Characteristic Curves for Multiple-Inspector Sampling Plans with Inspection Errors

Quality and Reliability Engineering International

WILEY

RESEARCH ARTICLE OPEN ACCESS

Characteristic Curves for Multiple-Inspector Sampling Plans with Inspection Errors

Domenico A. Maisano | Lucrezia Ferrara | Fiorenzo Franceschini

Politecnico di Torino, Torino, Italy

Correspondence: Fiorenzo Franceschini (fiorenzo.franceschini@polito.it)

Received: 27 May 2025 | **Revised:** 18 August 2025 | **Accepted:** 22 August 2025

Funding: This research was carried out under the MICS (Made in Italy, Circular and Sustainable) Extended Partnership and partially funded by the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza – Missione 4, Componente 2, Investimento 1.3, D.D. 1551.11-10-2022, PE00000004).

Keywords: acceptance sampling plan | aggregation | conformity assessment | inspection error | majority criterion | multiple inspectors | operating characteristic curve | unanimity criterion

ABSTRACT

Classical *single sampling plans* (SSPs) are designed assuming that product units sampled from a lot are inspected by one inspector, in the absence or presence of inspection errors. The scenario becomes more complicated when SSPs are applied assuming that multiple inspectors operate in parallel on the same sample, as occasionally required in high-value-added or highly customized industries (e.g., aerospace, defence, luxury goods, etc.). Current literature lacks a rigorous formulation of *operating characteristic* (OC) curves for SSPs involving multiple inspectors, who independently perform conformity assessments on the same product units. This paper addresses this gap by extending classical OC-curve theory to scenarios involving multiple inspectors, each characterized by distinct individual error rates (i.e., probabilities of misclassifying conforming units as defective ones and vice versa). The resulting analytical models are flexible enough to incorporate alternative aggregation criteria and sequences for consolidating individual inspectors' conformity assessments into an overall lot-disposition decision. Both *hypergeometric* (Type-A) and *binomial* (Type-B) formulations of the multiple-inspector OC curve are presented. Results show that in the absence of inspection errors, the multiple-inspector OC curve converges to the classical single-inspector curve. Moreover, aggregation criteria may significantly influence the OC-curve shape: a *majority* aggregation criterion yields robust OC curves, even under moderate to high error rates, whereas an *unanimity* criterion produces excessively severe OC curves, unless inspection errors are very low. Overall, the proposed analytical framework enables a realistic design of SSPs, particularly beneficial for sectors demanding highly reliable inspections.

1 | Introduction

Multiple-inspector sampling plans are conformity assessments in which individual product units of a sample—drawn from a production lot—undergo parallel, independent inspection by multiple inspectors [1–3]. In this scenario, the same sample units are inspected in turn by the multiple inspectors, who work in a *blind* mode—that is, without knowing the assessments given by the others. The presence of redundant inspections on the same sample units might appear, at first glance, as a possible waste of resources; however, such inspections are justified by the need for consistent verification in many high value-added, high-reliability and high-customization contexts, for example, in the aerospace, defence, nuclear, *haute couture* or luxury sector [4].

Laboratory experiments show that human redundancy (e.g., sequential inspections by two operators) may significantly improve defect detection in low-prevalence visual tasks, common in safety-critical domains, such as medical imaging and pharmaceutical inspection [5]. Similarly, industrial research demonstrates that two inspectors independently classifying the same unit under a 'two-reject' rule detect more defects than one inspector alone, while preserving acceptable false-defect rates [6]. In aerospace and defence manufacturing, redundant visual and dimensional inspections are standard practice for safety-critical parts like turbine blades or avionics assemblies, where undetected defects could be catastrophic [7]. In luxury manufacturing, multi-inspector schemes are applied to ensure flawless appearance and function in high-value products, such as *haute couture* garments or fine watchmaking, where brand reputation relies on perfect quality [8].

Inspections—particularly manual inspections, when lots undergo frequent changes in terms of product type, as in the case of relatively limited supplies of highly customized products—are inevitably subject to variability among inspectors (in terms of experience, training, level of concentration, etc.). In such a scenario, the use of parallel and redundant inspections allows to mitigate the inherent subjectivity and increase alignment among inspectors [9]. Second, the presence of multiple independent assessments on the same product units provides insights into the internal consistency and uniformity of inspection procedures, as well as the performance of each inspector. This highlights potential misalignments among inspectors in their respective conformity assessments, suggesting corrective action or ad hoc training interventions [8].

Although 100% inspection of the entire lot can be assumed, in many practical contexts it is preferable to limit inspection to a sample drawn from the lot, for both economic and technical reasons [10, 11]. It should also be considered that a 100% inspection by a single 'misaligned' (e.g., inexperienced) inspector could pose greater risks than redundant (thus more robust) inspections conducted on only part of the lot, but by multiple independent inspectors [8, 12].

The scientific literature describes a variety of schemes for implementing sampling plans, which are summarised in the literature review in Section 2. In this paper, explicit reference is made to *single sampling plans* (SSPs) by *attributes*, which represent the most basic and widespread scheme [11]. In summary, a SSP is characterized by two main parameters: sample size (n) and acceptance/rejection number (c). After extracting a sample of n units from a lot (of N units), the number d of defective (or nonconforming) units is counted: if $d \leq c$ the entire lot is accepted (*pass*), otherwise it is rejected (*fail*) [10, 11]. Such sampling plans are also codified in international standards such as ISO/DIS 2859-1:2025 [13], although they traditionally assume a single inspector who does not make inspection errors.

In practice, when conformity assessments are not entirely objective—due to the presence of inherently subjective elements—each inspector may commit two primary types of errors: (i) *false conforming*, where defective units misclassified as conforming, and (ii) *false defective*, where conforming units are misclassified as defective. As a result, the *fraction nonconforming*

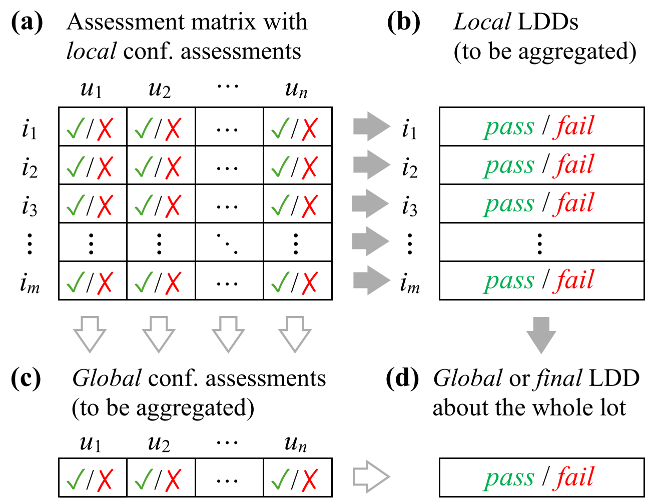

FIGURE 1 | Conceptual scheme illustrating a conformity-assessment matrix with two possible aggregation sequences for transforming its content into a final LDD for the whole lot. The aggregation can be performed either *first by rows then by column*, **(a)** → **(b)** → **(d)**, represented by grey arrows, or *first by columns then by row*, **(a)** → **(c)** → **(d)**, represented by white arrows (adapted from [8]). The final LDD outcome may differ based on the aggregation sequence and the specific column-wise aggregation criteria.

(also referred to as *fraction defective*, *percent defective* or *defectiveness*) resulting from the inspections carried out by each inspector—which we henceforth refer to as 'apparent'[1]—may differ from the 'true' one [14, 15].

At the operational level, the results of parallel multiple-inspector inspections can be organized in a *conformity-assessment matrix*, as shown in Figure 1 (1). Each row corresponds to an i-th inspector ($i_1, i_2, \ldots, i_m$), each column to a j-th sample unit ($u_1, u_2, \ldots, u_n$); each cell reports the outcome: '✓' (conforming) or '✗' (defective) of the conformity assessment of the j-th sample unit by the i-th inspector.

A final *lot-disposition decision* (LDD) of *pass*/*fail* for the entire lot must be derived from that matrix. There are many possible aggregation approaches. For example, each inspector can first determine a *local* LDD of *pass*/*fail* (row by row) and then the various local LDDs are combined into a single *global* LDD (aggregating by column). Alternatively, for each sample unit, the local conformity assessments can first be merged into a single global assessment (column-wise aggregation) and then the resulting unit-level assessments can be consolidated into one global *pass*/*fail* LDD for the entire lot (aggregation by row) [1, 8, 16]. A recent paper [8] exemplified some of these techniques and proposed a diagnostic indicator—Gwet's kappa (κ_G)—to assess the level of agreement (or disagreement) among inspectors.

The scientific literature provides a variety of *operating-characteristic curves* (OC curves) for sampling plans, which implicitly consider a single-inspector with no inspection errors: these OC curves express the probability of lot acceptance (P_a) as a function of the lot defectiveness (p) [11, 13]. OC curves are crucial tools for assessing the so-called *discriminatory power* of a specific sampling plan, also understood as *severity* from the *supplier*'s point of view, and level of *protection*, from the *customer*'s point of view [10, 11, 17]. The choice of the most appropriate sampling plan—determined by the selection of the relevant n and c parameters, so as to meet the needs of the parties involved (typically *supplier* and *customer*, as mentioned above)—always goes through the construction of the corresponding OC curve[2]. However, when shifting to multiple-inspector plans with possible inspection errors, the construction of OC curves becomes much more complex, for at least two reasons: (i) *sequences* and *criteria* of aggregation of multiple-inspector conformity assessments need to be taken into account, and (ii) the inspection *errors* of each inspector (e.g., probabilities of classifying 'false conforming' and 'false defective' units) need to be modelled.

Although there are a few studies in the literature on error-prone inspections [11, 18–20], they generally focus on single inspectors and do not formalize the construction of OC curves in the multiple-inspector case. In practice, this leads to a significant gap, because companies operating with multiple parallel inspections often continue to use OC curves from single-inspector sampling plans and without considering inspection errors, as if there were only one 'infallible' inspector. This creates a discrepancy between the actual procedures adopted and the quantitative tools to analyse them.

The main objective of this paper is to fill this gap, by proposing a rigorous formalization of the construction of OC curves for multiple-inspector SSPs in the presence of inspection errors. In particular, it aims to answer the three following research questions:

RQ1: What analytical model extends the classic definition of OC curves to the multiple-inspector context, incorporating the different propensity of each inspector to commit errors?

RQ2: To what extent do the multiple-inspector OC curves differ from the 'ideal' ones (with single inspector and no inspection errors) and how do the related parameters affect their shape and discriminatory power?

RQ3: Which configurations of sequences and criteria for aggregation are best in ensuring an effective and easily manageable multiple-inspector sampling plan?

The rest of the article is structured into five sections. Section 2 provides a concise review of the scientific literature concerning the approaches for modelling sampling plans in the presence of multiple inspectors and inspection errors. Section 3 formally defines the basic structure of multiple-inspector sampling plan and corresponding parameters. Section 4 develops new analytical models for the construction of the OC curves, analyzing several possible sequences and criteria for aggregation. The formulation of the new analytical models is considerably more complex than in the 'ideal' case, since both the errors of individual inspectors and the ways in which the multiple-inspector conformity assessments are aggregated must be taken into account [21]. Section 5 provides numerical examples of OC-curve construction, discussing their practical implications. The conclusions summarize contributions of originality, application aspects, limitations, and possible lines of future development. Finally, the appendix contains additional material detailing the previous analysis.

2 | Literature Review

Acceptance sampling plans have been developed in various configurations, including single, double, sequential and skip-lot plans, each balancing inspection effort and risks [11]. For example, double sampling plans allow a second sample when the first sample results are inconclusive, improving efficiency at high quality levels [22]. Fully sequential plans inspect units one by one, making acceptance/rejection decisions as soon as enough evidence is collected, often minimizing the average sample size [23]. Skip-lot plans further reduce inspection by occasionally skipping lot inspections once a supplier has a demonstrated history of good quality, thereby lowering costs while maintaining overall protection through periodic sampling [12, 24, 25].

Traditional acceptance sampling theory and standards (e.g., MIL-STD and ISO 2859) assume perfect inspections—that is, every defective unit in the sample is detected and no conforming unit is falsely rejected—under which these plans can achieve the desired producer's and consumer's risk points [13]. In practice, however, human or device inspectors are subject to misclassification errors, and inspection errors can significantly degrade the performance of acceptance sampling. An inspector may erroneously classify a conforming unit as defective or overlook a truly defective unit, analogous to Type-I and Type-II errors in hypothesis testing. These errors have long been recognized to flatten the OC curve and reduce the discriminatory power of sampling plans [26]. In other words, a plan under imperfect inspection will accept bad lots more often and reject good lots more often than predicted by the 'ideal' OC curve, in the absence of inspection errors.

Researchers have incorporated such inspection error probabilities into acceptance sampling models to evaluate their impact on lot acceptance criteria and to adjust plan parameters for mitigating risks. For instance, models have been proposed to optimize sampling plans economically by accounting for the costs of false decisions due to inspector errors. Overall, the literature confirms that even modest error rates can significantly affect key performance metrics (e.g., supplier and customer risk), underscoring the importance of considering inspector accuracy in plan design [26]. One approach to improve the reliability of inspections, especially for critical applications, is to employ multiple inspectors (or repeated inspections) for the same units. By having multiple inspectors independently examine each sampled unit, the chance that a defect goes undetected is reduced, and likewise the chance of wrongly scrapping a conforming unit can be minimized through cross-verification. Multiple-inspector sampling plans (also known as multiple inspection or repeated screening procedures) have been shown to enhance detection accuracy at the cost of additional inspection effort [27]. For example, expensive or safety-critical components (such as microchips or aerospace parts) are often subjected to repeated inspections to ensure that the probability of an undetected defect is extremely low [27]. Franceschini et al. [8] explore the problem of aggregating the assessments of multiple inspectors in an acceptance sampling context, highlighting how to combine individual inspection results into a final lot disposition decision. Early analyses of multiple-inspector scenarios suggest that agreement rules (e.g., requiring consensus or majority votes) will affect the composite OC curve of the plan. This emerging body of work on multiple-inspector plans extends the classic acceptance sampling theory by explicitly accounting for inspector-to-inspector redundancy and inconsistency.

It is important to clarify that the multi-inspector scenario in this study is different from the problem of multiple hypothesis tests on different attributes of the same unit. In the latter domain—a distinct branch of the literature—one conducts multiple statistical tests on various quality characteristics of a unit or lot, often requiring specialized error control procedures. Metrics such as the family-wise error rate (FWER) and false discovery rate (FDR) have been developed to manage the overall Type-I error in such multiple-testing situations [28–30]. These methods ensure, for example, that if ten independent features of a product are tested, the chances of any false defect signal or the proportion of false defect signals can be controlled across all tests. However, those approaches assume different tests or inspection criteria per unit and aim to avoid spurious rejections when examining numerous attributes simultaneously. In our context, by contrast, all inspectors are performing homologous checks for the same defect criterion on each unit—effectively repeating the same test multiple times to improve its accuracy. Thus, concepts like FDR and FWER are not directly applicable here, except by analogy.

The focus of the present research is on how multiple identical inspections (by human inspectors or test devices) can be combined to influence the OC curve and risks of an acceptance sampling plan. This fills a gap in the acceptance sampling literature: while single-inspector plans with inspection errors are well studied, comprehensive characterization of OC curves for multiple-inspector sampling plans has remained scarce. Our contribution builds on the above streams of research by deriving the OC curve for multiple-inspector plans with inspector errors and examining how inspector redundancy can restore or enhance the discriminatory power of acceptance sampling.

Finally, some authors have proposed modelling inspector behaviour through latent variables, such as in *structural equation models* (SEM) or latent class frameworks. These approaches are suitable when the true status of inspected units is unobserved, and inspector accuracy must be inferred from data [31]. In the present study, however, we assume known or empirically estimated error rates for each inspector (cf., Section 3.2), which allows for a direct analytical formulation. As such, latent modelling was considered unnecessary given our data and goals. Nonetheless, SEM approaches may prove valuable in future investigations involving richer datasets or scenarios where both inspector reliability and unit truth must be simultaneously inferred.

3 | Problem Formalization

This section recalls the configuration of an 'ideal' SSP for attributes without inspection errors and illustrates its extension to the multiple-inspector case, where each inspector is potentially prone to inspection errors.

3.1 | 'Ideal' SSP With a Single Inspector Without Inspection Errors

Let us consider a lot consisting of N product units, characterised by a certain defectiveness (or fraction nonconforming) $p = D/N$, D being the number of defective elements in the entire lot. A sample of size n is extracted from the lot, intended for inspection by a single inspector, using a classical SSP and assuming no inspection errors. The aim of the inspection is to determine whether the whole lot should be accepted or rejected, based on the following two-step rule (i.e., the same SSP's LDD recalled in Section 1):

1. the number of defectives (d) in the sample is determined through inspection;
2. if $d \leq c$, where c is the acceptance/rejection number, the whole lot is accepted (*pass*); otherwise, it is rejected (*fail*).

The choice of parameters (n, c) can be made directly, through targeted statistical risk analyses for the parties involved (*supplier* and *customer*), or indirectly, following established international standards such as ISO/DIS 2859-1:2025 [11, 13].

For lots of finite size (N) and in the absence of inspection errors, the OC curve of a SSP can be constructed using the formula based on the *hypergeometric* distribution [32]:

$$P_a = \mathbb{P}\,(d \leq c) = \sum_{d=0}^{c} \mathbb{P}\,(d) = \sum_{d=0}^{c} \frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}} \tag{1}$$

When n is much smaller than N (i.e., $n << N$ or $n/N \leq 0.1$), it is common practice to approximate the term $\mathbb{P}(d)$ in Equation (1)—that is, the probability of detecting exactly d defectives among the n sampled units—by replacing the *hypergeometric* distribution with the *binomial* one [32]:

$$P_a = \mathbb{P}\,(d \leq c) = \sum_{d=0}^{c} \mathbb{P}\,(d) = \sum_{d=0}^{c} \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d} \tag{2}$$

It can be noticed that the formula in Equation (2) does not take into account the specific value of N, but simply assumes that the condition $n << N$ holds. Figure A1 (in Appendix A.1) contains an example of OC curves for a sampling plan with parameters $N = 15$, $n = 3$ and $c = 1$. For the same sampling plan, we determined both the OC curve based on the hypergeometric distribution (Equation 1), also called Type-A OC curve, and the (approximate) OC curve based on the binomial distribution (Equation 2), also called Type-B OC curve.

3.2 | Extension to Multiple Inspectors With Inspection Errors

Compared to the 'ideal' SSP, two new elements are now introduced: (i) the presence of several inspectors, m in total, who individually and independently formulate their conformity assessments on the same sample of n units drawn from the lot, and (ii) the possibility that each inspector makes inspection errors. In particular, each i-th inspector (i = 1, 2, m) is characterised by two error rates:

a_i: probability of misclassifying as 'defective' (✗) a unit that is truly conforming ('false defective'). Consequently, $(1 - a_i)$ is the probability of (correctly) classifying as 'conforming' a unit that is truly conforming.

b_i: probability of misclassifying as 'conforming' (✓) a unit that is truly defective ('false conforming'). Accordingly, $(1 - b_i)$ is the probability of (correctly) classifying as 'defective' a unit that is truly defective.

The above two probabilities depend on the i-th inspector's skills for a certain type of product. They must be estimated empirically, for example, through training sessions for the inspectors themselves, comparisons between different inspections, inspections of training items whose actual defectiveness is known, and so on[3] [8].

3.3 | Conformity-Assessments Matrix and Aggregation

The results of multiple inspections can be summarised into the conformity-assessment matrix (see Figure 1 (1)). The next step is aggregating the conformity assessments (✓/✗) to obtain a global *pass*/*fail* decision of the lot. This aggregation includes two possible steps:

- *Aggregation by row*: the classical SSP's LDD (cf. Section 3.1) is applied row by row, leading to a (local) *pass*/*fail* decision.
- *Aggregation by column*: refers to *local* conformity assessments on a single product unit, by different (m) inspectors, to be aggregated into a *global* conformity assessment by means of an appropriate column-wise aggregation criterion. The most common column-wise aggregation criteria are *majority*, according to which the inspected unit is deemed conforming if a majority of the inspectors classify it as such, and *unanimity*—or *unanimous conformity* [8]—according to which the unit is deemed conforming only if all inspectors classify it as such. More generally, a threshold r may be introduced on the minimum number of conformity assessments to be able to classify the inspected unit as conforming.

The above aggregation steps can be combined in two possible sequences [8]:

- *Aggregation first by rows (then by column)*: a local *pass*/*fail* LDD is determined for each inspector, using the classical SSP's LDD; the (m) local LDDs are then aggregated by an appropriate column-wise aggregation criterion (e.g., by *majority*) into a single global LDD. This aggregation sequence is schematised in Figure 1 as **(a) → (b) → (d)**, represented by grey arrows.
- *Aggregation first by columns (then by row)*: first the *local* conformity assessments of individual inspectors are aggregated with reference to the individual units of the sample, resulting in corresponding *global* conformity assessments; these column-wise aggregations can be performed by specific criteria (e.g., *majority* or *unanimity*). The *global* conformity assessments are then aggregated by means of the classical SSP's LDD, resulting in a single *global* LDD. This aggregation sequence is schematized in Figure 1 as **(a) → (c) → (da)**, represented by white arrows.

The choice of column-wise aggregation *criteria* (e.g., *majority* vs. *unanimity*) and aggregation sequences (e.g., *first by rows* or *first by columns*) is often adopted empirically, based on established practices or on choices of operational simplicity. However, as will be seen in the next sections, such choices may have implications on the OC curve and thus on the discriminatory power of the SSP.

4 | Development of Analytical Models

This section develops analytical models for constructing the OC curves of a multiple-inspector SSP with inspection errors. It is structured into two subsections that develop the fundamental analytical models for calculating the OC curve with reference to the two aggregation sequences: (Section 4.1) *first by rows* and (Section 4.2) *first by columns*.

Before going into the construction of specific analytical models, it is helpful to recall the general problem. Let us consider a lot of size N with a 'true' defectiveness p. A sample of size n, containing d 'truly' defective units with probability $\mathbb{P}(d)$, is drawn from this lot. This probability can be calculated exactly using the hypergeometric distribution or approximately, if $n << N$, using the binomial distribution, as described in Equations (1) and (2) [32]. Next, m inspectors are considered and, for each i-th inspector, the two error rates defined in Section 3.2 are associated: a_i (probability of false defectives) and b_i (probability of false conforming units).

The scientific literature includes several contributions on SSPs in the presence of inspection errors, in which the misclassification behaviour of a hypothetical i-th inspector leads to the formulation of an *apparent-defectiveness* rate (p'_i), which distorts the 'true' defectiveness (p) through the 'distorted lenses' of the inspector's errors. This apparent defectiveness, specific to each i-th inspector and dependent on his/her individual error rates a_i and b_i, is often defined as [14, 33]:

$$p'_i = (1 - b_i) \cdot p + a_i \cdot (1 - p) \tag{3}$$

The first term at second member, $(1 - b_i) \cdot p$, represents the contribution of defective units that are correctly identified as such, while the second term, $a_i \cdot (1 - p)$, accounts for conforming units misclassified as defective (i.e., *false* defectives) by the i-th inspector.

This expression for p'_i can be substituted into the binomial formula (cf., Equation 2, assuming condition $n << N$ holds) to estimate an acceptance probability under error-prone conformity assessments, yielding a modified probability of acceptance, denoted as P'_{a_i}:

$$P'_{a_i} = \sum_{d=0}^{c} \binom{n}{d} \cdot \left(p'_i\right)^d \cdot \left(1 - p'_i\right)^{n-d} \tag{4}$$

The subscript 'i' in P'_{a_i} emphasizes that this acceptance probability is specific to the i-th inspector and reflects his/her subjective inspection errors; furthermore, the prime symbol (′) indicates that this is an alternative acceptance probability, distinct from the one (P_a, in Equation 2) computed under the assumption of no inspection errors. It is important to note, however, that this approach is only applicable in the case of binomial (Type-B) OC curves, not hypergeometric (Type-A) ones. We will return to this formulation later to analyse its applicability to multiple-inspector context.

4.1 | Case of Aggregation Sequence First by Rows

Let us assume that there are d 'true' defectives and thus $(n - d)$ 'true' conforming units in the sample. The i-th inspector will correctly classify k_1 ('true') defectives as such, k_1 being a binomially distributed variable: $k_1 \sim Bin(d, 1 - b_i)$. On the other hand, the same inspector will misclassify k_2 ('true') conforming units as 'false' defectives, k_2 being a binomially distributed variable: $k_2 \sim Bin(n - d, a_i)$. As a result, the total number of units classified as defective by the i-th inspector, $k = k_1 + k_2$, will follow a distribution obtained from the convolution of the above two binomial distributions [32, 34]. Specifically, the probability that the i-th inspector classifies a total of k defectives (either 'true' and/or 'false') is:

$$P_i\ (k|d) = \sum_{j=\max[0,\ k-(n-d)]}^{\min(k,\ d)} [P_i\,(k_1 = j) \cdot P_i\,(k_2 = k - j)]$$

$$P_i\ (k|d) = \sum_{j=\max[0,\ k-(n-d)]}^{\min(k,\ d)} \left\{ \left[\binom{d}{j} \cdot (1 - b_i)^j \cdot b_i^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1 - a_i)^{(n-d)-(k-j)} \right] \right\} \tag{5}$$

Let us focus on the extremes of the summation. The dummy variable j indicates how many of the d 'truly' defective units are correctly classified as defectives. Since, by definition, j cannot exceed the number of 'true' defectives, it follows that $j \leq d$. Furthermore, the quantity $(k - j)$ counts how many of the $(n - d)$ 'true' conforming units are misclassified as defective; thus the additional constraint $(k - j) \leq (n - d)$ holds. Furthermore, neither j nor $(k - j)$ can be negative. All the above constraints result in: $0 \leq j \leq d$ and $0 \leq (k - j) \leq (n - d)$; in particular, the second condition implies: $j \leq k$ and $j \geq k - (n - d)$. Combining these constraints, explain the expressions $j \leq \max[0, k - (n - d)]$, which corresponds to the lower index of the summation, and $j \leq \min(k, d)$, which corresponds to the upper index of the summation in Equation (5).

The probability that the i-th inspector classifies no more than c defectives out of n, that is, the probability that $k \leq c$, is obtained as:

$$R_i\ (c,\ d, a_i,\ b_i) = \sum_{k=0}^{c} P_i\,(k|d) \tag{6}$$

If it is desired that all m inspectors *unanimously* classify at least r conforming units, the probability of this occurring is the product of their respective terms R_i:

$$\prod_{i=1}^{m} R_i\,(c,\ d, a_i,\ b_i) \tag{7}$$

Equation (7) assumes that the inspected sample contains exactly d 'truly' defective units (as introduced at the beginning of this subsection). To compute the general probability that *all* inspectors classify *no more than c* units as defective, the total probability theorem must be applied [32]. Specifically, the weighted sum of the terms in Equation (7) should be taken over all possible realizations of $d = 0, 1, \ldots, n$, using as weights the corresponding probabilities $\mathbb{P}(d)$—that is, the probability that exactly d defective units are present in the sample; these probabilities were previously introduced in Equations (1) and (2), depending on whether the hypergeometric or binomial distribution is used:

$$\sum_{d=0}^{n} \left\{ \left[\prod_{i=1}^{m} R_i\,(c, d, a_i,\ b_i) \right] \cdot \mathbb{P}\,(d) \right\} \tag{8}$$

In a more general form, if it is required that—not necessarily *unanimity*—but at least q inspectors out of m satisfy the condition of classifying no more than c defectives, it is necessary to sum over all combinations of inspectors of cardinality $s \geq q$ that satisfy the previous condition:

$$\sum_{d=0}^{n} \left\{ \sum_{s=q}^{m} \left[\sum_{A \subseteq \{1,\ \ldots,\ m\},\ \ |A|=s} \left\{ \left[\prod_{i \in A} R_i\,(c, d, a_i,\ b_i) \right] \cdot \left[\prod_{i \notin A} (1 - R_i\,(c, d, a_i,\ b_i)) \right] \right\} \right] \cdot \mathbb{P}\,(d) \right\} \tag{9}$$

s being the numerosity of the subset (A) of inspectors satisfying the imposed condition and q being an imposed threshold for the previous parameter. In order to define all possible combinations in which $q \leq s \leq m$, a contingency table covering them should be considered. For example, if we have $m = 5$ inspectors and set $q = 3$, we have the $\sum_{s=q}^{m} \binom{m}{s} = 10 + 5 + 1 = 16$ possible combinations shown in Table 1.

Two particularly important cases are: the one where $q = m$, which determines the condition of *unanimity*, where Equation (9) reduces to Equation (8); the one where $q = \left\lfloor \frac{m}{2} \right\rfloor + 1$, which determines the condition of *majority*.

It is noted that the condition of classifying no more than c defectives coincides with the SSP's LDD at the local level of the individual inspector. In this case, the expression in Equation (9) would correspond precisely to the P_a for the multiple-inspector sampling plan in the presence of inspection errors by the individual inspectors.

Moreover, it is easily verified that if $a_i \to 0$ and $b_i \to 0\ \forall i$, then the possible values of the term R_i (cf. Equation 6) are:

$$R_i\,(c, d, a_i,\ b_i) \to \begin{cases} 0 & \text{if } d > c \\ 1 & \text{if } d \leq c \end{cases} \tag{10}$$

Consequently, the acceptance probabilities in Equations (8) and (9) would coincide with that of an 'ideal' SSP with c, confirming that, in the absence of errors, the multiple-inspector model falls back to the ideal case. Thus, in this ideal condition, the number of inspectors, the aggregation criterion (and even the sequence, as we will see later on) has no effect on the OC curve, which

TABLE 1 | Example of a contingency table for $m = 5$ inspectors, covering all possible combinations in which at least $q = 3$ inspectors fulfil a given condition of interest. 'Y' and 'N' respectively indicate that the condition is fulfilled or not fulfilled for a certain inspector.

Inspector 1	Inspector 2	Inspector 3	Inspector 4	Inspector 5	$s = \|A\|$
Y	Y	Y	N	N	3
Y	Y	N	Y	N	3
Y	Y	N	N	Y	3
Y	N	Y	Y	N	3
Y	N	Y	N	Y	3
Y	N	N	Y	Y	3
N	Y	Y	Y	N	3
N	Y	Y	N	Y	3
N	Y	N	Y	Y	3
N	N	Y	Y	Y	3
Y	Y	Y	Y	N	4
Y	Y	Y	N	Y	4
Y	Y	N	Y	Y	4
Y	N	Y	Y	Y	4
N	Y	Y	Y	Y	4
Y	Y	Y	Y	Y	5

remains the same as in the ideal SSP. The complete formulae of the acceptance probability as function of N, n, c, m, a_i and b_i (for $i = 1, \ldots, m$) are provided in Appendices A.2.1–A.2.4.

We note that the model introduced so far employs a formalism different from the one most widely used in the literature, which is based on the concept of *apparent defectiveness* (p_i', c.f., the beginning of Section 4). This raises a natural question: can the approach based on the apparent defectiveness be adapted to the multiple-inspector case? Conceptually, one might define a (local) probability of acceptance (P'_{a_i}) for each i-th inspector (cf. Equation 4) and then aggregate these probabilities into a global lot-acceptance probability. The following expression gives the probability that at least q out of the m inspectors simultaneously reach an acceptance decision:

$$P'_a = \sum_{\forall A \mid q \leq s \leq m} \left(\sum_{A \subseteq \{1, \ldots, m\},\, |A|=s} \left\{ \left[\prod_{i \in A} P'_{a_i} \right] \cdot \left[\prod_{i \notin A} \left(1 - P'_{a_i}\right) \right] \right\} \right) \tag{11}$$

As previously discussed, the subset A refers to the s out of m inspectors who perform a local LDD, with q being the imposed threshold for the minimum number of such inspectors required. Consistent with Equation (9), this model accommodates both types of column-wise aggregation criteria: when $q = m$, the condition corresponds to *unanimity*; when $q = \lfloor \frac{m}{2} \rfloor + 1$, the condition corresponds to *majority*.

A primary limitation of the alternative analytical model in Equation (11) is that it applies exclusively to Type-B OC curves, as it relies on the binomial approximation under the assumption that $n \ll N$. But is this model truly consistent with the inspection scenario under consideration? Unfortunately, the answer is negative. The model overlooks a fundamental aspect of the actual setting—namely, that all m inspectors independently assess the *same* sample of n product units. In contrast, the alternative model implicitly assumes that each i-th inspector independently inspects a different random sample of n units drawn from the lot. Consequently, no constraint enforces that inspectors share a common sample and the model effectively treats their assessments as mutually independent. This assumption diverges significantly from the practical scenario of interest, where all inspectors must examine exactly the same set of sample units.

A counterexample further confirms the inadequacy of the alternative model in Equation (11). In the ideal case with no inspection errors (i.e., $a_i = b_i = 0\ \forall i$), all inspectors would make identical local LDDs and thus the global LDD—regardless of the column-wise aggregation criterion used (*majority* or *unanimity*)—should match that of the 'ideal' SSP, implying $P'_a = P_a$. However, under the alternative model in Equation (11), each inspectors' local acceptance probability equals P_a (i.e., $P'_{a_i} = P_a\ \forall i$) and the global probability becomes $P'_a = \sum_{q=s}^{m} \left[\binom{q}{m} \cdot P_q^a \cdot (1 - P_a)^{m-q} \right] \neq P_a$.

4.2 | Case of Aggregation Sequence First by Columns

Let us consider the other aggregation sequence, in which the (*local*) conformity assessments related to each individual sample unit (reported in a specific j-th column of the assessment matrix, cf. Figure 1 (1)) are first aggregated into a *global* conformity assessment; subsequently, the n different global assessments are aggregated by applying the SSP's LDD.

Let us proceed with a 'backward' description, assuming that the column-wise aggregation—to which we will return later—has been performed and that the aggregation by row of the *global* conformity assessments (✓/✗) of each j-th unit must then be performed. Assuming the sample consists of n units, of which d are 'true' defective, the remaining $(n-d)$ units will be 'true' conforming. The number k of units classified in global terms as ('true' and/or 'false') defectives is—as seen in Section 4.1—given by the sum of the number of 'true' defectives (k_1) and that of 'false' defectives (k_2). The probability that, at the level of global conformity assessments, exactly k conforming units are identified is given by:

$$P\,(k|d) = \sum_{j=\max[0,\ k-(n-d)]}^{\min(k,\ d)} [P\,(k_1 = j)\cdot P\,(k_2 = k-j)] \tag{12}$$

which corresponds to the convolution of the two binomial distributions: $k_1 \sim Bin(d, 1-b)$, that is, the distribution of the number of 'true' defectives that are globally (and correctly) classified as such, and $k_2 \sim Bin(n-d, a)$, that is, the distribution of the number of 'true' conforming units that are globally misclassified as defectives. Thus, Equation (12) can be reformulated as:

$$P\,(k|d) = \sum_{j=\max[0,\ k-(n-d)]}^{\min(k,\ d)} \left\{ \left[\binom{d}{j} \cdot (1-b)^j \cdot b^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot a^{(k-j)} \cdot (1-a)^{(n-d)-(k-j)} \right] \right\} \tag{13}$$

It can be seen that Equation (13) is almost identical to the second formula in Equation (5), except for the two terms a and b, instead of a_i and b_i, respectively. The term a depicts the probability of misclassifying a conforming unit as 'defective' at a *global* level; consequently, $(1-a)$ will be the probability of classifying a conforming unit correctly. The term b depicts the probability of misclassifying a defective unit as 'conforming' at a *global* level; consequently, $(1-b)$ will be the probability of classifying the defective unit correctly.

In order to obtain a global LDD of *pass*, resulting in the acceptance of the entire lot, it is necessary that the defectives are $k \le c$. The probability that the group of inspectors will classify no more than c defectives out of n, in case the sample includes d ('true') defective units is thus:

$$\sum_{k=0}^{c} P\,(k|d) \tag{14}$$

Finally, to construct the acceptance probability (P_a) over the entire range $d = 0, 1, \ldots, n$, the total probability theorem is again used [32], obtaining:

$$\sum_{d=0}^{n} \left\{ \left[\sum_{k=0}^{c} P\,(k|d) \right] \cdot \mathbb{P}\,(d) \right\} \tag{15}$$

where $\mathbb{P}(d)$ is the probability that the sample contains d 'true' defective units, based on the binomial or hypergeometric distribution, depending on whether the condition $n << N$ is verified or not (cf. Equations 1 and 2).

Let us now take a step back and return to the previously left-out terms a and b of Equation (13). The term a represents the aggregate probability that a 'true' conforming unit will be misclassified as 'defective' at a global level, taking into account the corresponding error probabilities (a_i) at the (local) level of individual inspectors. If the criterion of *unanimity* is used for the column-wise aggregation of the inspectors' local conformity assessments, the probability that all m inspectors correctly classify a given j-th unit as 'conforming' is:

$$\prod_{i=1}^{m} (1-a_i) = 1-a \tag{16}$$

Note that the quantity at the second term $(1-a)$ is the aggregated probability that, after aggregating the inspectors' local conformity assessments by unanimity, the correct classification of the (true) conforming unit as such will occur. From Equation (16), it can be obtained:

$$a = 1 - \prod_{i=1}^{m} (1-a_i) \tag{17}$$

We note that aggregation by unanimity is rather stringent as it results in a value of a systematically higher than the a_i values of the individual inspectors.

Let us now consider the more general case in which global conformity is no longer determined by unanimity (which is a very stringent condition) but if at least q out of m inspectors classify the given unit as conforming. In this more general (and less restrictive) condition, it follows:

$$\sum_{s=q}^{m} \left[\sum_{A\subseteq\{1,\ \ldots,\ m\},\ |A|=s} \left\{ \left[\prod_{i\in A} (1-a_i) \right] \cdot \left(\prod_{i\notin A} a_i \right) \right\} \right] = 1-a \tag{18}$$

where A is the subset of s inspectors over m, which correctly classify the ('true') conforming unit (cf. Equation 9). The specific case of aggregation via the *unanimity* criterion would imply $q = m$, $A = \{1, 2, \ldots, m\}$, $s = |A| = m$, causing Equation (18) to reduce to Equation (16). In the specific case of aggregation by the majority criterion, however, we have $q = \left\lfloor \frac{m}{2} \right\rfloor + 1$.

Let us now focus on term b. If a unit is 'truly' defective, the i-th inspector will have a probability $(1-b_i)$ of correctly classifying it as 'defective', and a probability b_i of misclassifying it as 'conforming'. If the *unanimity* aggregation criterion is adopted, there would only be a *global* misclassification of the 'truly' defective unit as 'conforming' if all inspectors were wrong at the same time. Such an event would have a probability:

$$\prod_{i=1}^{m} b_i = b \tag{19}$$

By definition, the *global* probability b will tend to be very low and certainly lower than the probability b_i of the individual i-th inspector.

Let us now consider the more general case in which global conformity is determined no longer by unanimity but if at least q out of m inspectors misclassify the unit of interest as being 'conforming'. In this condition, one has:

$$\sum_{s=q}^{m}\left[\sum_{A\subseteq\{1,\ \dots,\ m\},\ \ |A|=s}\left\{\left[\prod_{i\in A} b_i\right]\cdot\left[\prod_{i\notin A}(1-b_i)\right]\right\}\right]=b \tag{20}$$

In the specific case of unanimity, where $q = m$, $A = \{1, 2, \dots, m\}$, $s = |A| = m$, Equation (20) reduces to Equation (19). In addition, note that the case of no inspection errors by all inspectors—which means $a_i = 0$ and $b_i = 0\ \forall i$—determines $a = 0$ and $b = 0$, regardless of whether the column-wise aggregation criterion by majority or unanimity is applied. In this specific case, Equation (13) would reduce to:

$$P(k|d)=\begin{cases}1, & \text{if } k=d\\ 0, & \text{if } k\neq d\end{cases} \tag{21}$$

Moreover, it follows that if $d \le c$, the entire lot is accepted (probability 1); otherwise, it is rejected (probability 0). Under this condition, Equation (15) would reduce to the P_a formula of an 'ideal' SSP with parameters n and c, as shown in Equations (1) and (2). This means that, in the absence of inspection errors, neither the number of inspectors, nor the aggregation criterion, nor the aggregation sequence have any effect on the OC curve, which remains the same as the 'ideal' SSP.

Similar to the above, $\mathbb{P}(d)$ is the probability that the sample actually contains d 'true' defective units, based on the binomial or hypergeometric distribution, depending on the satisfaction of the condition $n << N$ (cf. Equations 1 and 2) [32]. The complete formulae of P_a as function of N, n, c, m, a_i and b_i (for $i = 1, \dots, m$) are given in Appendices A.2.5–A.2.8.

At this point, it is worth exploring whether an alternative analytical approach—grounded in the concept of *apparent defectiveness* (cf. the beginning of Section 4)—might serve as an approximation to the rigorous model so far developed. Following this direction, one could introduce global error rates (a and b), derived from the aggregation of local inspector error rates (a_i and b_i), according to the chosen column-wise criterion (*majority* or *unanimity*). These global error rates, as defined in Equations (18) and (20), can then be used to compute a (global) 'apparent' defectiveness (p''), reflecting a distortion of the 'true' defectiveness p:

$$p'' = (1-b)\cdot p + a\cdot(1-p) \tag{22}$$

The double-prime symbol (″) indicates that this *apparent* defectiveness is distinct from the one previously defined for individual inspectors (cf., p'_i in Equation 3). Furthermore, the absence of the subscript 'i' implies that this apparent defectiveness is defined at a *global* level, taking into account the aggregated behaviour of all m inspectors. Substituting p'' into the classical binomial acceptance formula (Equation 2) yields an approximate acceptance probability (P''_a):

$$P''_a=\sum_{d=0}^{c}\binom{n}{d}\cdot(p'')^{d}\cdot(1-p'')^{n-d} \tag{23}$$

Numerical simulations confirm that, when $\mathbb{P}(d)$ is evaluated with the binomial distribution, the model in Equation (23) yields results virtually indistinguishable from those of Equation (15). Accordingly, Equation (23) can be viewed as a simplified substitute for the more rigorous Equation (15), although its applicability is limited to Type-B OC curves.

It is worth noting that, while the model in Equation (15) explicitly accounts for the individual effects of 'false conforming' and 'false defective' classifications (as driven by the corresponding error rates), the approximated model in Equation (23) incorporates an 'apparent' defectiveness (p''), which already embeds both types of errors. Despite these conceptual differences, the two analytical approaches ultimately converge to equivalent results.

Next sections will provide some practical examples of OC-curve construction, discussing how the choices of (i) the aggregation sequence and (ii) the column-wise aggregation criterion can affect the discriminatory power of the resulting multiple-inspector sampling plan.

5 | Examples of Construction of OC Curves

This section provides some practical examples of the construction of OC curves for a multiple-inspector sampling plan in the presence of inspection errors, as well as a preliminary discussion of the implications of different sequences and aggregation criteria.

Referring back to the example in Section 3, let us assume that a *haute-couture* company receives from a supplier production lots of $N = 15$ garments each. To perform the acceptance inspection, a SSP is used with parameters $n = 3$ and $c = 1$, established based on the AQL = 1.5% value chosen according to ISO/DIS 2859-1:2025 [13]. In the 'ideal' case of no inspection errors (and single inspector), this SSP would result in the same Type-A OC curve illustrated in Figure A1 (see Section 3.1). However, the *haute-couture* context can be characterized by inspections with inherently subjective features and non-infrequent inspection errors, which suggest the involvement of five inspectors ($m = 5$) to inspect the same garments in parallel. Consequently, the actual OC curve of the plan (which includes multiple inspectors with relevant inspection errors) may differ even considerably from that of the original SSP.

Table 2 shows the profiles of five fictitious inspectors, each with different levels of experience and corresponding propensities to make errors. For each inspector, the error rates a_i (probability of 'false defectives') and b_i (probability of 'false conforming' units) are reported for three scenarios of increasing error rates—*low* (L), *moderate* (M), and *high* (H)—reflecting decreasing familiarity with the items under inspection. In scenarios M and H, the error rates are approximately 5 and 10 times higher, respectively, than in scenario L, in order to emphasise the impact of severe inspection errors. While scenarios L and M remain within plausible ranges, scenario H employs deliberately inflated error levels, in some cases reaching several tens of percent, to stress-test the robustness of multiple-inspector sampling plans under extreme error conditions.

5.1 | Aggregation Sequence First by Rows

Consider the case where the content of the assessment matrix is aggregated with a sequence *first by rows (then by column)*. Using

TABLE 2 | Profiles and error rates (a_i and b_i) of five hypothetical inspectors (i_1 to i_5) across three scenarios: *low* (L) error rates, assuming that inspectors have great familiarity with the inspected items, *moderate* (M) error rates, assuming that inspectors have partial familiarity with the inspected items, and *high* (H) error rates, assuming that inspectors have minimal familiarity with the inspected items. For the same scenario, the more experienced inspectors (e.g., i_1 and i_2) tend to show lower error rates, whereas the less experienced inspectors (e.g., i_3, i_4 and i_5) show significantly higher error rates. Furthermore, the error rates in scenarios M and H are respectively five and ten times higher than in scenario L, for all inspectors.

Inspector	Profile	a_i			b_i		
		L	***M***	***H***	***L***	***M***	***H***
i_1	i_1 is an experienced inspector with a long career in the *haute-couture* industry. i_1 rarely makes mistakes, thanks to a deep knowledge of materials and production techniques.	1.2%	6%	12%	2.3%	11.5%	23%
i_2	i_2 is an expert with a keen eye for detail. i_2 rarely makes mistakes, due to an in-depth understanding of the production process and conformity criteria.	1.7%	8.5%	17%	1.1%	5.5%	11%
i_3	i_3 has some experience in inspection but is not entirely familiar with certain fabrics and complex manufacturing processes. i_3 tends to attribute some false defects but is good at identifying actual nonconformities.	6.0%	30%	60%	2.8%	14%	28%
i_4	i_4 is a novice and struggles to detect certain defects, although tends to report nonconformities even where there are none.	5.6%	28%	56%	8.4%	42%	84%
i_5	i_5 is the least experienced of the team and often fails to identify the most hidden defects. Unfortunately, insecurity not infrequently leads i_5 to report false defectives.	4.5%	22.5%	45%	8.9%	44.5%	89%

the inspectors' error rates (Table 2) and imposing $c = 1$, P_a is calculated as a function of $p = D/N$, based on the hypergeometric distribution (Type-A OC curve). In particular, Equation (8) can be used in the case of column-wise aggregation with the *unanimity* criterion and Equation (9) in the case of column-wise aggregation with the *majority* criterion. Figure 2 shows the resulting OC curves for the above three error rate scenarios (L, M, H), compared with the OC curve of the 'ideal' SSP (dashed line).

Let us first focus on the OC curves in the case of applying the column-wise *majority* aggregation criterion (see the three upper graphs in Figure 2). In scenario L, with relatively small errors, the OC curve appears completely superimposed on that of the 'ideal' SSP. Even in scenario M, with medium to high errors, a close proximity between the two OC curves is maintained. In scenario H, with enormous error rates, the two OC curves deviate more. In particular, for low p-values, a_i-related errors (propensity to classify false defectives) lead to some conforming units being classified as defective, reducing the P_a (lower curve). For high values of p, b_i-related errors (propensity to classify 'false' conforming units) can generate an 'upward' effect on the curve, partially reversing the typical monotonically decreasing trend. It can be shown that, by taking the error rates to extremes, in the limiting case where $a_i = b_i = 100\%$, a curve complementary to the 'ideal' error-free OC curve would even be obtained (see Figure 3).

Let us now consider the OC curves in the case of applying the column-wise *unanimity* aggregation criterion (cf., the three graphs at the bottom of Figure 2). Apart from the L error-rates scenario, in which the OC curve remains relatively close to that in the case of no inspection errors, in the other scenarios the OC curve drops dramatically, with low P_a values even at very small p values. This suggests that aggregation by unanimity is unduly severe in the presence of non-negligible inspection errors. This aggregation criterion may therefore only be reasonable when inspectors have a relatively low propensity to make errors and extreme caution is desired.

5.2 | Aggregation Sequence First by Columns

Let us now turn to the aggregation sequence *first by columns (then by row)* (cf. Section 4.2). Again, the OC curves are constructed for the same parameters $N = 15$, $n = 3$, $c = 1$ and $m = 5$, comparing the column-wise aggregation criteria by *majority* and *unanimity* (see Figure 4).

Overall, the OC-curves are very similar to those described in the case of aggregation sequence first by rows (cf. Section 5.1). The OC curves in the case of column-wise aggregation by *unanimity* appear even more severe and deviate more from those ones in the absence of errors, except in scenario L. In the case of column-wise aggregation by *majority*, the OC curves are almost identical.

5.3 | Further Remarks

All the above examples refer to Type-A OC curves (based on hypergeometric distribution), considering finite lots of $N = 15$ units each, sample parameters $n = 3$, $c = 1$ and $m = 5$ inspectors.

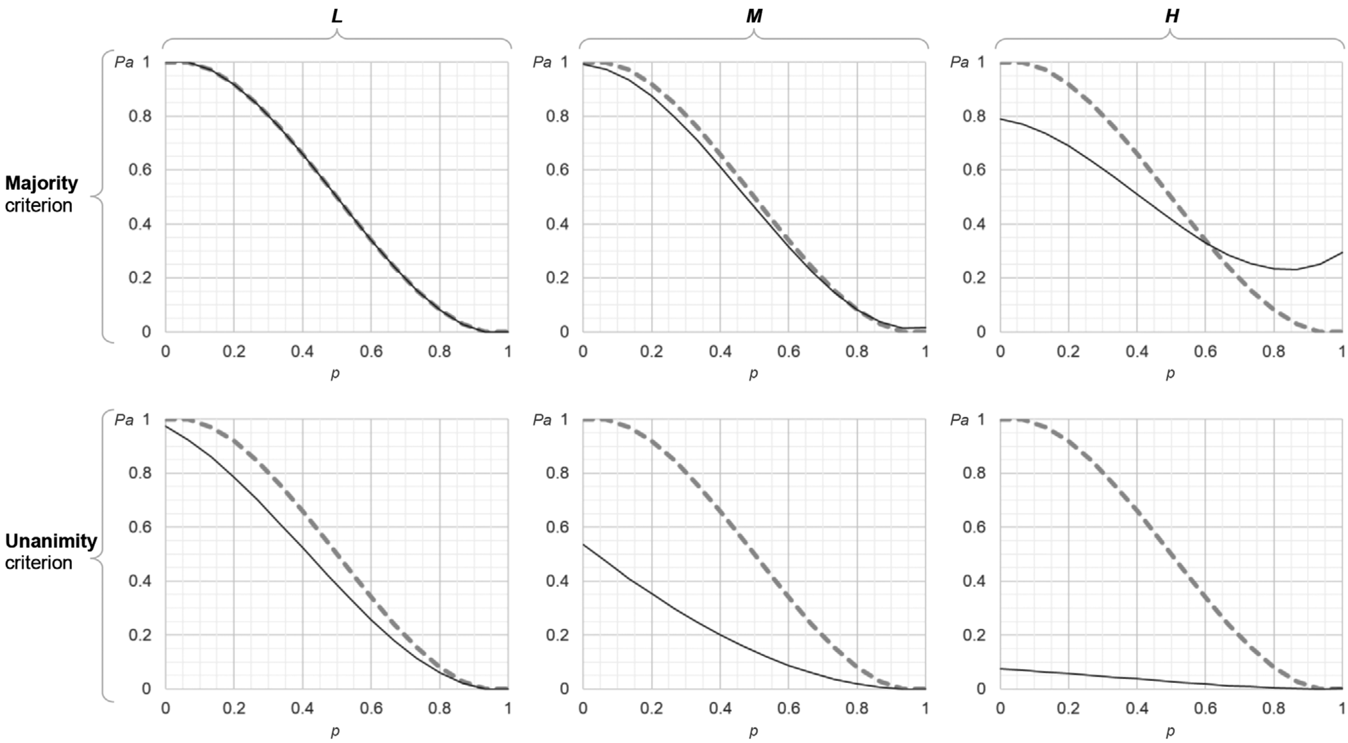

FIGURE 2 | Type-A OC curves (based on the *hypergeometric* distribution) for a SSP ($N = 15$, $n = 3$ and $c = 1$) involving $m = 5$ inspectors, using the sequence of aggregation *first by rows*. Three scenarios are considered, characterized by inspectors with increasing inspection error rates (a_i and b_i): *L*, *M* and *H* (cf. Table 2). The three upper graphs refer to the column-wise aggregation by *majority*, while the three lower graphs correspond to the column-wise aggregation by *unanimity*. The dashed OC curve represents the ('ideal') SSP with no inspection errors.

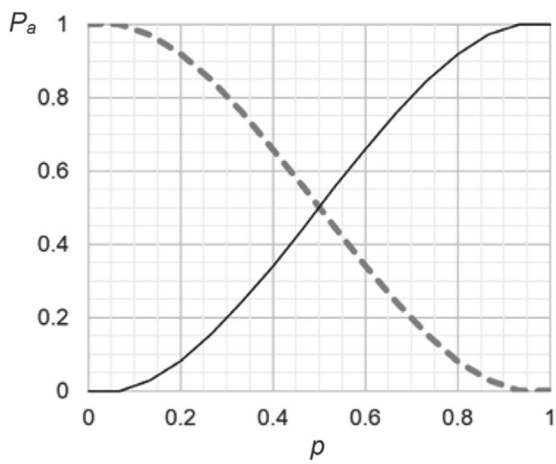

FIGURE 3 | Example of (Type-A) OC curves in the extreme case of $a_i = b_i = 100\%\ \forall i$, for a SSP with $N = 15$, $n = 3$, $c = 1$ and $m = 5$ inspectors. The OC curve (solid line) is complementary to that of the SSP with no inspection errors (dashed line), as the outcome of each inspection is systematically 'overturned'.

Type-B OC curves can be also constructed, approximating the distribution with a binomial (i.e., under the assumption that N is relatively large that the condition $n \ll N$ is met) and keeping the same values of the other parameters ($n = 3$, $c = 1$ and $m = 5$). The results, highlighted in Figures A2 and A3 (in Appendix A.3) being almost entirely superimposable on Figures 3 and 4, without showing any particular deviations worthy of further investigation.

In summary, the proposed examples—although limited to a single reference SSP—highlights some general aspects. First, the column-wise aggregation of *majority* tends to keep the OC curve relatively stable and close to the 'ideal' OC curve in the absence of errors, as long as the inspectors' error rates are not excessively high. In contrast, the column-wise aggregation of *unanimity* produces OC curves that tend to be excessively severe, unless the inspectors' error rates are relatively small (scenario *L*).

In addition, there does not seem to be any substantial difference between the results obtained by applying the two possible aggregation sequences (i.e., the sequence *first by rows* or the sequence *first by columns*); in other words, this aspect seems to affect the final OC curve much less than the column-wise aggregation criterion (*majority* vs. *unanimity*). When applying the column-wise aggregation criterion for majority, we finally notice that error rates a_i tend to make the OC curve lower for low values of p while error rates b_i tend to make it higher for high values of p. In extreme cases of very high errors, a 'reversal' of the characteristic downward trend of the OC-curve is even observed.

From a practical point of view, adopting a multiple-inspector plan with column-wise aggregation by majority can be a very effective strategy to contain the negative impact of inspection errors by individual inspectors: in this way, a relatively 'robust' OC curve can be obtained anyway, because it is close to what it would be in the absence of errors, which is generally agreed between the parties involved (typically *supplier* and *customer*).

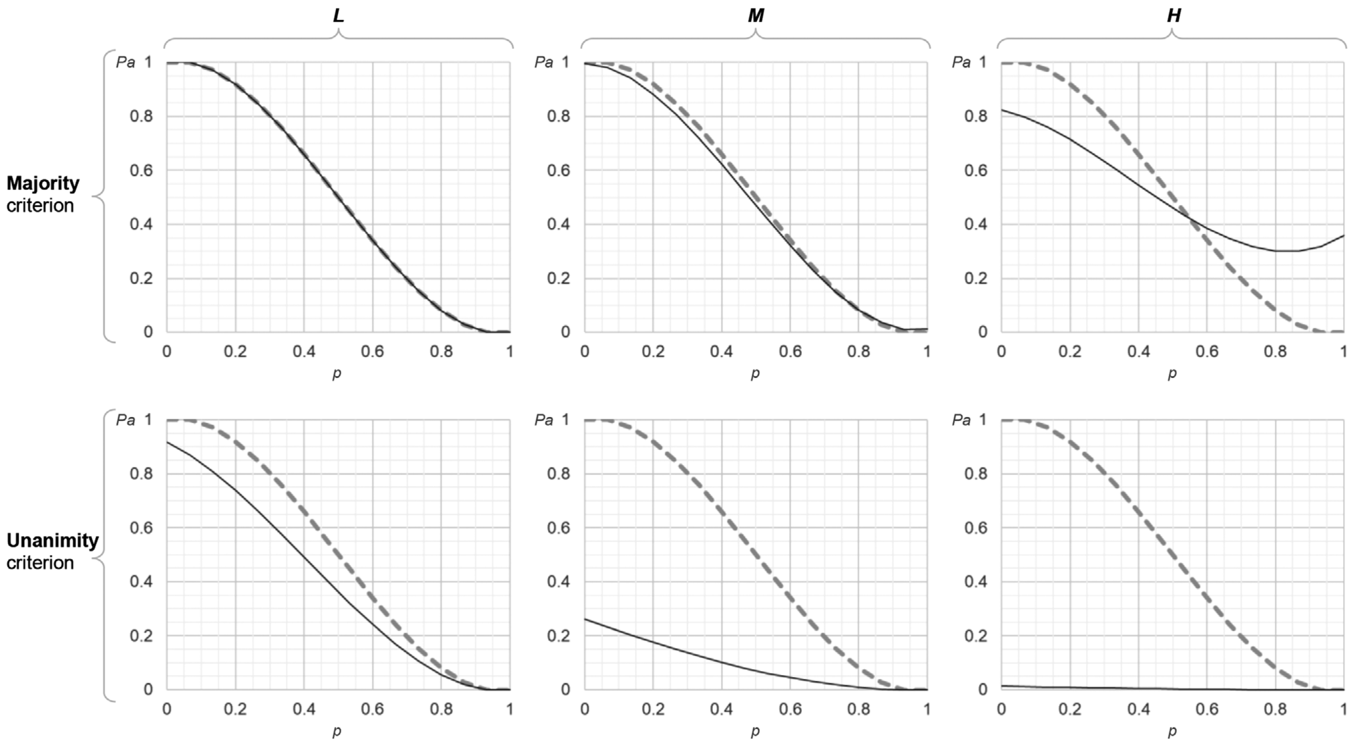

FIGURE 4 | Type-A OC curves (based on the *hypergeometric* distribution) for a SSP ($N = 15$, $n = 3$ and $c = 1$) involving $m = 5$ inspectors, using the sequence of aggregation *first by columns*. Different scenarios are considered, characterized by inspectors with increasing inspection error rates (a_i and b_i): *L*, *M* and *H* (cf. Table 2). The three upper graphs refer to column-wise aggregation by *majority*, while the three lower graphs correspond to column-wise aggregation by *unanimity*. The dashed OC curve represents the 'ideal' SSP scenario (single inspector, with no inspection errors).

In contrast, aggregation by unanimity, although it seems to offer the utmost caution, risks lowering the P_a excessively, even for very low values of p.

6 | Conclusions

This paper introduced an analytical framework for constructing the OC curve of a SSP in a multiple-inspector context with inspection errors. The classical SSP's parameters (n, c) were extended to include multiple inspectors (m), each with defined error rates a_i and b_i (i.e., the i-th inspector's probability of assigning false defectives and false conforming units, respectively). The proposed analytical models encompass both Type-A OC curves (based on the hypergeometric distribution) and Type-B OC curves (based on a binomial approximation), thereby maintaining mathematical rigour even in the multiple-inspector setting (cf., ***RQ1***) [11, 32]. This extension helps bridge a gap in the acceptance sampling literature and provides a basis for quantitatively evaluating how inspector errors and conformity-assessment aggregation policies affect the performance of sampling plans [8].

As part of this study, two alternative *sequences* for aggregating the inspectors' conformity assessments were examined: one *first by rows*, and the other one, *first by columns*. Each sequence was combined with two different column-wise aggregation criteria: *majority* criterion (requiring a consensus of more than half of the inspectors) and a *unanimity* criterion (requiring all inspectors to agree on the unit conformity). The analysis revealed that the choice of the column-wise aggregation criterion has a far greater influence on the OC curve than the choice of the aggregation sequence (cf., ***RQ2***). Specifically, the column-wise *majority* criterion yields an OC curve that remains close to the 'ideal' one (with no inspection errors), as long as inspectors' error rates are moderate, whereas the *unanimity* criterion results in an exaggeratedly severe OC curve, except in the extreme case where all inspectors' error rates (a_i, b_i) approach zero.

An important result of this preliminary study is the relative robustness observed with the *majority* criterion, which is probably due to a statistical compensation effect among the inspectors' independent errors. In practical terms, this means that even when inspectors exhibit relatively significant error rates (a_i and b_i)—and these rates are not estimated with high accuracy or through a rigorous evaluation of their associated uncertainty—the mere use of multiple inspectors combined with a column-wise majority aggregation ensures that the resulting OC curve will remain very close to the ideal one. This is a particularly relevant finding, since in many practical situations estimating a_i and b_i accurately can be complex and challenging [8]. By contrast, the *unanimity* criterion, although maximally cautious, can make the OC curve exaggeratedly severe even at very low p levels.

The analytical model developed in this research can be used to calculate the OC curve for any given set of SSP's parameters (n, c), number of inspectors m, and individual inspector error rates (a_i, b_i). This capability enables quality engineers to quantitatively assess the risks to suppliers and customers under a variety of multiple-inspector configurations and to design and implement appropriate inspection strategies (cf., ***RQ3***). Appendix A.2 provides complete formulae for constructing OC curves in the exemplified cases.

The multiple-inspector logic at the core of this study remains relevant even as inspection processes become increasingly automated. Modern quality-control systems often employ multiple independent sensors/devices, replicating the redundancy of human multi-inspector schemes to enhance detection reliability; for example, using data-fusion across multiple sensors in smart manufacturing [35]. In such cases, each 'automated inspector' may have distinct error tendencies, and combining their outputs—analogous to parallel human inspections—can improve defect detection and reduce the risk of oversight.

Despite its usefulness, the proposed model assumes that the errors by different inspectors are independent. In reality, correlated error patterns might emerge even in the absence of any interaction among inspectors—for example, if several inspectors share a tendency to misclassify the same type of flaw, their errors will not be statistically independent. This may motivate relaxing the independence assumption in future research and extending the model to account for correlated error tendencies among inspectors. Addressing inter-inspector correlation in this way would broaden the model's applicability to a wider range of inspection scenarios. Furthermore, it would be worthwhile to perform a more structured comparison of the two aggregation sequences (*first by rows* vs. *first by columns*)—which, according to the present study, have only a marginal influence—and to revisit the model in cases where inspectors' error rates are allowed to vary rather than being treated as fixed values.

Finally, the methodology and results presented here are directly applicable to high-value, high-risk inspection contexts. For instance, multiple-inspector sampling plans can be particularly beneficial in industries such as aerospace, defence or luxury manufacturing, where products may exhibit high variability and exceptionally stringent quality requirements.

Nomenclature

Acronyms

AQL	acceptance quality level
H	high error-rate scenario
ISO	International Organization for Standardization
L	low error-rate scenario
LDD	lot-disposition decision
LTPD	lot tolerance percent defective
M	moderate error-rate scenario
OC curve	operating characteristic curve
RQ	research question
SSP	single sampling plan

Variables

a_i	probability of 'false' defectives by the i-th inspector
A	subset of inspectors classifying a unit as ('truly' of 'falsely') conforming
b_i	probability of 'false' conforming units by the i-th inspector
c	acceptance number of a single sampling plan
d	true number of defective units within the sample
D	true number of defective units within the whole lot
$i_1, i_2, \ldots$	generic inspector involved in the sample inspection
k	number of defective units identified by the i-th inspector in the sample (i.e., $k_1 + k_2$)
k_1	number of 'true' defectives identified by the i-th inspector in the sample
k_2	number of 'false' defectives identified by the i-th inspector in the sample
n	sample size of a single sampling plan
N	lot size
m	number of inspectors involved in the sample inspection
p	*defectiveness* or *fraction nonconforming* or *fraction defective* of a lot (i.e., D/N);
p'_i	'Apparent' defectiveness of a lot, from the (local) perspective of the i-th inspector
P_a	probability of acceptance of a lot
P'_{a_i}	probability of acceptance of a lot, from the (local) perspective of the i-th inspector
$\mathbb{P}(d)$	probability that a sample contains exactly d defectives
q	number of inspectors classifying at least r defectives
r	conventional threshold on the k value
s	conventional threshold on the q value
$u_1, u_2, \ldots$	generic product unit of the inspected sample
✓	result of conformity in the assessment of a single product unit
✗	result of nonconformity in the assessment of a single product unit

Author Contributions

The authors have provided an equal contribution to the drafting of the paper.

Acknowledgments

This research was carried out under the MICS (Made in Italy, Circular and Sustainable) Extended Partnership and partially funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza – Missione 4, Componente 2, Investimento 1.3, D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Open access publishing facilitated by Politecnico di Torino, as part of the Wiley - CRUI-CARE agreement.

Ethics Statement

The authors respect the Ethical Guidelines of the Journal.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Endnotes

[1] The adjective 'apparent' indicates that the defectiveness perceived by a specific inspector does not necessarily match the 'true' defectiveness, but may be distorted by inspection errors, as formalised in Section 3.2. In other words, it is as if the inspector's perception of reality were altered by a set of lenses, the distortion of which increases with the inspector's propensity to commit inspection errors. The contrasting use of the adjectives 'apparent' and 'true' to describe the defectiveness of samples/lots has also been adopted in other studies in the scientific literature [11, 14].

[2] The supplier's perspective is typically focused on a relatively low defectiveness (p) rate—that is, the *Acceptable Quality Level* (AQL)—at which the lot's probability of acceptance should be very high, tolerating only a small risk of rejection, that is, the so-called 'supplier's risk' α, ideally on the order of only a few percentage points. On the other hand, the customer concentrates on a higher but still tolerable p rate—the *Lot Tolerance Percent Defective* (LTPD)—which is desired to correspond to a relatively low probability of acceptance, (i.e., the so-called 'customer's risk' β, also ideally limited to a few percentage points. The combination of these dual perspectives leads to the definition of an OC curve that, at least as a first approximation, satisfies both supplier and customer, serving as an agreement [11]. In the rest of this paper, however, we will not refer explicitly to the preliminary phase of defining the AQL, LTPD and the associated α and β risks. Instead, we will focus on how OC curves defined for single inspectors can change in the case of multiple inspectors, when taking inspection errors into account.

[3] *Gage Repeatability and Reproducibility* (GR&R) studies can be used to assess the variability of inspectors prior to actual inspections [7]. In this paper, however, no formal GR&R study was performed; instead, it is assumed that each inspector's error rates (a_i and b_i) are estimated through separate training trials on homologous items before the inspection phase, similar to the approach described by Franceschini et al. [8] in the *haute couture* context. The systematic integration of GR&R data is identified as a useful extension for future work.

References

1. A. Aurum, H. Petersson, and C. Wohlin, "State-of-the-Art: Software Inspections After 25 Years," *Software Testing, Verification and Reliability* 12 (2002): 133–154, https://doi.org/10.1002/stvr.243.

2. Y. H. Chun, "Improved Method of Estimating the Product Quality After Multiple Inspections," *International Journal of Production Research* 54, no. 19 (2016): 5686–5696, https://doi.org/10.1080/00207543.2015.1128128.

3. C. Mazza and J. L. Alvarez, "Haute Couture and Prêt-à-Porter: The Popular Press and the Diffusion of Management Practices," in *The Aesthetic Turn in Management* (Routledge, 2017), 157–178.

4. Y. H. Chun, "Improving Product Quality by Multiple Inspections: Prior and Posterior Planning of Serial Inspection Procedures," *IIE Transactions* 41, no. 9 (2009): 831–842, https://doi.org/10.1080/07408170802389324.

5. D. H. Cymek, "Effects of Blinded and Nonblinded Sequential Human Redundancy on Inspection Effort and Inspection Outcome in Low Prevalence Visual Search," *Scientific Reports* 14 (2024): 23003.

6. C. G. Drury, M. H. Karwan, and D. R. Vanderwarker, "The Two-Inspector Problem," *IIE Transactions* 18, no. 2 (1986): 174–181.

7. W. N. van Wieringen and J. de Mast, "Measurement System Analysis for Binary Data," *Technometrics* 50, no. 4 (2008): 468–478.

8. F. Franceschini, D. Maisano, and L. Mastrogiacomo, "Acceptance Sampling With Multiple Inspectors: Critical Aspects in Aggregating Individual Assessments," *Quality Engineering* 37, no. 2 (2025): 318–329.

9. J. Chandra and S. Schall, "The Use of Repeated Measurements to Reduce the Effect of Measurement Errors," *Iie Transactions* 20 (1998): 83–87.

10. E. G. Schilling and D. V. Neubauer, *Acceptance Sampling in Quality Control*, 2nd ed. (CRC Press, 2009).

11. D. C. Montgomery, *Introduction to Statistical Quality Control*, 8th ed. (Wiley, 2019).

12. R. S. Kenett, S. Zacks, and P. Gedeck, "Sampling Plans for Batch and Sequential Inspection," in *Industrial Statistics: A Computer-Based Approach With Python* (Springer International Publishing, 2023), 397–442.

13. ISO/DIS 2859-1:2025, Sampling Procedures for Inspection by Attributes – Part 1: Sampling Schemes Indexed by Acceptance Quality Limit (AQL) for Lot-by-Lot Inspection, Draft Int. Standard, Jan. 2025 (International Organization for Standardization, 2025).

14. S. O. Duffuaa and A. El-Ga'aly, "Impact of Inspection Errors on the Formulation of a Multi-Objective Optimization Process Targeting Model Under Inspection Sampling Plan," *Computers & Industrial Engineering* 80 (2015): 254–260.

15. F. Franceschini, D. Maisano, and L. Mastrogiacomo, "Empirical Analysis and Classification of Database Errors in Scopus and Web of Science," *Journal of Informetrics* 10, no. 4 (2016): 933–953.

16. J. C. Knight and A. Mayers, "An Improved Inspection Technique," *Communications of the ACM* 36, no. 11 (1993): 51–61.

17. W. E. Deming, *Out of the Crisis, Reissue* (MIT press, 2018).

18. W. G. Ferrell Jr. and A. Chhoker, "Design of Economically Optimal Acceptance Sampling Plans With Inspection Error," *Computers & Operations Research* 29, no. 10 (2002): 1283–1300.

19. E. B. Jamkhaneh, B. Sadeghpour-Gildeh, and G. Yari, "Inspection Error and Its Effects on Single Sampling Plans With Fuzzy Parameters," *Structural and Multidisciplinary Optimization* 43 (2011): 555–560.

20. T. Gadrich and E. Bashkansky, "A Bayesian Approach to Evaluating Uncertainty of Inaccurate Categorical Measurements," *Measurement* 91 (2016): 186–193.

21. F. Franceschini, D. Maisano, and L. Mastrogiacomo, *Rankings and Decisions in Engineering: Conceptual and Practical Insights*. International Series in Operations Research & Management Science Series (Springer International Publishing, 2022).

22. ISO 28596:2022, *Sampling Procedures for Inspection by Attributes – Two-Stage Sampling Plans for Auditing and for Inspection Under Prior Information* (International Organization for Standardization, 2022).

23. A. Wald, *Sequential Analysis* (John Wiley & Sons, 1947).

24. H. F. Dodge, "Skip-lot Sampling Plan," *Journal of Quality Technology* 9, no. 3 (1977): 143–145.

25. A. Hald, *Statistical Theory of Sampling Inspection by Attributes* (Academic Press, 1981).

26. M. Jaraiedi and W. H. Iskander, "Posterior Analysis of Inspector Accuracy," *International Journal of Quality & Reliability Management* 6, no. 2 (1989).

27. Y. H. Chun, "Economic Design of Multiple Inspection Plans via the Expectation–Maximization Algorithm," *Journal of the Operational Research Society* 74, no. 3 (2022): 762–776.

28. S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics* 6, no. 2 (1979): 65–70.

29. Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)* 57, no. 1 (1995): 289–300.

30. N. Das and S. K. Bhandari, "Bound on FWER for Correlated Normal," *Statistics & Probability Letters* 168 (2021): 108943.

31. O. Danila, S. H. Steiner, and R. J. MacKay, "Assessing a Binary Measurement System With Varying Misclassification Rates Using a

Latent Class Random Effects Model," *Journal of Quality Technology* 44, no. 3 (2012): 179–191.

32. S. M. Ross, *Introductory Statistics* (Academic Press (Elsevier), 2017).

33. R. J. Burke, R. D. Davis, F. C. Kaminsky, and A. E. P. Roberts, "The Effect of Inspector Errors on the True Fraction Non-Conforming: An Industrial Experiment," *Quality Engineering* 7, no. 3 (1995): 543–550, https://doi.org/10.1080/08982119508918802.

34. R. L. Trader and H. Fenwick Huss, "Prediction in the Presence of Imperfect Inspections," *Communications in Statistics-Simulation and Computation* 14, no. 2 (1985): 425–440.

35. J. Lee, B. Bagheri, and H. A. Kao, "A Cyber-Physical Systems Architecture for Industry 4.0-Based Manufacturing Systems," *Manufacturing Letters* 3 (2015): 18–23.

Appendix

A.1 Example of OC Curves for a SSP

Figure A1 illustrates the construction of Type-A and Type-B OC curves for a SSP with $N = 15$, $n = 3$ and $c = 1$ (cf. Section 3.1).

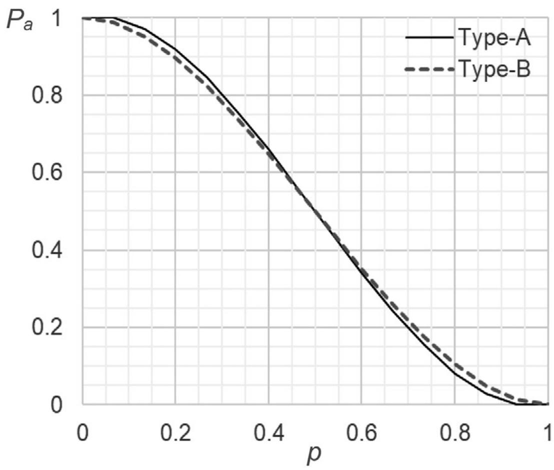

FIGURE A1 | Example OC curves for a SSP with parameters $N = 15$, $n = 3$ and $c = 1$. The Type-A OC curve (solid line) is based on hypergeometric distribution, while the Type-B OC curve (dashed line) is based on binomial distribution. The slight deviation between the two curves is attributable to the fact that the condition $n << N$ is not met in this case, and therefore the Type-B OC curve does not approximate the (exact) Type-A OC curve very well [32].

A.2 Complete Formulae of Exemplified OC Curves

This section details the formulae used to construct the OC curves exemplified in the article, depending on the sequence of aggregation (*first by rows* or *first by columns*), column-wise aggregation criterion (by *majority* or *unanimity*) and type (i.e., Type-A or Type-B) of OC curves.

A.2.1 Type-A OC Curves, Aggregation First by Rows, With Majority Column-Wise Aggregation

Starting from the expression of P_a in Equation (9), referring to the sequence of aggregation *first by rows* and the column-wise *majority* criterion, we substitute $r = n - c$, so as to refer to the parameters of the SSP, whose LDD is used in the aggregations by rows. Next, we substitute $\mathbb{P}\,(\boldsymbol{d}) = \dfrac{\binom{\boldsymbol{D}}{\boldsymbol{d}}\binom{\boldsymbol{N}-\boldsymbol{D}}{\boldsymbol{n}-\boldsymbol{d}}}{\binom{\boldsymbol{N}}{\boldsymbol{n}}}$ in order to obtain a Type-A OC curve based on the hypergeometric model (cf. Equation 1). Thus, it is obtained:

$$
\begin{aligned}
P_a = \sum_{d=0}^{n} \Bigg\{ \sum_{s=q}^{m} \Bigg[\sum_{A\subseteq\{1,\,\dots,\,m\},\ |A|=s} \Bigg\{ \Bigg[\prod_{i\in A} \Bigg(\sum_{k=n-c}^{n} \Bigg\{ \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)} \Bigg\{ \Bigg[\binom{d}{j} \cdot (1-b_i)^{j} \cdot b_i^{d-j} \Bigg] \cdot \Bigg[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1-a_i)^{(n-d)-(k-j)} \Bigg] \Bigg\} \Bigg\} \Bigg) \Bigg] \\
\cdot \Bigg[\prod_{i\notin A} \Bigg(1 - \Bigg[\sum_{k=n-c}^{n} \Bigg\{ \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)} \Bigg\{ \Bigg[\binom{d}{j} \cdot (1-b_i)^{j} \cdot b_i^{d-j} \Bigg] \cdot \Bigg[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1-a_i)^{(n-d)-(k-j)} \Bigg] \Bigg\} \Bigg\} \Bigg] \Bigg) \Bigg] \Bigg\} \Bigg] \cdot \frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}} \Bigg\}
\end{aligned}
\tag{A1}
$$

A.2.2 Type-B OC Curves, Aggregation First by Rows, With Majority Column-Wise Aggregation

This case is analogous to the previous one (Section A.2.1), the only difference being that one substitutes $\mathbb{P}\,(d) = \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d}$, in order to obtain a Type-B OC curve based on the binomial model (cf. Equation 2). It is thus obtained:

$$
\begin{aligned}
P_a = \sum_{d=0}^{n} \Bigg\{ \sum_{s=q}^{m} \Bigg[\sum_{A\subseteq\{1,\,\dots,\,m\},\;|A|=s} \Bigg\{ \Bigg[\prod_{i\in A} \Bigg(\sum_{k=n-c}^{n} \Bigg\{ \sum_{j=\max[0,\;k-(n-d)]}^{\min(k,\;d)} \Bigg\{ \Bigg[\binom{d}{j} \cdot (1-b_i)^j \cdot b_i^{d-j} \Bigg] \cdot \Bigg[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1-a_i)^{(n-d)-(k-j)} \Bigg] \Bigg\} \Bigg\} \Bigg) \Bigg] \\
\cdot \Bigg[\prod_{i\notin A} \Bigg(1 - \Bigg[\sum_{k=n-c}^{n} \Bigg\{ \sum_{j=\max[0,\;k-(n-d)]}^{\min(k,\;d)} \Bigg\{ \Bigg[\binom{d}{j} \cdot (1-b_i)^j \cdot b_i^{d-j} \Bigg] \cdot \Bigg[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1-a_i)^{(n-d)-(k-j)} \Bigg] \Bigg\} \Bigg\} \Bigg] \Bigg) \Bigg] \Bigg\} \Bigg] \cdot \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d} \Bigg\}
\end{aligned}
\tag{A2}
$$

A.2.3 Type-A OC Curves, Aggregation First by Rows, With Unanimity Column-Wise Aggregation

Starting from the expression of P_a in Equation (8), referring to the use of the aggregation sequence *first by rows* and the column-wise *unanimity* criterion, we substitute $r = n - c$, in order to refer to the parameters of the SSP, whose LDD is used in the aggregations by rows. We then substitute $\mathbb{P}\,(\boldsymbol{d}) = \dfrac{\binom{\boldsymbol{D}}{\boldsymbol{d}}\binom{\boldsymbol{N}-\boldsymbol{D}}{\boldsymbol{n}-\boldsymbol{d}}}{\binom{\boldsymbol{N}}{\boldsymbol{n}}}$ in order to obtain a Type-A OC curve based on the hypergeometric model (cf. Equation 1). Thus, resulting in:

$$
P_a = \sum_{d=0}^{n} \left\{ \left[\prod_{i=1}^{m} \sum_{k=n-c}^{n} \left\{ \sum_{j=\max[0,\;k-(n-d)]}^{\min(k,\;d)} \left\{ \left[\binom{d}{j} \cdot (1-b_i)^j \cdot b_i^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1-a_i)^{(n-d)-(k-j)} \right] \right\} \right\} \right] \cdot \frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}} \right\}
\tag{A3}
$$

A.2.4 Type-B OC Curves, Aggregation First by Rows, With Unanimity Column-Wise Aggregation

This case is analogous to the previous one (Section A.2.3), the only difference being that one substitutes $\mathbb{P}\,(d) = \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d}$, in order to obtain a Type-B OC curve based on the binomial model (cf. Equation 2). It is thus obtained:

$$
P_a = \sum_{d=0}^{n} \left\{ \left[\prod_{i=1}^{m} \sum_{k=n-c}^{n} \left\{ \sum_{j=\max[0,\;k-(n-d)]}^{\min(k,\;d)} \left\{ \left[\binom{d}{j} \cdot (1-b_i)^j \cdot b_i^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot a_i^{(k-j)} \cdot (1-a_i)^{(n-d)-(k-j)} \right] \right\} \right\} \right] \cdot \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d} \right\}
\tag{A4}
$$

A.2.5 Type-A OC Curves, Aggregation First by Columns, With Majority Column-Wise Aggregation

Starting with the expression for P_a in Equation (13) and using Equations (18) and (20)—which refer to a and b in the case of aggregation *first by columns*, under the column-wise *majority* criterion—we set $r = n - c$ to reflect the SSP's LDD for row-based aggregation. Next, we substitute $\mathbb{P}\,(\boldsymbol{d}) = \dfrac{\binom{\boldsymbol{D}}{\boldsymbol{d}}\binom{\boldsymbol{N}-\boldsymbol{D}}{\boldsymbol{n}-\boldsymbol{d}}}{\binom{\boldsymbol{N}}{\boldsymbol{n}}}$, in order to obtain a Type-A OC curve based on the hypergeometric model (cf. Equation 1). It is finally obtained:

$$
\begin{aligned}
P_a = \sum_{d=0}^{n} \Bigg\{ \Bigg[\Bigg[\sum_{k=0}^{c} \sum_{j=\max[0,k-(n-d)]}^{\min(k,d)} \Bigg\{ \Bigg[\binom{d}{j} \cdot \Bigg(1 - \sum_{s=\frac{m}{2}+1}^{m} \Bigg[\sum_{A\subseteq\{1,\dots,m\},|A|=s} \Bigg\{ \Bigg[\prod_{i\in A} b_i \Bigg] \cdot \Bigg[\prod_{i\notin A} (1-b_i) \Bigg] \Bigg\} \Bigg] \Bigg)^{j} \\
\cdot \Bigg\{ \sum_{s=\frac{m}{2}+1}^{m} \Bigg[\sum_{A\subseteq\{1,\dots,m\},|A|=s} \Bigg\{ \Bigg[\prod_{i\in A} b_i \Bigg] \cdot \Bigg[\prod_{i\notin A} (1-b_i) \Bigg] \Bigg\} \Bigg] \Bigg\}^{d-j} \Bigg] \Bigg] \cdot \Bigg[\binom{n-d}{k-j} \cdot \Bigg\{ 1 - \sum_{s=\frac{m}{2}+1}^{m} \Bigg[\sum_{A\subseteq\{1,\dots,m\},|A|=s} \Bigg\{ \Bigg[\prod_{i\in A} (1-a_i) \Bigg] \cdot \Bigg(\prod_{i\notin A} a_i \Bigg) \Bigg\} \Bigg] \Bigg\}^{(k-j)} \\
\cdot \Bigg(1 - \sum_{s=\frac{m}{2}+1}^{m} \Bigg[\sum_{A\subseteq\{1,\dots,m\},|A|=s} \Bigg\{ \Bigg[\prod_{i\in A} (1-a_i) \Bigg] \cdot \Bigg(\prod_{i\notin A} a_i \Bigg) \Bigg\} \Bigg] \Bigg)^{(n-d)-(k-j)} \Bigg] \Bigg] \Bigg\} \Bigg] \cdot \frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}} \Bigg\}
\end{aligned}
\tag{A5}
$$

A.2.6 Type-B OC Curves, Aggregation First by Columns, With Majority Column-Wise Aggregation

This case is analogous to the previous one (Section A.2.5), the only difference being that one substitutes $\mathbb{P}(d) = \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d}$, in order to obtain a Type-B OC curve based on the binomial model (cf. Equation 2). It is thus obtained:

$$P_a = \sum_{d=0}^{n} \left\{ \left[\sum_{k=0}^{c} \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)} \left\{ \left[\binom{d}{j} \cdot \left(1 - \sum_{s=\lfloor \frac{m}{2} \rfloor+1}^{m} \left[\sum_{A \subseteq \{1,\,\dots,\,m\},\,|A|=s} \left\{ \left[\prod_{i\in A} b_i \right] \cdot \left[\prod_{i\notin A} (1-b_i) \right] \right\} \right] \right)^{j} \right.\right.\right.\right.$$
$$\left. \cdot \left\{ \sum_{s=\lfloor \frac{m}{2} \rfloor+1}^{m} \left[\sum_{A \subseteq \{1,\,\dots,\,m\},\,|A|=s} \left\{ \left[\prod_{i\in A} b_i \right] \cdot \left[\prod_{i\notin A} (1-b_i) \right] \right\} \right] \right\}^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot \left\{ 1 - \sum_{s=\lfloor \frac{m}{2} \rfloor+1}^{m} \left[\sum_{A \subseteq \{1,\,\dots,\,m\},\,|A|=s} \left\{ \left[\prod_{i\in A} (1-a_i) \right] \cdot \left(\prod_{i\notin A} a_i \right) \right\} \right] \right\}^{(k-j)} \right.$$
$$\left.\left.\left.\left. \cdot \left(1 - \sum_{s=\lfloor \frac{m}{2} \rfloor+1}^{m} \left[\sum_{A \subseteq \{1,\,\dots,\,m\},\,|A|=s} \left\{ \left[\prod_{i\in A} (1-a_i) \right] \cdot \left(\prod_{i\notin A} a_i \right) \right\} \right] \right)^{(n-d)-(k-j)} \right] \right\} \right] \cdot \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d} \right\} \tag{A6}$$

A.2.7 Type-A OC Curves, Aggregation First by Columns, With Unanimity Column-Wise Aggregation

Starting with the expression for P_a in Equation (13) and using Equations (17) and (19)—which refer to a and b in the case of aggregation *first by columns*, under the column-wise *unanimity* criterion—we set $r = n - c$ to reflect the SSP's LDD for row-based aggregation. Next, we substitute $\mathbb{P}(\boldsymbol{d}) = \frac{\binom{\boldsymbol{D}}{\boldsymbol{d}}\binom{\boldsymbol{N}-\boldsymbol{D}}{\boldsymbol{n}-\boldsymbol{d}}}{\binom{\boldsymbol{N}}{\boldsymbol{n}}}$, in order to obtain a Type-A OC curve based on the hypergeometric model (cf. Equation 1). It is finally obtained:

$$P_a = \sum_{d=0}^{n} \left\{ \left[\sum_{k=0}^{c} \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)} \left\{ \left[\binom{d}{j} \cdot \left(1 - \prod_{i=1}^{m} b_i \right)^{j} \cdot \left\{ \prod_{i=1}^{m} b_i \right\}^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot \left\{ 1 - \prod_{i=1}^{m} (1-a_i) \right\}^{(k-j)} \cdot \left(\prod_{i=1}^{m} (1-a_i) \right)^{(n-d)-(k-j)} \right] \right\} \right] \cdot \frac{\binom{D}{d}\binom{N-D}{n-d}}{\binom{N}{n}} \right\} \tag{A7}$$

A.2.8 Type-B OC Curves, Aggregation First by Columns, WITH Unanimity Column-wise Aggregation

This case is analogous to the previous one (Section A.2.7), the only difference being that one substitutes $\mathbb{P}(d) = \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d}$, in order to obtain a Type-B OC curve based on the binomial model (cf. Equation 2). It is thus obtained:

$$P_a = \sum_{d=0}^{n} \left\{ \left[\sum_{k=0}^{c} \sum_{j=\max[0,\,k-(n-d)]}^{\min(k,\,d)} \left\{ \left[\binom{d}{j} \cdot \left(1 - \prod_{i=1}^{m} b_i \right)^{j} \cdot \left\{ \prod_{i=1}^{m} b_i \right\}^{d-j} \right] \cdot \left[\binom{n-d}{k-j} \cdot \left\{ 1 - \prod_{i=1}^{m} (1-a_i) \right\}^{(k-j)} \cdot \left(\prod_{i=1}^{m} (1-a_i) \right)^{(n-d)-(k-j)} \right] \right\} \right] \cdot \binom{n}{d} \cdot p^d \cdot (1-p)^{n-d} \right\} \tag{A8}$$

A.3 Type-B OC Curves

Figures A2 and A3 contain the Type-B OC curves corresponding to the specific examples developed in Sections. 5.1 and 5.2.

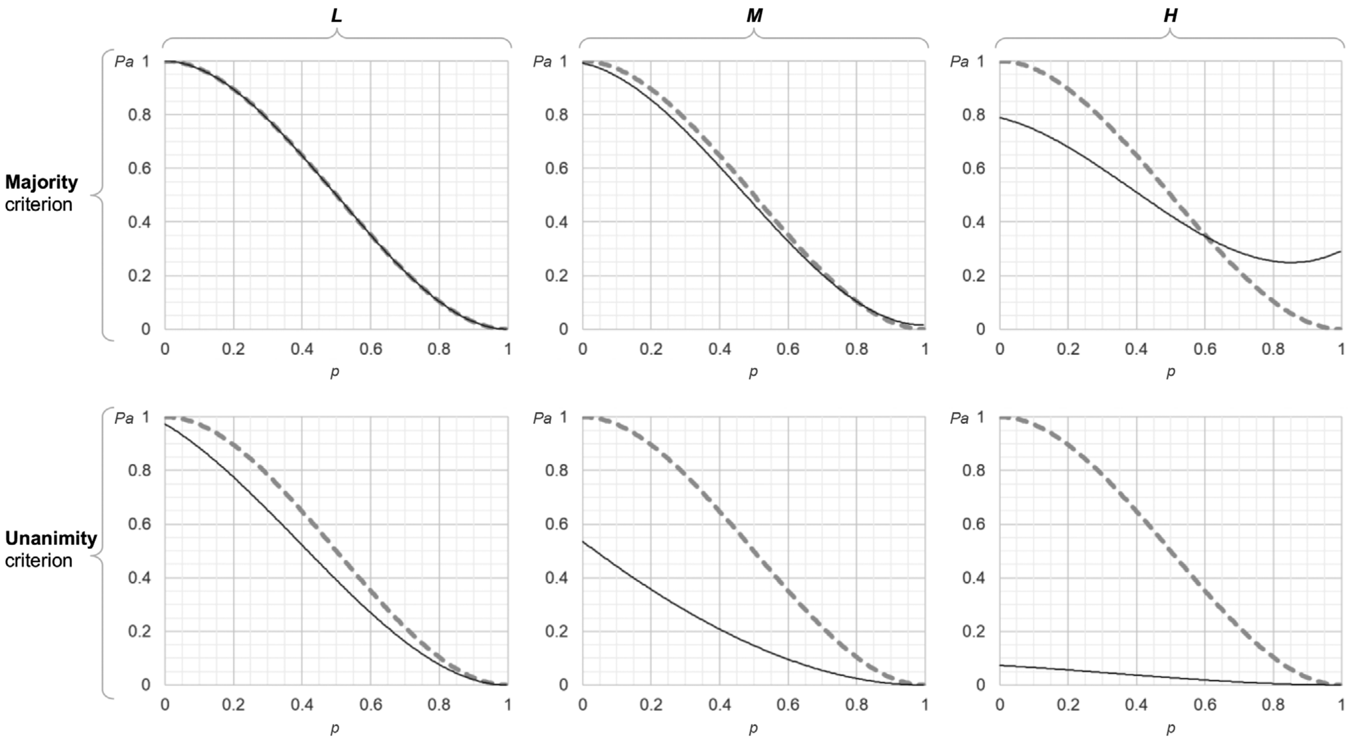

FIGURE A2 | Type-B OC curves (based on the *binomial* distribution) for a SSP ($N = 30$, $n = 3$ and $c = 1$) involving $m = 5$ inspectors, using the sequence of aggregation *first by rows*. Three scenarios are considered, characterized by inspectors with increasing inspection error rates (a_i and b_i): *L*, *M* and *H* (cf. Table 2). The three upper graphs refer to column-wise aggregation by *majority*, while the three lower graphs correspond to column-wise aggregation by *unanimity*. The dashed OC curve represents the 'ideal' SSP with no inspection errors.

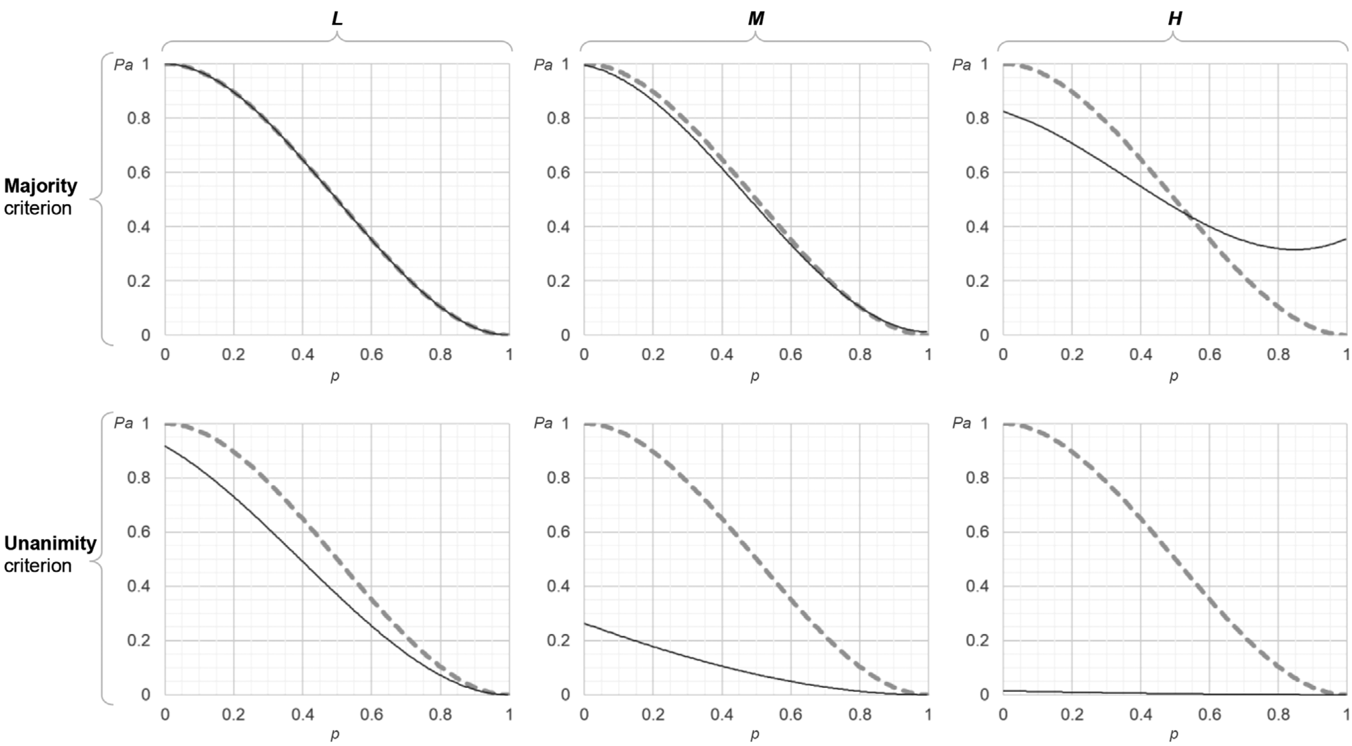

FIGURE A3 | Type-B OC curves (based on the *binomial* distribution) for a SSP ($N = 30$, $n = 3$ and $c = 1$) involving $m = 1$ inspectors, using the sequence of aggregation *first by columns*. Three scenarios are considered, characterized by inspectors with increasing inspection error rates (a_i and b_i): *L*, *M* and *H* (cf. Table 2). The three upper graphs refer to column-wise aggregation by *majority*, while the three lower graphs correspond to column-wise aggregation by *unanimity*. The dashed OC curve represents the 'ideal' SSP with no inspection errors.