Brief Report

Using CA-62 Marker in a Pilot Screening Project in Kazakhstan, 2021: Baseline Data and Scenario-Based Modeling of Estimated Specificity

Sergey Tsurkan , Evgueni Klinski , Anna Prostyakova * , Janneta Tcherkassova

Posted Date: 6 May 2026

doi: 10.20944/preprints202604.2003.v2

Keywords: CA-62; screening; asymptomatic cohort; epithelial cancers; estimated specificity; open data

 doi:10.20944/preprints202604.2003.v2

Brief Report

Using CA-62 Marker in a Pilot Screening Project in Kazakhstan, 2021: Baseline Data and Scenario-Based Modeling of Estimated Specificity

Sergey Tsurkan [1], Evgueni Klinski [2], Anna Prostyakova [3,*] and Janneta Tcherkassova [4]

[1] JVS Diagnostic LLC, Moscow, Russia
[2] UCM Technologies, Toronto, Ontario, Canada
[3] Laboratory Polymers for biology, Shemaykin-Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia
[4] JVS Diagnostic LLC, Moscow, Russia
* Correspondence: prostyakova@gmail.com

Abstract

The CLIA-CA-62 assay is an in vitro diagnostic device registered in Russia and Kazakhstan for measuring a marker specific to epithelial carcinomas. This pilot project aimed to assess CA-62 utility for primary cancer screening in an asymptomatic cohort in Kazakhstan. The trial was interrupted in January 2022 for reasons unrelated to the scientific program before clinical outcomes could be obtained. Available baseline data were therefore used to characterize the CA-62 value distribution and perform a scenario-based assessment of estimated assay specificity at a reference value of 5,000 U/mL. The analysis included 1,214 quantitative CA-62 measurements from asymptomatic healthcare workers aged 45–70 years, collected during annual preventive examinations between September and October 2021. The distribution was markedly right-skewed, with 92.5% of samples in the normal zone (median: 3,371 U/mL; IQR: 1,965–4,415 U/mL; 95th percentile: 6,309 U/mL). At the 5,000 U/mL cutoff, 7.5% of results (91/1,214) were elevated. Scenario-based modeling assuming cancer prevalence of 0.5–2.5% and assay sensitivity of 65–95% yielded an estimated specificity of 92.79–94.75%. These findings provide an analytical foundation for prospective verification of CA-62 in primary screening settings.

Keywords: CA-62; screening; asymptomatic cohort; epithelial cancers; estimated specificity; open data

1. Introduction

Early detection of malignant neoplasms remains one of the major unresolved challenges in clinical oncology. According to the Global Cancer Observatory (IARC/WHO), approximately 20 million new cancer cases and 9.7 million cancer deaths occurred worldwide in 2022 [1]. Five-year survival at stage I exceeds 80-90% for most solid epithelial tumors, whereas for most cancers at stage IV it remains below 20% [2]. At the same time, more than 70% of non-small cell lung cancers and over 90% of pancreatic cancers are diagnosed at advanced stages, when there are very few alternatives for curative treatment [2].

Established screening modalities, including mammography, PSA testing, colonoscopy, and low-dose chest CT scan, have well-recognized limitations, including organ specificity, invasiveness, high false-positive rates, and limited trough output [3,4]. Circulating tumor DNA-based approaches are promising; however, their sensitivity at stages I-II remains constrained by the low release of tumor material into the bloodstream at early stages [5]. One conceptual approach is the detection of markers linked to common characteristics of malignant transformation, such as loss of epithelial differentiation, rather than individual mutations or organ-specific proteins.

 doi:10.20944/preprints202604.2003.v2

CA-62 is a glycosylated cell-surface antigen associated with an N-glycoside epitope characteristic of poorly differentiated epithelial cells. The biological rationale for this marker is based on the concept of poor differentiation as a common feature of malignant transformation across tumors of epithelial origin, regardless of organ site [6]. Previous experimental studies related to the development of CA-62 marker suggested that common epithelial antigenic features may be shared across different tumor tissues [15]. At the time the Kazakhstan pilot project was initiated, the CA-62 test kit had been registered as an in vitro diagnostic medical device in both the Russian Federation (Registration Certificate No. 2020/9880) and the Republic of Kazakhstan (RK-IVD-5 No. 019561).

Previous Retrospective blinded validation studies have reported high diagnostic performance of CA-62 in early-stage breast cancer (sensitivity, 97-100%; specificity, 97%; AUC, 0.989) [7], non-small cell lung cancer in combination with CEA and CYFRA 21-1 (sensitivity, 93%; specificity, 100%; AUC, 0.990) [8], renal cell carcinoma (sensitivity, 89%; specificity, 97%; AUC, 0.94) [9], and prostate cancer [10]. A separate publication on treatment monitoring in breast cancer using CA-62 marker also supported the potential value of the marker for cancer monitoring [11].

Population-based data on the distribution of CA-62 in real asymptomatic cohorts, however, has remained largely unavailable. This has limited the ability to estimate expected specificity and the frequency of false-positive results under screening conditions, both of which are key parameters in evaluating the feasibility of broader assay implementation [3,12].

In 2021, a pilot screening trial using CA-62 was initiated in Nur-Sultan, Kazakhstan, among medical personnel who were required to undertake annual mandatory preventive examinations. The original study design specified referrals of individuals with results above 5,000 U/mL for confirmatory evaluation. In January 2022, the study was interrupted before the planned clinical outcomes could be obtained, for reasons unrelated to the scientific program. Under these circumstances, analysis of the accumulated primary dataset became the most scientifically appropriate objective.

The present study had two objectives: first, to define the baseline CA-62 value distribution in this asymptomatic group and make the de-identified primary dataset accessible to the public; the second objective is to perform a scenario-based assessment of the estimated specificity of the assay at the reference value of 5,000 U/mL on the basis of the observed proportion of positive results.

2. Materials and Methods

2.1. Study Design

A pilot clinical prospective cohort study on screening among the generally healthy population of healthcare workers of Nur-Sultan (RK) aged 45 to 70 years based on the quantitative measurement of the CA-62 level to identify groups of high risk with early stages of epithelial cancer of various primary localizations and the referral of identified patients to a specialized medical institution for confirmation of the diagnosis.

2.2. Participants

The pilot project included asymptomatic healthcare workers aged 45-70 years who underwent mandatory annual medical examinations at healthcare institutions in Nur-Sultan, Kazakhstan, between September and December 2021. This cohort was selected for the evaluation of the screening algorithm in a population under regular preventive observation, thereby reducing the likelihood of including individuals with clinical symptoms or previously undetected advanced cancers. Standard contraindications to vending the venipuncture were the only exclusion criteria. All local organizational, participant-facing, and personal-data-related aspects of the pilot cohort study in Kazakhstan, including participant selection, obtaining informed consent, sample collection, sample coding, and local data handling, were managed by MEC Lab LLP (Nur-Sultan). The samples and associated study data used for the present analysis were obtained by MEC Lab LLP during the local conduct of the pilot project and were provided to JVS Diagnostics LLC in coded form without

 doi:10.20944/preprints202604.2003.v2

personal identifiers. JVS Diagnostics LLC performed independent measurements on the material provided. JVS Diagnostics LLC covered the cost of these measurements.

2.3. Methods

Quantitative measurement of CA-62 was performed using a competitive chemiluminescent immunoassay (CLIA-CA-62, JVS Diagnostics LLC) in a centralized authorized laboratory. In this assay format, acridinium-labeled CA-62 conjugate competes with endogenous antigen for binding to immobilized antibodies; chemiluminescence intensity is inversely proportional to the antigen concentration in the sample. Results were expressed in arbitrary units (U/mL) within an inverse-logarithmic calibration framework. The assay coefficient of variation was below 10% across the working range of 1,250-10,000 U/mL. The reference value of 5,000 U/mL represents an operational analytical level established during development of the competitive assay at 95% specificity in healthy donor samples and should not be interpreted as an absolute physicochemical concentration of a molecular analyte.

2.4. Statistical Analysis

Descriptive statistics included the minimum, maximum, median, quartiles, and selected percentiles of the distribution. The proportion of positive results was defined as the proportion of measurements exceeding 5,000 U/mL. Since CA-62 values are expressed in arbitrary units using an inverse-logarithmic calibration system, and the classification near the operational cutoff can be affected by assay-related analytical variability, a confidence interval was not calculated for the observed proportion exceeding the cutoff.

2.5. Scenario-Based Modeling of Specificity

Based on the observed positive proportion ($r = 91/1{,}214 \approx 0.075$), estimated specificity was calculated using the following formula, derived algebraically from Bayesian relations:

$$Sp = 1 - (r - Se \times p) / (1 - p),$$

where r is the observed proportion of results above 5,000 U/mL, p is the assumed prevalence of cancer in the cohort (0.5-2.5%), and Se is the assumed assay sensitivity (65-95%). A 5 × 7 scenario matrix was calculated without imputation of missing data and without outlier removal. PPV and NPV were also calculated for illustrative purposes.

2.6. Public Data Release

A de-identified primary dataset containing a record identifier, JVS sample code, specimen type, and CA-62 value in U/mL is being deposited in a public repository with assignment of a permanent DOI. Internal participant codes and clinical fields are not included in the publicly released dataset.

3. Results

Of the 1,220 source records, 1,214 contained a quantitative CA-62 measurement and were included in the final analytical dataset. Records without quantitative value were excluded.

The distribution of CA-62 values at baseline was markedly right-skewed (Figure 1). The principal descriptive statistics are shown in Table 1.

Table 1. Descriptive statistics of the CA-62 distribution (n = 1,214).

Parameter	Value
Number of observations (n)	1,214
Minimum, U/mL	10.0
Maximum, U/mL	10,462.3
Median, U/mL	3,371.2
Q1 (25th percentile), U/mL	1,965.1

 doi:10.20944/preprints202604.2003.v2

Q3 (75th percentile), U/mL	4,415.0
95th percentile, U/mL	6,309.4
Results >5,000 U/mL, %	7.5% (91/1,214)
Number of observations (n)	1,214
Minimum, U/mL	10.0

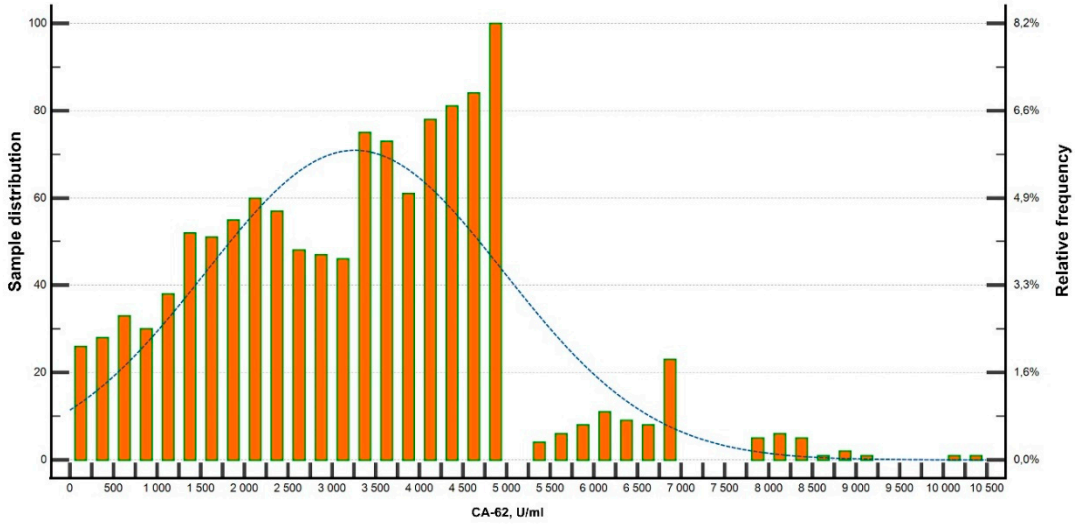

Figure 1. Sample distribution and relative frequency (%) based on CA-62 measurements. Blue dots – normal distribution pattern.

Most values were clustered on the left between 1,000 and 5,000 U/mL (92.5%), whereas above 5,000 U/mL the density of the distribution declined sharply, forming a limited right tail. Notably, the visible inflection in the distribution coincided with the historically established reference value of the assay. This observation should not be interpreted as retrospective cutoff value selection on the basis of the present data; rather, it reflects internal consistency between the population-level dataset and an analytical cutoff established previously using independent data. At the reference value of 5,000 U/mL, 91 of 1,214 results (7.5%) exceeded the cutoff.

Across the full modeled range of sensitivity (65-95%) and prevalence (0.5-2.5%), estimated specificity remained within 92.79-94.75%. The corresponding PPV ranged from 4.3% to 31.7%, whereas NPV remained high at 99.1-99.97% across all scenarios considered. (Table 2)

Table 2. Scenario-based modeling of the estimated specificity (Sp, %) of CA-62 at the reference value of 5,000 U/mL.

Assumed cancer prevalence, p, %	Se = 65%	Se = 70%	Se = 75%	Se = 80%	Se = 85%	Se = 90%	Se = 95%
0.5	92.79	92.82	92.84	92.87	92.89	92.92	92.94
1.0	93.08	93.14	93.19	93.24	93.29	93.34	93.39
1.5	93.38	93.46	93.53	93.61	93.68	93.76	93.84
2.0	93.68	93.78	93.88	93.98	94.09	94.19	94.29

* Note. Calculated as Sp = 1 - (r - p × Se) / (1 - p), where r = 7.5% is the observed proportion of CA-62 results above 5,000 U/mL in the baseline cohort; p = assumed cancer prevalence; Se = assumed assay sensitivity.

4. Discussion

The principal result of the present prospective cohort study is the first publicly available population-based distribution of CA-62 values obtained in a real asymptomatic occupational cohort. Despite early interruption of the project, the resulting primary dataset provided two levels of

 doi:10.20944/preprints202604.2003.v2

information: a descriptive characterization of the distribution and a scenario-based estimate of specificity linking the population-based data to previously reported clinical properties of the marker.

The observed proportion of positive results (7.5%) was somewhat higher than the expected 4-5% that might be inferred from a 95% specificity estimate obtained in healthy donor reference samples. This discrepancy is consistent with what may be expected when moving from controlled validation settings to real-world clinical practice: healthcare workers aged 45-70 years do not constitute a strictly healthy reference population and may have chronic inflammatory or other proliferative conditions that could influence CA-62 marker's levels [7,10]. In addition, minor batch-related variability within the dataset cannot be excluded. Across the assumed ranges of sensitivity and prevalence, scenario-based modeling suggested that estimated specificity would remain above 92%, broadly consistent with previously published validation series.

The consistently high NPV (>99% in all scenarios) is a clinically relevant implication of these calculations. In the context of screening asymptomatic populations, a high modeled negative predictive value suggests that a negative CA-62 result may be associated with a low probability of active malignancy under the stated assumptions. This feature could make CA-62 attractive as a potential first-line filter before resorting to more costly imaging techniques, a concept that has already been explored in the literature on cancer screening strategies [12,13].

The composition of the cohort is also methodologically relevant. Routine preventive check-ups reduce, but do not eliminate, the probability (the chance?) that undetected cancers have existed in the population. For this reason, the healthcare-workers cohort should be considered as an approximate screening group with moderate oncological risk rather than as a reference group of completely healthy individuals.

It should be emphasized that the present study does not establish observed specificity in this cohort and does not represent a complete validation of a screening strategy. The scenario-based modeling is illustrative and rests on assumptions regarding sensitivity and prevalence that may not fully hold under real-world circumstances. A valid prospective assessment will require full outcome verification.

Making the primary dataset publicly available guarantees that the calculations reported here can be independently reproduced and creates a foundation for future meta-analyses as comparable population data emerge. The FAIR principles—findability, accessibility, interoperability, and reusability—are increasingly adopted in biomarker research [14], and this publication adheres to that framework.

5. Conclusions

The baseline CA-62 measurements from 1,214 asymptomatic healthcare workers cohort aged 45–70 in the Kazakhstan pilot screening (Oct-Dec 2021) are scientifically valuable on their own. These CA-62 values displayed a clearly right-skewed pattern of sample distribution, with the majority of patients (92.5%) clustered on the left "normal values" and a sharp drop in frequency beyond previously defined operational reference cutoff value of 5,000 U/mL. By applying the baseline CA-62 value distribution to this asymptomatic cohort, we identified a subset of patients (7.5%) with elevated marker levels, likely representing high-risk individuals with underlying pathological conditions such as cancer, as it was found previously [7–10].

A scenario-based modeling across sensitivity values of 65-95% and assumed cancer prevalence of 0.5-2.5% yielded estimated specificity of 92.79-94.75% together with consistently high NPV (>99%). Public release of the primary dataset ensures independent reproducibility of the results and provides an analytical basis for subsequent prospective verification of CA-62 in primary screening applications in asymptomatic populations.

6. Limitations

 doi:10.20944/preprints202604.2003.v2

Early interruption of the study made it impossible to link CA-62 values with clinical outcomes in this healthcare-workers cohort. Demographic data and medical history were not available in the transmitted files. Possible batch-related variability limits generalizability. Scenario-based modeling is based on assumptions regarding clinical parameters and does not substitute for prospective validation.

Author Contributions: Conceptualization, S.T.. and J.T.; methodology, E.K.; formal analysis, A.P.; data curation, S.T.; writing—original draft preparation, S.T., E.K; writing—review and editing, J.T.; visualization, A.P.; supervision, J.T.; project administration, S.T. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Committee on the Evaluation of Medical Technologies of the Joint Commission on the Quality of Medical Services of the Ministry of Health of the Republic of Kazakhstan (protocol code №29, date of approval 29.07. 2020).”

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: Tsurkan S and Tcherkassova J are the employees of JVS Diagnostics LLC, the manufacturer of the CA-62 assay. All other authors declare no conflict of interest.

References

1. Bray F., Laversanne M., Sung H. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. doi: 10.3322/caac.21834
2. Sung H., Ferlay J., Siegel R.L. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-249. doi: 10.3322/caac.21660
3. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46. doi: 10.1158/2159-8290.CD-21-1059
4. Etzioni R., Urban N., Ramsey S. et al. The case for early detection. Nat Rev Cancer. 2003;3(4):243-252. doi: 10.1038/nrc1041
5. Lennon A.M., Buchanan A.H., Kinde I. et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science. 2020;369(6499):eabb9601. doi: 10.1126/science.abb9601
6. Tcherkassova J.R., Tsurkan S.A., Kondratiev V.B., Moro-Vidal R. Cancer antigen for early cancer detection. Patent WO2021215955A1. 2021.
7. Sekacheva M, Boroda A, Fatyanova A, Rozhkov A, Bagmet N. Clinical validation of the novel CLIA-CA-62 assay efficacy for early-stage breast cancer detection. Front Oncol. 2023 May 3;13:1009863.
8. Tcherkassova J, Klinski E, Tsurkan S, Prostyakova A, Boroda A, Pirogova Y, Bagmet L, Sekacheva M. Diagnostic significance of the novel biomarker combination for early-stage non-small cell lung cancer detection: results of the blind clinical study. Front Oncol. 2025 May 19;15:1508563.
9. Tcherkassova J, Tsurkan S, Prostyakova A, Klinski E, et al. CA-62: A new biomarker for the detection of early-stage renal cell carcinoma. J Clin Oncol. 2024;42(16_suppl):4526. doi:10.1200/JCO.2024.42.16_suppl.4526.
10. Tcherkassova J.R., Tsurkan S.A., Prostyakova A.I. et al. Potential clinical application of the CA-62 cancer antigen in the differential diagnosis of prostate cancer and benign prostatic hyperplasia. Research and Practical Medicine. 2023;10(1):10-26.
11. Khakimova G.G., Tcherkasova J.R., Tsurkan S.A., Fedchikov G.A., Suganov N.V., Gorbunova V.A. A pilot clinical trial to monitor response to chemotherapy using the CA-62 marker of epithelial carcinomas. Siberian J Oncol (2019) 18(5):18–28.
12. Woloshin S., Schwartz L.M. How a charity oversells mammography. BMJ. 2012;345:e5132.
13. Esserman L.J., Thompson I.M. Jr, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. JAMA. 2013;310(8):797-798

 doi:10.20944/preprints202604.2003.v2

14. Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
15. Tcherkassova J.R., Tsurkan S.A., Smirnova G.B., Borisova Yu.A., Treshchalina E.M. Expression of alpha-fetoprotein receptors in human tumor HOMOGENATES SW620, T47D and HepG2 cell culture. Russian Journal of Biotherapy. 2016;15(1):116-117. [in Russian]