer	Fashion	Europe
8	Team Manager Stock Movement	Grocery	Europe
9	Internal Audit Manager	Grocery	Europe
10	Accuracy and loss business partner	Fashion	Europe
11	Project Manager Group Operations	Fashion	Europe
12	Team leader central supply function	Grocery	Europe
13	Head of Retail Operations	Pharmaceutical	Europe
14	RFID and Analytics Manager	Fashion	Europe
15	Retail Strategy Leader	Grocery	North America
16	Process Lead	Online Grocery	Europe
17	Senior Director Retail Operations	Grocery	North America
18	Stock Optimization and Retail Audit Manager	Household hardware	Europe
19	Solution Analyst Supply Chain	Grocery	Europe
20	Lead Analytics Manager	Grocery	Europe
21	National Director of Hypermarket Format	Grocery	Europe
22	Program Manager	Grocery	Oceania
23	Manager Loss and Fraud Prevention	Grocery	Europe
24	Head of Front Store Operations & Innovation Team	Pharmaceutical	North America
25	Replenishment Director	Grocery	Europe

Table A2. Overview and classification of the pertinent literature.

<table>
<tr><th>Authors</th><th>Method</th><th>Measure</th><th>Description</th><th>Unit of measure</th><th>Addressed in the interviews</th></tr>
<tr><td>Barratt, Kull, and Sodero (2018)</td><td>Case Study/Simulation</td><td>Signed relative</td><td>(Actual inventory – System inventory)/actual inventory</td><td>volume</td><td>No</td></tr>
<tr><td>Bertolini et al. (2015)</td><td>Case Study</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>Best et al. (2022)</td><td>Simulation</td><td>Signed difference</td><td>Actual inventory – system inventory</td><td>volume</td><td>Yes</td></tr>
<tr><td>Brooks and Wilson (1995)</td><td>Conceptual</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>Bruccoleri, Cannella, and La Porta (2014)</td><td>Simulation</td><td>Absolute relative</td><td>|Actual inventory – System inventory|/Actual inventory</td><td>volume</td><td>Yes</td></tr>
<tr><td>Chaitien and Ramingwong (2024)</td><td>Case Study</td><td>Signed relative</td><td>(Actual inventory – System inventory)/System inventory</td><td>volume</td><td>No</td></tr>
<tr><td>Chuang and Oliva (2015)</td><td>Empirical study/ Simulation</td><td>Absolute relative</td><td>|Actual inventory – system inventory|/Unit of simulation time</td><td>volume</td><td>No</td></tr>
<tr><td>Chuang et al. (2022)</td><td>Mathematical modelling/Empirical study</td><td>Binary</td><td>A SKU is either correct or incorrect.</td><td>volume</td><td>Yes</td></tr>
<tr><td>Chuang, Oliva, and Liu (2016)</td><td>Empirical study</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>Collins et al. (2006)</td><td>Case study</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>DeHoratius and Raman (2008)</td><td>Empirical study</td><td>Absolute, Absolute relative</td><td>|System inventory – actual inventory|; |System inventory – actual inventory|/Daily units sold</td><td>volume, value</td><td>Yes</td></tr>
<tr><td>DeHoratius et al. (2023)</td><td>Empirical study/ Simulation</td><td>Signed difference, Absolute relative</td><td>System inventory – actual inventory, |System inventory – actual inventory|/average system inventory</td><td>volume</td><td>Partly</td></tr>
<tr><td>Destro et al. (2023)</td><td>Simulation</td><td>Absolute relative</td><td>|System inventory – Physical inventory|/Physical inventory</td><td>volume</td><td>No</td></tr>
<tr><td>Ernst, Guerrero, and Roshwalb (1993)</td><td>Empirical study/ Simulation</td><td>Signed relative</td><td>|Actual inventory – System inventory|/System inventory; summed up over all SKUs and then divided by the number of SKUs</td><td>volume</td><td>Yes</td></tr>
<tr><td>Fleisch and Tellkamp (2005)</td><td>Simulation</td><td>Absolute relative</td><td>|Actual inventory – System inventory |/Average physical inventory</td><td>volume</td><td>No</td></tr>
<tr><td>Gemechu et al. (2021)</td><td>Empirical study</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>Goyal et al. (2016)</td><td>Empirical study</td><td>Absolute</td><td>|System inventory – actual inventory|</td><td>volume</td><td>Yes</td></tr>
<tr><td>Graff (1987)</td><td>Conceptual</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>Gumrukcu, Rosetti, and Buyurgan (2008)</td><td>Simulation</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
<tr><td>Gutesa, Jebena, and Kebede (2024)</td><td>Empirical</td><td>Binary relative</td><td>Total number of accurate records/Number of records checked</td><td>volume</td><td>Yes</td></tr>
</table>

(*Continued*)

Table A2. (Continued).

Authors	Method	Measure	Description	Unit of measure	Addressed in the interviews
Hardgrave, Aloysius, and Goyal (2009)	Empirical study	Binary with range	The authors define three categories: correct, close (±2 units), and inaccurate	volume	Yes
Hardgrave, Aloysius, and Goyal (2013)	Empirical study	Absolute	|Actual inventory – System inventory|	volume	Yes
Ishfaq and Raja (2020a)	Empirical study/ Simulation	Signed difference, Binary, Binary relative	Actual inventory – system inventory; A SKU is either correct or incorrect; Total number of accurate records/Number of records checked	volume	Yes
Ishfaq and Raja (2020b)	Empirical study/ Simulation	Signed difference, Binary, Binary relative	Actual inventory – system inventory; A SKU is either correct or incorrect; Total number of accurate records/Number of records checked	volume	Yes
Kang and Gershwin (2005)	Empirical study/ Simulation	Binary relative, Binary with range	Total number of accurate records/Number of records checked, Binary with range; the range the authors consider is ±5 items	volume	Yes
Kull et al. (2013)	Empirical study/ Simulation	Signed relative	(System inventory – actual inventory)/actual inventory Considered if actual inventory > 0, 0 otherwise	volume	No
Meyer (1991)	Case Study	Binary relative	Total number of accurate records/Number of records checked	volume	Yes
Miller (1997)	Conceptual	Binary with range; Binary relative	Binary with range; ABC classification used to determine range, range in percent; Total number of accurate records/Number of records checked	volume	Yes
Millet (1994)	Case Study	Binary relative	Total number of accurate records/Number of records checked	volume	Yes
Bernard (1985)	Conceptual	Binary with range; Binary relative	Binary with range; ABC classification used to determine range, range in percent; Total number of accurate records/Number of records checked	volume	Yes
Raman, DeHoratius, and Ton (2001)	Case Study	Absolute	|Actual inventory – System inventory|	volume	Yes
Rekik, Syntetos, and Glock (2019a)	Empirical Study/Mathematical modelling	Signed difference	System inventory – actual inventory	volume	Yes
Rinehart (1960)	Case Study	Binary with range	A record is considered accurate if the error is not larger than 1% of the record balance or if the monetary value of the difference is not more than \$1.	volume, value	Partly
Shabani et al. (2021)	Mathematical modelling/Case Study	Signed relative	(System inventory – actual inventory)/number of units received	volume	No
Sheppard and Brown (1993)	Empirical study	Binary with range	A record is considered accurate if the error is not larger than four units	volume	Yes
Wang, Chen, and Xie (2010)	Case Study	Binary relative	Total number of accurate records/Number of records checked	volume	Yes
Wilson (1995)	Conceptual	Binary relative	Total number of inaccurate records/Number of records checked	volume	Yes