scientific reports

Check for updates

OPEN

Calibration of two validated SARS-CoV-2 pseudovirus neutralization assays for COVID-19 vaccine evaluation

Yunda Huang[1,2,3✉], Oleg Borisov[4], Jia Jin Kee[1], Lindsay N. Carpp[1], Terri Wrin[5], Suqin Cai[5], Marcella Sarzotti-Kelsoe[6,7], Charlene McDanal[6], Amanda Eaton[6], Rolando Pajon[8], John Hural[1], Christine M. Posavad[1], Katherine Gill[9], Shelly Karuna[1], Lawrence Corey[1,10], M. Juliana McElrath[1,11], Peter B. Gilbert[1,2,12], Christos J. Petropoulos[5] & David C. Montefiori[6✉]

Vaccine-induced neutralizing antibodies (nAbs) are key biomarkers considered to be associated with vaccine efficacy. In United States government-sponsored phase 3 efficacy trials of COVID-19 vaccines, nAbs are measured by two different validated pseudovirus-based SARS-CoV-2 neutralization assays, with each trial using one of the two assays. Here we describe and compare the nAb titers obtained in the two assays. We observe that one assay consistently yielded higher nAb titers than the other when both assays were performed on the World Health Organization's anti-SARS-CoV-2 immunoglobulin International Standard, COVID-19 convalescent sera, and mRNA-1273 vaccinee sera. To overcome the challenge this difference in readout poses in comparing/combining data from the two assays, we evaluate three calibration approaches and show that readouts from the two assays can be calibrated to a common scale. These results may aid decision-making based on data from these assays for the evaluation and licensure of new or adapted COVID-19 vaccines.

Multiple vaccines have shown short-term efficacy against coronavirus disease-19 (COVID-19) in phase 3 randomized placebo-controlled trials[1–7]; one has been fully approved by the US Food and Drug Administration (FDA)[8], three have been issued an Emergency Use Authorization by the FDA[9], and six have been given an Emergency Use Listing from the World Health Organization (WHO)[10]. To expediently meet the worldwide demand for sufficient doses (including booster doses) of safe and efficacious vaccines, and for the licensure of new or modified vaccines, the FDA and the WHO have recommended that immunobridging trials may be used to assess the effectiveness of booster vaccines, vaccines in new populations (e.g., pediatric) or against emerging SARS-CoV-2 variants[11,12], thus potentially negating the need for additional large-scale placebo-controlled randomized trials (which are becoming increasingly difficult to conduct, especially in high-income countries[13]) or active-control randomized trials[14]. In such immunobridging trials, vaccine-induced immune responses of an established immune marker surrogate endpoint[15–17] are typically assessed to ensure that levels deemed sufficient for clinical efficacy are achieved. Neutralizing antibodies (nAbs) are correlates of protection for many licensed vaccines[17], and an accumulating body of evidence[18–27] supports their evaluation in immunobridging trials for COVID-19 vaccines.

In each of the five US government (USG)-sponsored phase 3 trials, nAb responses are assessed by one of two different pseudovirus-based SARS-CoV-2 neutralization assays, one performed at Duke University and one at Monogram Biosciences. Each pseudovirus neutralization assay has been rigorously validated in accordance with

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [3]Department of Global Health, University of Washington, Seattle, WA, USA. [4]Biomedical Advanced Research and Development Authority, Washington, DC, USA. [5]LabCorp-Monogram Biosciences, South San Francisco, CA, USA. [6]Department of Surgery, Duke University Medical Center, Durham, NC, USA. [7]Department of Immunology, Duke University Medical Center, Durham, NC, USA. [8]Moderna, Inc., Cambridge, MA, USA. [9]Desmond Tutu HIV Centre, University of Cape Town, Cape Town, South Africa. [10]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, USA. [11]Departments of Medicine and Laboratory Medicine, University of Washington, Seattle, WA, USA. [12]Department of Biostatistics, University of Washington, Seattle, WA, USA. ✉email: yunda@scharp.org; david.montefiori@duke.edu

the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)/FDA guideline[28,29] and these assays are planned for use in additional efficacy trials or immunobridging trials of COVID-19 vaccine candidates for potential regulatory approvals. If the Duke and Monogram assays do not render comparable readouts or are not calibrated (cross-validated) to a common scale, nAb response data from these assays may misguide regulatory decisions on these COVID-19 vaccine candidates.

To facilitate the harmonization of SARS-CoV-2 nAb data across multiple laboratories and assay types, the WHO developed an International Standard (WHO I.S.) for anti-SARS-CoV-2 immunoglobulin consisting of a pool of 11 COVID-19 convalescent plasma samples and provided as lyophilized material that, after reconstitution, was assigned a unitage of 1,000 international units per mL (IU/mL) for neutralization assays[30]. Here we describe the Duke and Monogram nAb assays and compare titer readouts between the two assays using a diverse set of samples that include the WHO I.S., COVID-19 convalescent serum samples, and serum samples from Moderna mRNA-1273 vaccine recipients. Because the ultimate goal is to ensure comparability of nAb readouts in vaccine recipients, we also present and compare statistical approaches for calibrating readouts from the two assays to support future vaccine evaluation and development.

Results

Table 1 summarizes and compares the Duke and Monogram assays. Similarities of both assays include utilizing lentiviral vector that (1) is pseudotyped with full-length G614 Spike protein (the SARS-CoV-2 surface protein responsible for virus entry via interaction with the human ACE2 receptor[31,32]) and (2) contains a firefly luciferase (Luc) reporter gene. Thus, in both assays a quantitative luminescence readout is directly proportional to the amount of pseudovirus entry into target cells. In both assays, neutralizing antibodies in the assayed sample inhibit pseudovirus entry and reduce luminescence signal in a dose-dependent manner, providing the basis for quantifying the neutralizing activity of the sample. Both assays are established to measure ID50 and ID80 neutralizing titers and are fully validated in terms of all required assay performance parameters[28,29] for convalescent serum and partially validated in terms of specificity, linearity, precision and limits of quantitation of the assay for vaccine incurred samples, in order to extend validation for the purpose of testing clinical trial specimens.

Differences in pseudovirus preparation, cell lines, plate layouts, and dilution schemes are also noted in Table 1. Among the notable differences is the way that Transmembrane Protein Serine 2 (TMPRSS2) is introduced into the assay. SARS-CoV-2 fuses with host cells by binding to the ACE2 receptor; the host cell TMPRSS2 protease cleaves the Spike protein at the S1/S2 sites to facilitate fusion[33]. In the Duke assay, TMPRSS2 is expressed during pseudovirus production, whereas in the Monogram assay TMPRSS2 is expressed on the surface of HEK 293 T/ACE2 target cells. This difference can impact relative neutralization titers measured in these assays, with higher titers seen in HEK 293 T/ACE2/TMPRSS2 cells (Montefiori, unpublished).

For assay comparison and establishment of the calibration algorithms, nAbs were measured in the first WHO I.S. for anti-SARS-CoV-2 immunoglobulin (20/136), as well as in 248 convalescent serum samples by both the Duke and Monogram labs following each assay format. For demonstration of the performance of the proposed calibration methods, nAbs were measured by both labs in 90 mRNA-1273 vaccine recipient serum samples [30 participants at three time points each: Day 1 (baseline), Day 29 (4 weeks post first vaccination), and Day 57 (4 weeks post second vaccination)].

Among the 90 samples from vaccine recipients, all Day 1 (pre-vaccination) samples had non-detectable nAbs by both labs, indicating a high level of specificity of the assays. Moderate nAb titers were detected after the first vaccination at Day 29 among almost all of the participants for ID50 (30 out of 30, 100% by Duke, 27 out of 30, 90% by Monogram) and the majority of participants for ID80 (80% by Duke, 70% by Monogram). nAb titers at Day 57 were detected among all participants who received the second dose (100% for ID50 and ID80 by both labs), and Day 57 titers were on average more than 20-fold higher than Day 29 nAb responses in both assays (Fig. 1, Table 2). Specifically, geometric mean ID50 titers (standard deviation of $\log_e$-transformed titers, SD) were 65.7 (0.8) at Day 29 and 1463.5 (1.1) at Day 57 as measured by the Duke lab and 178.3 (1.0) at Day 29 and 4120.7 (1.0) at Day 57 as measured by the Monogram lab. Geometric mean ID80 titers were 16.2 (0.8) at Day 29, and 330.5 (0.9) at Day 57 as measured by the Duke lab; 60.9 (0.8) at Day 29 and 1522.5 (1.0) at Day 57 as measured by the Monogram lab. Moderate correlations were seen between the Day 29 and Day 57 nAb titers (Figure S1). Day 29 and Day 57 nAb titers were treated separately in the calibration process, in order to cover different dynamic ranges of possible assay readouts.

Prior to calibration, we observed that ID50 and ID80 titer readouts from the Monogram lab were consistently higher than those from the Duke lab measured in convalescent sera, vaccine recipient sera, and the WHO I.S. sample (Fig. 1, Table 2). Specifically, the fold difference of individual-level ID50 titers from Monogram versus Duke across individuals had a geometric mean (SD) of 2.8 (0.7) in convalescent samples, 2.7 (0.4) and 2.8 (0.4) in Day 29 and Day 57 vaccine recipient samples, respectively, and the fold difference of the geometric (arithmetic) mean ID50 titers across replicates of the WHO I.S. sample was 4.9 (3.7); the fold difference of individual-level ID80 titers Monogram versus Duke across individuals had a geometric mean (SD) of 5.2 (0.7) in convalescent samples, 3.7 (0.5) and 4.6 (0.2) in Day 29 and Day 57 vaccine recipient samples, respectively, and the fold difference of the geometric (arithmetic) mean ID80 titers across replicates of the WHO I.S. sample was 7.5 (6.6). These differences in readouts between the two assays are an example of the need for calibration of the responses to the same scale prior to cross-assay comparisons or cross-assay merging of results. The fact that fold differences between the two assays are largely consistent across convalescent samples and vaccine recipient samples suggests that the differences between the assays are not specific to naturally acquired or vaccine-elicited nAbs. In addition, the range of readouts in convalescent samples generally covers that in vaccine recipient samples; this supports the use of those convalescent samples to develop the algorithms for the calibration of titers in the vaccine recipient samples. On the other hand, the WHO I.S., which is a single sample comprised of a pool of

Component	Duke	Monogram
Purpose	Measures ability of antibody in a serum/plasma/monoclonal or purified antibody sample to neutralize SARS-CoV-2 pseudovirus	
Principal	– SARS-CoV-2 pseudotyped virus expresses firefly luciferase upon infecting HEK 293 T cells, with relative light units (RLU) of luminescence directly proportional to the infectivity of the virus inoculum – Neutralizing antibodies inhibit pseudovirus infection of cells, resulting in lower RLUs – Serial dilutions of antibodies are used to construct a dose–response curve to quantify potency for each sample, recorded as 50% inhibitory dose (ID50) and 80% inhibitory dose (ID80)	
Format and capacity	Performed in 96-well plate format and includes controls	
	Up to 4 test samples in duplicate per plate	Up to 6 test samples in singlicate titrations per plate
Virus	Lentivirus:	
	Pseudotyped with SARS-CoV-2 G614 full-length Spike protein and packaged with TMPRSS2 serine protease and pHR' CMV Luc luciferase reporter genes	Pseudotyped with SARS-CoV-2 G614 full-length Spike protein and packaged with HIV genomic vector, pRTV1.FlucP.CNDO-ΔU3, containing a luciferase reporter gene in place of HIV envelope
Cell Line	HEK 293 T cell line:	
	– Stably overexpressing the human ACE2 cell surface receptor protein (defined as 293 T/ACE2), – Cells may be used up to passage 46 or 4 months in culture	– Transiently transfected to express human ACE2 cell surface receptor protein along with a serine airway protease, TMPRSS2, for processing the S protein – Cells may be used up to passage 30 or as long as data maintains high quality
Controls	– Assay Plate Positive Control – anti-spike RBD mAb spiked into normal human serum. Included on every test plate – Assay Run Positive Controls – convalescent serum samples with high, medium, and low titers – Assay Run Negative Control – normal human serum (NHS). – Included on at least one plate in each assay run	– SARS-CoV-2 Positive Control (CPC#) – pool of COVID-19 convalescent plasma with positive SARS-CoV-2 neutralization titer. Included on every test plate – SARS-CoV-2 Negative Control (CNC#) – pool of COVID-19 convalescent plasma with negative SARS-CoV-2 neutralization titer. Included on every test plate – Daily Positive Controls – convalescent serum samples with high, medium, and low titers tested once per day – Specificity Control Pseudovirus – a lentiviral pseudovirus with Influenza H10N3 envelope tested with every sample and control to detect non-specific inhibition
Equipment	– 96-well Poly-L-Lysine or regular plates – GloMax Navigator Microplate Luminometer (Promega) – Manual liquid handling	– 96 well opaque Luminometer plates – Luminoskan Luminometers – Semi-automated liquid handling
Assay Range	1:10 to 1:781,250 or 1:20 to 1:1,562,500	1:40 to 1:787,320
Sample types	Heat-inactivated (56 °C, 30 min) serum (preferred), plasma, convalescent and vaccine incurred samples plus monoclonal antibodies and purified antibodies	Heat-inactivated (56 °C, 60 min) serum, plasma, convalescent and vaccine incurred samples plus monoclonal and purified antibodies
Basic steps	– Each sample is diluted serially five-fold in duplicate for a total of 8 dilutions starting at a 1:10 or 1:20 dilution (dilution after adding virus but before adding cells) – Virus is added to the plate (except cell control wells) and incubated for 45–90 min in humidified 37 °C and 5% CO_2 incubator – 293 T/ACE2 cells are harvested and 100 µl added to the plate at a density of 10,000 cells/well – Plates are incubated for 71–73 h in humidified 37 °C and 5% CO_2 incubator – Add Bright-Glo luciferase reagent to lysed cells and luminescence measured with a luminometer	– Each sample is diluted initially 1:20 and then diluted serially three-fold to yield a total of 10 concentrations. The starting dilution of 1:20 becomes the final, reported as 1:40 after addition of the virus – Virus is added to the plate and incubated for 1–2 h in humidified 37 °C and 7% CO_2 incubator – Transfected 293/ACE2/TMPRSS2 cells are harvested and diluted to 10,000 cells/well and added to the plates – Plates are incubated for 3 days in humidified 37 °C and 7% CO_2 incubator – Steady Glo luciferase reagent is added to plate and luminescence measured with a luminometer
Calculations	Percent neutralization calculated as difference between RLUs of virus- and cell-only wells and test sample wells:	
	$\%Neutralization = 100\% \times (1 - \frac{Avg.RLU(Virus+Sample+Cells)}{Avg.RLU(Virus+Diluent+Cells)})$ – Average RLU of cell control wells (without virus) are subtracted from RLUs of all other wells on a plate	$\%Neutralization = 100\% \times (1 - \frac{RLU(Virus+Sample+Cells)}{Avg.RLU(Virus+Diluent+Cells)})$
	ID50 and ID80 neutralizing antibody titers are expressed as the reciprocal of the sample dilution required to reduce RLU by 50% and 80%, respectively	
Acceptance Criteria/ Plate level	– Positive Controls within acceptance range for ID50 or ID80 titers – Signal of virus-only wells is in range of RLU > 48,000 to < 1,000,000 – %CV ≤ 35% for virus only control wells	– Positive Control, CPC#, titer falls within expected range for the assay (± 3 SD of the mean) and within 30% of all CPC#s titers generated for the day – RLUs fall between ~ 10,000 and ~ 500,000
Acceptance Criteria/ Sample level	– Relative Percent Difference ≤ 30% between duplicate wells for each sample with ≥ 40% neutralization – Neutralization curves must cross 50% neutralization 0–1 time	– Percent inhibition values of serial dilutions must follow a logical progression, e.g. the 1:40 serum dilution should exhibit equal or more inhibition of viral infection than the 1:120 serum dilution. Once the data points decrease below maximum inhibition and are on the linear part of the curve, each succeeding data point (higher dilution) should be less inhibiting than the preceding point (lower dilution) – Graphs are rejected when antibody inhibition profiles (curves) have all or most (> 80%) of the points not falling on or within 20% of the extrapolated linear portion of the curve
Acceptance Criteria/ each Assay Run	ID50 or ID80 titers for Assay Run Positive and Negative Controls must be within acceptance range	Positive Control, CPC#, titer falls within expected range for the assay (± 3 SD of the mean) and within 30% of all CPC#s titers generated for the day
Validation Status	– Full validation for convalescent samples – Partial validation for vaccine incurred samples	

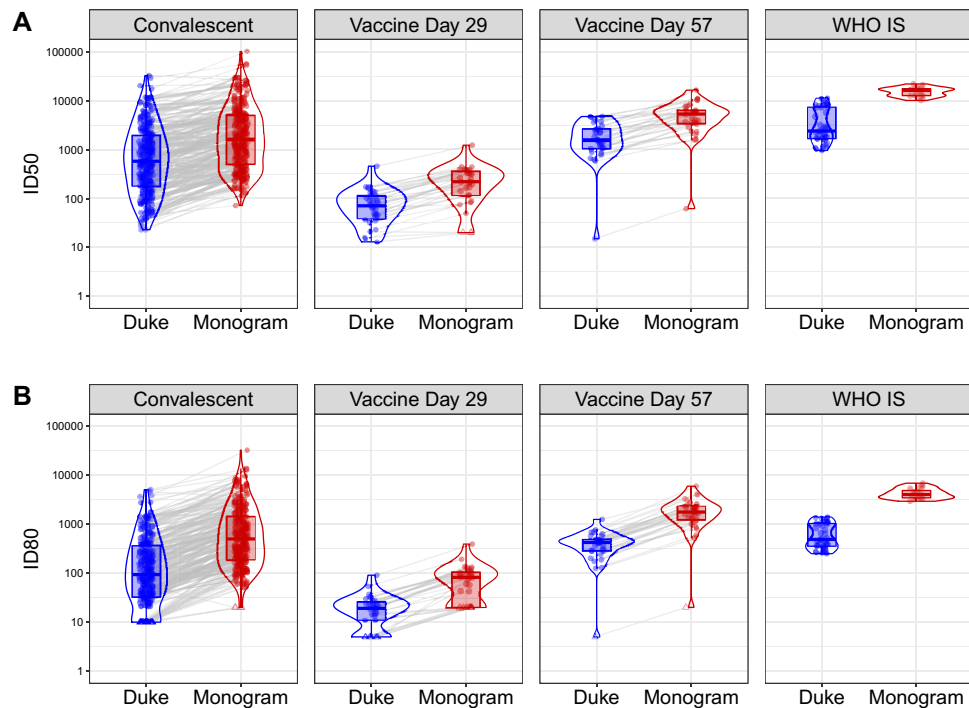**Table 1.** Overview comparison of the Duke versus Monogram assays.

Figure 1. Distributions of pre-calibration ID50 (Panel **A**) and ID80 (Panel **B**) titers measured by the Duke and Monogram labs of convalescent patient samples, vaccine recipient samples, and the WHO International Standard sample. For the convalescent samples, the assay lower limit of detection (LLoD) is set at 20 for Duke and 40 for Monogram, the reciprocal of the starting dilution of each assay (these titers are also considered as the positivity cutoff values). For the vaccine samples collected at Day 29 and Day 57, 4 weeks post the first and the second Moderna doses, respectively, and the WHO I.S. sample, the assay LLoD is set at 10 for Duke and 40 for Monogram, the reciprocal of the starting dilution of each assay (these titers are also considered as the positivity cutoff values). Circle-shape points indicate titers > LLoD; triangle-shape points indicate titers ≤ LLoD.

convalescent plasma samples, does not cover the entire range of titers and hence resulted in an expected narrower range of replicate values.

We evaluated three different approaches for the calibration of vaccine-induced neutralizing antibody response titers measured by the Duke and Monogram labs. In sum, Approach 1 uses tiers of the WHO I.S. sample to 'standardize' nAb titers measured by each lab to a common international unit (IU50/ml or IU80/ml for ID50 and ID80, respectively)[34], whereas Approaches 2 and 3, via two different statistical algorithms (details in Online Methods), use titers of an independent pool of clinical (e.g. convalescent) samples measured by both labs to calibrate titers measured by each lab to a common scale still in titer units. We next describe each Approach in greater detail.

For Approach 1, calibration factors were calculated using the arithmetic mean, geometric mean or median titers of the WHO I.S. sample measured by both labs (Table S1). Using the arithmetic mean-based calibration factor, ID50 (in IU50/ml) and ID80 (in IU80/ml) readouts of the Day 29 and Day 57 vaccine-induced responses from the two labs rendered the highest agreement, compared to the geometric- and median-based factors, with a concordance correlation coefficient (CCC) of 0.75 (95% CI: 0.60, 0.85) for ID50, and 0.72 (0.56, 0.82) for ID80 (Figure S2). Therefore, the arithmetic mean-based calibration factor is used in Approach 1 in all subsequent analyses. In Approach 2, readouts from the two labs were calibrated to the same scale (with the Monogram readouts set to be the reference) via an algorithm developed using titers of the convalescent samples. This approach assumes a bivariate normal (BIVN) distribution of the two assay readouts and incorporates the measurement error of each assay obtained in assay validation experiments. Similar to Approach 2, readouts from the Duke lab in Approach 3 were calibrated to the scale of the Monogram lab via an algorithm also developed using titers of the convalescent samples. Unlike Approach 2, Approach 3 is based on regressing (REG) the log-transformed titers of the Monogram assay (reference lab) on the Duke assay without incorporating assay measurement errors (Figure S3).

We next compared the pre-calibration vaccine recipient titer values from the two labs to the calibrated values obtained by each of the three Approaches. Because both labs measured the same set of samples, we consider that the closer the calibrated values from the two labs are to each other, the better the performance of the calibration is. For all three Approaches, the calibrated Day 57 vaccine recipient values from the two labs (Fig. 2, Table 3) are much closer to each other compared to the pre-calibration values (Fig. 1, Table 2). This is reflected in the geometric mean fold-differences (Monogram vs. Duke) being closer to 1, i.e. for ID50 the fold-difference (SD) was 0.8 (0.4) for titers calibrated by Approach 1, 1.0 (0.4) for Approach 2, and 1.2 (0.4) for Approach 3 (Table 3). The ID80 geometric mean fold-differences (SD) were similarly close to 1, specifically, 0.7 (0.2) for Approach 1,

	Convalescent (n = 248)	Vaccine Day 29 (n = 30)	Vaccine Day 57 (n = 30)	Vaccine Day 57 versus Day 29	WHO International Standard
ID50					
Duke					
Mean (SD)*	607.3 (1.6)	65.7 (0.8)	1463.5 (1.1)	22.3 (1.1)	3046.7 (0.8)
Median	581.0	72.5	1547.0	22.4	2421.5
Range	23.0–32,362.0	13.0–456.0	15.0–4714.0	0.5 – 158.4	957.0–11,101.0
Response Rate**	99%	100%	100%	NA	100%
Monogram					
Mean (SD)*	1692.0 (1.4)	178.3 (1.0)	4120.7 (1.0)	23.1 (1.1)	14,997.5 (0.2)
Median	1604.0	219.5	5274.0	23.5	16,133.0
Range	72.0–100,563.0	20.0–1222.0	62.0–16,280.0	0.4 – 163.2	10,044.0–21,785.0
Response Rate**	100%	90%	100%	NA	100%
Monogram versus Duke fold difference					
Mean (SD)*	2.8 (0.7)	2.7 (0.4)	2.8 (0.4)	NA	4.9 (NA)
Median	3.0	2.8	2.7		6.7
Range	0.3–26.0	1.0–7.1	1.2–6.3		2.0–10.5
ID80					
Duke					
Mean (SD)*	103.7 (1.6)	16.2 (0.8)	330.5 (0.9)	20.3 (0.9)	566.7 (0.6)
Median	94.5	19.5	421.5	22.7	489.0
Range	10.0–5078.0	5.0–91.0	5.0–1255.0	1.0–112.0	250.0–1391.0
Response Rate**	83%	77%	97%	NA	100%
Monogram					
Mean (SD)*	536.8 (1.4)	60.9 (0.8)	1522.5 (1.0)	25.0 (1.1)	4259.7 (0.2)
Median	504.5	82.0	1758.5	27.5	4062.5
Range	20.0–32,557.0	20.0–393.0	20.0–5984.0	0.5–130.5	2950.0–6892.0
Response Rate**	98%	70%	97%	NA	100%
Monogram versus Duke fold difference					
Mean (SD)*	5.2 (0.7)	3.7 (0.5)	4.6 (0.2)	NA	7.5 (NA)
Median	5.1	4.0	4.5		8.3
Range	0.4–75.2	1.3–15.0	2.9–7.8		5.0–11.8

Table 2. Summary of pre-calibration ID50 and ID80 titers measured by the Duke and Monogram labs and their fold-differences for convalescent patient sera, Day 29 vaccinee sera, Day 57 vaccinee sera, and the WHO International Standard sample. *Geometric mean of titers or titer ratios and standard deviation of the $\log_e$-transformed titers or titer ratios. **Proportion of samples with an ID50 or ID80 titer above the assay LLoD.

1.1 (0.2) for Approach 2, and 1.1 (0.3) for Approach 3. Across the three Approaches, calibrated Day 57 vaccine recipient values from the two labs are the closest to each other when Approach 2 is used.

For calibrated Day 29 vaccine recipient titer values, Approach 2 again yielded highly comparable results across the two labs, particularly for ID50, with a geometric mean fold-difference (Monogram vs. Duke) (SD) of 1.0 (0.4) for ID50 and 1.4 (0.5) for ID80 (Table 3). Compared to using Approach 2, Day 29 titers were less comparable across the two labs when either Approach 1 or Approach 3 was used for calibration; however, both Approach 1 and Approach 3 still brought the values of the two labs closer together compared to the non-calibrated values (reflected by the geometric mean fold-differences being closer to 1 in Table 3 compared to Table 2). Thus, any of the three calibration Approaches decreases the difference across the two assays in the Day 57 or Day 29 vaccine recipient titers.

We next assessed the level of agreement across labs (expressed as the CCC) of each of the three calibration Approaches, when including both Day 57 and Day 29 vaccine recipient titers in the analysis. Using Approach 2 for calibration (yielding BIVN-ID50 and BIVN-ID80 titers) achieved the highest agreement among the three Approaches, with a CCC (95% CI) of 0.87 (0.85, 0.87) for BIVN-ID50 and 0.85 (0.81, 0.88) for BIVN-ID80 (Fig. 3). While agreement was slightly less when either Approach 1 or Approach 3 was used for calibration, the CCC values still indicated reasonable agreement: By Approach 1, CCC (95% CI) for ID50 (IU50/ml) = 0.75 (0.60, 0.85) and for ID80 (IU80/ml) = 0.72 (0.56, 0.82); by Approach 3, CCC (95% CI) for REG-ID50 (Duke) with observed ID50 (Monogram) = 0.77 (0.71, 0.82) and for REG-ID80 (Duke) with observed ID80 (Monogram) = 0.68 (0.62, 0.75) for ID80 titers. Thus, Approach 3 rendered calibration performance largely comparable to that of Approach 1.
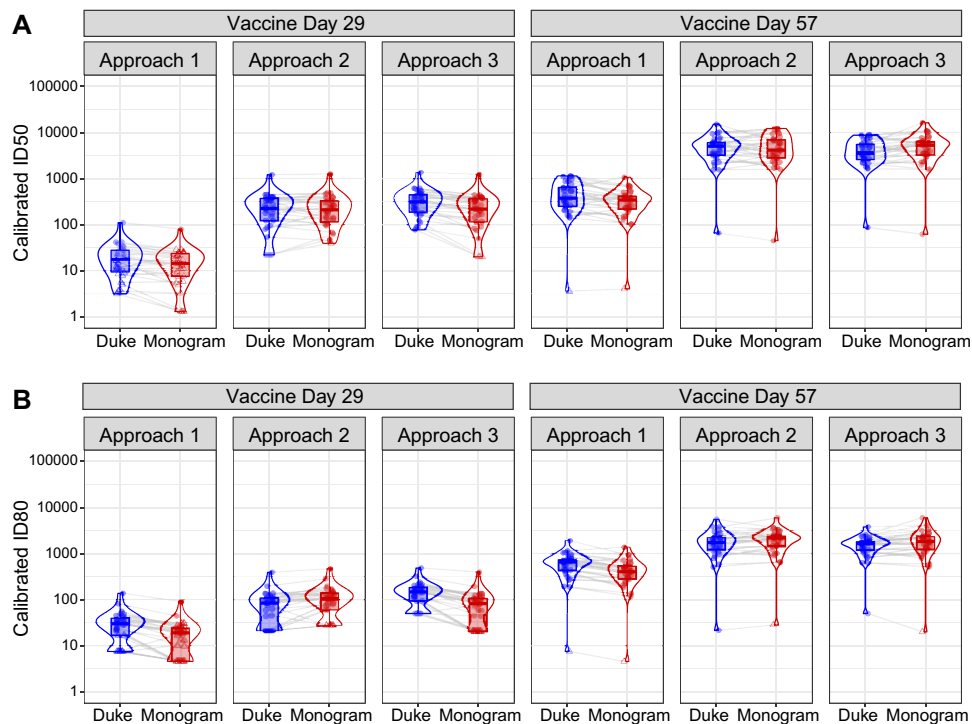
5

Figure 2. Distributions of calibrated ID50 (Panel **A**) and ID80 (Panel **B**) titers of vaccine recipient samples collected at Day 29 and Day 57, estimated by the three calibration approaches. In Approach 1, titers from each lab are calibrated to the WHO International Standard using the arithmetic mean-based calibration factor, expressed in International Units (IU50/ml for ID50; IU80/ml for ID80); in Approach 2, titers from each lab are calibrated to a common scale using an independent pool of clinical samples, assuming a bivariate normal distribution for the readouts from the two labs; in Approach 3, titers from the non-reference lab (Duke) calibrated to the reference lab (Monogram) using an independent pool of clinical samples, based on regressing titers from the reference lab on the non-reference lab. Circle-shape points indicate pre-calibration titers > LLoD; triangle-shape points indicate pre-calibration titers ≤ LLoD.

Discussion

Neutralizing antibody data obtained from multiple trials of different vaccines can support the establishment of nAbs as a correlate of protection, and if demonstrated, may facilitate approval of future vaccines through immunobridging. Evaluating samples from different vaccine trials is not often possible in a single assay and laboratory. In the case of SARS-CoV-2, nAb data were generated at one of two independent laboratories for each of the five USG-sponsored vaccine efficacy trials; this approach is planned for future trials, as well. Because assay readouts in vaccine recipient sera from these two laboratories are on average three to five-fold different, our study addresses this imperative need via calibrating data from the two assays to the same scale.

Our first main conclusion is that the WHO I.S. sample (Approach 1) demonstrated satisfactory performance comparable to that of the panels of convalescent sera in calibrating COVID-19 vaccine-induced nAb titers measured by these laboratories. This finding has important implications for future development and evaluation of COVID-19 vaccines because the WHO I.S. is more widely available and hence provides a straightforward way to compare and combine data across different platforms, and/or different developers and laboratories. Another main conclusion is that any of the three different calibration approaches can be used to bring readouts from different assays to the same scale for further analyses across studies, but that, due to the distinct advantages and disadvantages of each approach (summarized in Table S2), the specific approach that will be best suited for a given analysis depends on the availability of data to establish the calibration algorithms. For our motivating application—cross-protocol immunogenicity and correlates analyses of the 5 USG COVID-19 trials, where all nAb data are obtained by either the Duke or the Monogram assay—Approach 2 is recommended. However, for analyses including nAb data not measured by either the Duke or Monogram assay, then Approach 1 (using International Units) is likely the most feasible option, because not all labs have access to the convalescent samples that are necessary for Approach 2. We discuss other implementation considerations for the three Approaches below.

Among the three approaches, Approach 1 is the easiest to implement and likely requires the fewest resources as long as data on the WHO I.S. sample are available from the lab from a sufficient number of vials (e.g., > 20–30) and replicates. One requirement is to choose an appropriate summary statistic to calculate the calibration factor. In the dataset presented in this study, the calibration factor based on the arithmetic mean of ID50 or ID80 titers rendered the best performance. This is largely due to the fact that the corresponding WHO standard-based calibration factor for each lab translates to conversion factors, i.e., the ratio of readouts in international units of 3.7

Calibrated ID50	Vaccine Day 29 (n = 27)			Vaccine Day 57 (n = 30)		
	Approach 1	Approach 2	Approach 3	Approach 1	Approach 2	Approach 3
Duke						
Mean (SD)*	15.9 (0.8)	185.8 (1.0)	284.2 (0.7)	353.9 (1.1)	3955.5 (1.0)	3426.3 (0.9)
Median	17.5	227.4	307.5	374.1	5030.2	3582.3
Range	3.1–110.3	22.1–1210.8	77.5–1344.5	3.6–1140.0	66.4–15,078.3	86.9–8757.2
Monogram						
Mean (SD)*	11.6 (1.0)	190.9 (0.8)	178.3 (1.0)	268.9 (1.0)	3911.0 (1.0)	4120.7 (1.0)
Median	14.3	210.0	219.5	344.2	4128.0	5274.0
Range	1.3–79.8	39.4–1257.3	20.0–1222.0	4.0–1062.5	45.3–12,208.5	62.0–16,280.0
Monogram versus Duke fold difference						
Mean (SD)*	0.7 (0.4)	1.0 (0.4)	0.6 (0.5)	0.8 (0.4)	1.0 (0.4)	1.2 (0.4)
Median	0.7	1.0	0.7	0.7	1.0	1.2
Range	0.3–1.9	0.4–2.9	0.2–1.5	0.3–1.7	0.4–2.2	0.7–3.0
Calibrated ID80	Vaccine Day 29 (n = 27)			Vaccine Day 57 (n = 30)		
	Approach 1	Approach 2	Approach 3	Approach 1	Approach 2	Approach 3
Duke						
Mean (SD)*	24.4 (0.8)	63.5 (0.8)	125.8 (0.6)	496.1 (0.9)	1460.2 (1.0)	1330.0 (0.7)
Median	29.3	84.8	145.1	632.7	1680.3	1609.0
Range	7.5–136.6	21.5–390.4	50.0–484.7	7.5–1883.8	21.5–5538.9	50.0–3779.6
Monogram						
Mean (SD)*	13.9 (0.8)	87.9 (0.8)	60.9 (0.8)	347.3 (1.0)	1636.7 (0.9)	1522.5 (1.0)
Median	18.7	105.0	82.0	401.1	2072.5	1758.5
Range	4.6–89.6	28.0–468.1	20.0–393.0	4.6–1365.0	28.0–5975.6	20.0–5984.0
Monogram versus Duke fold difference						
Mean (SD)*	0.6 (0.5)	1.4 (0.5)	0.5 (0.5)	0.7 (0.2)	1.1 (0.2)	1.1 (0.3)
Median	0.6	1.3	0.5	0.7	1.1	1.2
Range	0.2–2.3	0.4–3.8	0.2–1.5	0.4–1.2	0.7–1.8	0.4–2.3

Table 3. Summary of calibrated ID50 and ID80 titers from the Duke and Monogram labs and their fold-differences for Day 29 vaccinee sera and Day 57 vaccinee sera. In Approach 1, titers from each lab are calibrated to the WHO International Standard using the arithmetic mean-based calibration factor, expressed in International Units per ml (IU50/ml for ID50; IU80/ml for ID80); in Approach 2, titers from each lab are calibrated to a common scale using an independent pool of clinical samples, assuming a bivariate normal distribution for the readouts from the two labs; in Approach 3, titers from the non-reference lab (Duke) calibrated to the reference lab (Monogram) using an independent pool of clinical samples, based on regressing titers from the reference lab on the non-reference lab. *Geometric mean of titers or titer ratios and standard deviation of the $\log_e$-transformed titers or titer ratios.

for ID50 and 6.6 for ID80 titers between Monogram versus Duke labs, that are in line with the fold-differences observed in the convalescent patient samples and in the vaccine recipient samples. This explains why this simple multiplicative-factor approach rendered reasonable calibration performance. Meanwhile, our findings also highlight that external neutralizing antibody titers from convalescent or vaccine sera generated by both labs on a common set of samples can inform decisions on which summary statistic to use in Approach 1 by evaluating the conversion factor against the additional data sources. Currently, Approach 1 is being used in the analyses of nAb data generated from two of the USG trials, one set of data from the Duke lab and the other by the Monogram lab. These analyses provide an important validation that assay-specific differences are minimized in the resulting calibrated data based on Approach 1 across the two USG trials, so that potential differences or similarity between the vaccine candidates in terms of their immune responses are not masked by differences between the two assays.

For both Approaches 2 and 3, a common pool of convalescent or vaccine samples will need to be assayed by both labs in order to establish the calibration algorithm, where Approach 2 requires readouts from both labs to be calibrated to a common scale, and Approach 3 calibrates the readouts from one lab to the other reference lab. We find that Approach 2 rendered the best calibration performance, followed by comparable performance between Approaches 1 and 3 for both ID50 and ID80 titers. These results suggest that precision in calibration can be improved by accounting for assay measurement error and the differences between the two assays across a range of response magnitudes (i.e., as done in Approach 2). In our case, measurement errors were characterized in assay validation experiments using both convalescent samples and vaccine samples by both labs, and assay differences in magnitude were characterized using convalescent samples with dynamic ranges of titers that were comparable to the ranges of titers of vaccine trial samples requiring calibration. Therefore, Approach 2 is recommended when paired-sample data are available from the two assays/labs (as is the case for the analyses of assay
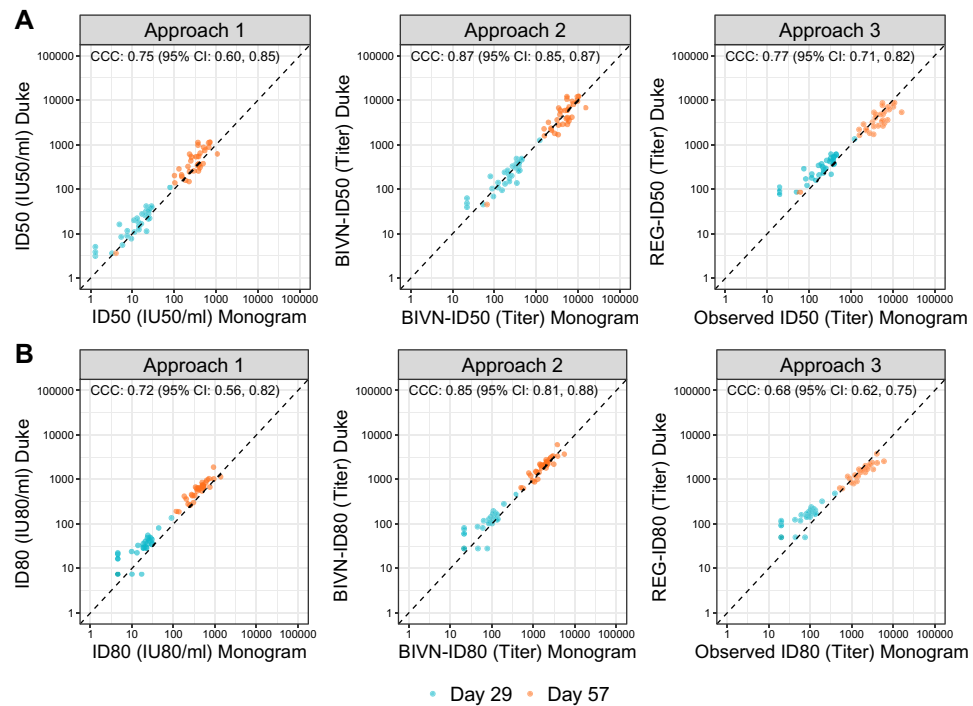
Figure 3. Scatterplots of calibrated ID50 (Panel **A**) and ID80 (Panel **B**) titers from the Duke and Monogram labs demonstrating performance of the three calibration approaches using vaccine recipient samples collected at Day 29 (turquoise) and Day 57 (orange), 4 weeks post the first and the second mRNA-1273 vaccine doses, respectively. The arithmetic mean-based calibration factor was used in Approach 1. The concordance correlation coefficient (CCC) and 95% confidence intervals indicate the level of agreement between the x- and y-axis values. ID50, ID80: ID50, ID80 titers calibrated to the WHO International Standard, expressed in International Units per ml (IU50/ml for ID50; IU80/ml for ID80) (Approach 1). BIVN-ID50, BIVN-ID80: ID50, ID80 titers from each lab calibrated to a common scale using an independent pool of clinical samples, assuming a bivariate normal distribution for the readouts from the two labs (Approach 2). REG-ID50, REG-ID80: ID50, ID80 titers from the non-reference lab (Duke) calibrated to the reference lab (Monogram) using an independent pool of clinical samples, based on regressing titers from the reference lab on the non-reference lab (Approach 3).

data from the 5 USG COVID-19 vaccine trials, where both the Duke and Monogram labs assayed a common set of convalescent samples for the calibration of their readouts in vaccine samples).

There are several limitations of this study. First, for evaluating the performance of the calibration approaches, we only considered a limited number of samples from trials of the mRNA-1273 vaccine. This potentially limits our ability to generalize findings from the calibration study to other COVID-19 vaccine platforms. We plan to assay a subset of samples from phase 3 trials of other COVID-19 vaccines to further validate the calibration approaches when those data become available. Second, to develop the calibration algorithms, both labs must have their assays performed on a common pool of independent samples that are expected to cover a similar dynamic range as the readouts requiring calibration. Access to such a common pool may be challenging for many labs. In that case, Approach 1 might be the most feasible option as long as the WHO I.S. sample is available in sufficient quantity. Third, for all three approaches, we focused the calibration only for titers above the lower limit of detection (LLoD) of the reference assays. Methods that account for calibration of responses below the LLoD, especially in the context of two assays having different LLoDs, are currently under development. Fourth, this study only concerns data from two specific labs measuring nAbs. Further validation is needed to investigate the performance of the three approaches applied to the calibration of other assays and/or immune responses. Although the described calibration approaches are only applied to the calibration of assays measuring anti-SARS-CoV-2 nAb responses induced by candidate vaccines, a more general use is amenable when assay data from multiple sources are of interest for comparison or combination. Fifth, the two nAb assays used in this study were both pseudovirus-based using lentiviral packaging. Future studies could also compare nAb responses obtained by the pseudovirus assay to those obtained by other SARS-CoV-2 neutralization platforms, e.g., pseudovirus-based neutralization using other viral vectors, or neutralization of live SARS-CoV-2. Lastly, this study focused on a single variant of SARS-CoV-2 (G614) that dominated the pandemic during the earliest phase 3 trials (e.g., Pfizer and Moderna trials), but has since been replaced by other variants, most recently the Delta variant[35]. Although the G614 form of Spike remains an important comparator, additional efforts are needed to establish new standards for current and future SARS-CoV-2 variants of concern.

Our study has a number of strengths. First, three distinct statistical approaches for calibration, each with different data requirements and/or data assumptions, are discussed to provide researchers versatile tools to help

guide decisions for which approach may be optimal for calibrating a given data set. Second, each assay was formally validated for the assessment of COVID-19 vaccine-induced or naturally acquired neutralizing antibody immune responses, according to current regulatory guidelines and standards. Third, to our knowledge, this is the first study that uses convalescent sera and the WHO I.S. sample to establish calibration algorithms and evaluate the performance of different calibration approaches for samples from COVID-19 vaccine trials. When more data become available, each approach could be readily updated to improve calibration accuracy and precision. Lastly, although the described calibration approaches were only illustrated for the analysis of data from two labs, all three approaches can be easily generalized to the calibration of data from more than two assays or labs.

In summary, correlates and immuno-bridging studies that could accelerate the development and evaluation of new and modified vaccines are at the heart of COVID-19 prevention efforts. We presented calibration approaches to support valid comparisons and combination of key immune response data from these studies to minimize the influence of assay-specific differences and optimize the characterization of future vaccine candidates.

Online methods

Ethics declaration. The work described here was approved by the Duke University Health System Institutional Review Board (Duke University) through protocol IDs Pro00093087 and Pro00105358. All experiments were performed in accordance with the relevant guidelines and regulations. All participants whose serum samples were assayed in this work provided informed consent.

Pseudovirus neutralization assays. Both assays have been formally validated in terms of their accuracy, precision (repeatability and intermediate precision), specificity, sensitivity, linearity, and limits of quantitation. Table 1 provides a comparison of the assays' components and major steps. Further details are provided for each assay below.

Duke. Neutralization was measured in a formally validated and FDA approved assay that utilized lentiviral particles pseudotyped with full-length SARS-CoV-2 Spike protein and containing the TMPRSS2 gene for production of the host serine protease and a firefly luciferase (Luc) reporter gene for quantitative measurements of infection by relative luminescence units (RLU).

The assay was performed in 293 T/ACE2.MF cells provided by Drs. Michael Farzan and Huihui Mu. Pseudoviruses were prepared, titrated and used for measurements of neutralizing antibodies essentially as described previously[36]. Briefly, an expression plasmid encoding codon-optimized full-length spike of the Wuhan-1 strain (VRC7480) was provided by Drs. Barney Graham and Kizzmekia Corbett at the Vaccine Research Center, National Institutes of Health (U.S.A.). Mutations, including the D614G mutation, were introduced into VRC7480 by site-directed mutagenesis[36] using the QuikChange Lightning Site-Directed Mutagenesis Kit from Agilent Technologies (Catalog # 210,518). All mutations were confirmed by full-length spike gene sequencing by Sanger Sequencing, using Sequencher and SnapGene for sequence analyses. Pseudovirions were produced in HEK293T/17 cells (ATCC cat. No. CRL-11268) by transfection using Fugene 6 (Promega Cat#E2692) and a combination of spike plasmid, TMPRSS2 plasmid, lentiviral backbone plasmid (pCMV ΔR8.2) and firefly Luc reporter gene plasmid (pHR' CMV Luc)[37] in a 1:17:17 ratio in Opti-MEM (Life Technologies). Transfection mixtures were added to pre-seeded HEK 293 T/17 cells in T-75 flasks containing 12 ml of growth medium and incubated for 16–20 h at 37 °C. Medium was removed and 15 ml of fresh growth medium added. Pseudovirus-containing culture medium was collected after an additional 2 days of incubation and clarified of cells by low-speed centrifugation and 0.45 μm micron filtration. TCID50 assays were performed as described previously[36]. For measurements of neutralization, a pre-titrated dose of virus was incubated with 8 serial fivefold dilutions of serum samples (1:10 or 1:20 starting dilution) in duplicate in a total volume of 150 μl for 1 h at 37 °C in 96-well flat-bottom poly-L-lysine-coated culture plates. 293 T/ACE2-MF cells were detached from T75 culture flasks using TrypLE Select Enzyme solution, suspended in growth medium (100,000 cells/ml) and immediately added to all wells (10,000 cells in 100 μL of growth medium per well). One set of 8 wells received cells + virus (virus control) and another set of 8 wells received cells only (background control). After 66–72 h of incubation, medium was removed by gentle aspiration and 30 μl of Promega 1X lysis buffer was added to all wells. After a 10-min incubation at room temperature, 100 μl of Bright-Glo luciferase reagent was added to all wells. After 1–2 min, 110 μl of the cell lysate was transferred to a black/white plate. Luminescence was measured using a GloMax Navigator luminometer (Promega). Serum samples were heat-inactivated for 30 min at 56 °C prior to assay. Neutralization titers are the inhibitory dilution (ID) of serum samples at which RLUs were reduced by either 50% (ID50) or 80% (ID80) compared to virus control wells after subtraction of background RLUs.

Monogram. Neutralizing antibody activity was measured in a formally validated and FDA approved assay that utilized lentiviral particles pseudotyped with full-length SARS-CoV-2 Spike protein and containing a firefly luciferase (Luc) reporter gene for quantitative measurements of infection by relative luminescence units (RLU). The backbone vector used in pseudovirus creation, F-lucP.CNDOΔU3, encodes the HIV genome with firefly luciferase replacing the HIV *env* gene. A codon-optimized version of the full-length spike gene of the Wuhan-1 SARS-CoV-2 strain (MN908947.3) (GenScript) was cloned into the Monogram proprietary *env* expression vector, pCXAS-PXMX, for use in the assay. The D614G spike mutation was introduced into the original Wuhan sequence by site-directed mutagenesis. Sequences of the spike gene and expression vector were confirmed by full-length sequencing using Illumina MiSeq NGS.

Pseudovirus stock was produced in HEK 293 cells via a calcium phosphate transfection using a combination of spike plasmid (pCXAS-SARS-CoV-2-D614G) and lentiviral backbone plasmid (F-lucP.CNDOΔU3). Transfected 10 cm² plates were re-fed the next day and harvested on Day 2 post transfection. The pseudovirus stock

(supernatant) was collected, filtered and frozen at < − 70 °C in single-use aliquots. Pseudovirus infectivity was screened at multiple dilutions using HEK293 cells transiently transfected with ACE2 and TMPRSS2 expression vectors. RLUs were adjusted to ~ 50,000 for use in the neutralization assay. Neutralization was performed in white 96-well plates by incubating pseudovirus with 10 serial threefold dilutions of serum samples for one hour at 37 °C. Serum samples were heat-inactivated for 60 min at 56 °C prior to assay.

The dilution series was based on a 1:20 starting dilution which was reported as 1:40 after addition of virus. HEK293 target cells, which had been transfected the previous day with ACE2 and TMPRSS2 expression plasmids, were detached from 10 cm^2 plates using trypsin/EDTA and re-suspended in culture medium to a final concentration that accommodated the addition of 10,000 cells per well. Cell suspension was added to the serum-virus mixtures and assay plates were incubated at 37 °C in 7% CO_2 for 3 days. On the day of assay read, Steady Glo (Promega) was added to each well. Reactions were incubated briefly and luciferase signal (RLU) was measured using a luminometer. Neutralization titers represent the inhibitory dilution (ID) of serum samples at which RLUs were reduced by either 50% (ID50) or 80% (ID80) compared to virus control wells (no serum wells).

The Monogram assay employs a specificity control which is created using the same HIV backbone/Luc sequence used in the SARS-CoV-2 pseudovirus. The envelope is 1949 Influenza A H10N3. It is unlikely for human sera to have antibodies against this rare avian influenza virus. The specificity control is designed to detect non-antibody factors (e.g., ART therapy) that could inhibit SARS-CoV-2 pseudovirus and result in false positive measurements of antibody neutralization. Positive anti-SARS-CoV-2 nAb activity was defined as an anti-SARS-CoV-2 nAb titer > 3 times greater than the titer of the same serum tested with the specificity control.

Serum samples. A serum sample is considered as having a detectable or positive neutralization antibody response when its neutralization ID50 or ID80 titer is greater than the lower limit of detection (LloD), which is set at the reciprocal of the starting dilution of the respective assay.

mRNA-1273 vaccine recipients:. Neutralization activity was assayed in a total of 90 serum samples collected in a phase 1 trial[38,39] (NCT04283461) that used the same mRNA-1273 vaccine, dose, and schedule as that used in the Moderna phase 3 trial[3] (NCT04470427). Sera were collected from 30 vaccine recipients at day 0, day 29 (4 weeks post-first dose), and day 57 (4 weeks post-second dose). Samples were stored at − 80 °C, thawed, and heat-inactivated for 30 min (Duke) or 60 min (Monogram) at 56 °C. Heat-inactivated samples were stored at 4 °C (Duke) or at − 80 °C (Monogram) until assayed.

Convalescent patients:. Neutralization activity was also assayed in a total of 248 serum samples collected in HVTN 405/HPTN 1901 (NCT04403880), an observational cohort study that enrolled individuals 18 years or older who had a positive SARS-CoV-2 test and recovered from a spectrum of infection and COVID-19 severity, from asymptomatic to requiring advance medical care in the intensive care unit, within 1–8 weeks of enrollment (if symptomatic). The laboratory was blinded to the clinical status of the donors but not to HIV-1 infection status. Only convalescent serum samples from HIV-1-uninfected individuals were used in the present analyses because antiretroviral drugs, which individuals with HIV may be taking, can inhibit replication of the HIV backbone of the pseudovirus used in the assay. Samples were stored at − 80 °C, thawed, and heat-inactivated for 30 min (Duke) or 60 min (Monogram) at 56 °C. Heat-inactivated samples were stored at 4 °C (Duke) or at − 80 °C (Monogram) until assayed.

WHO anti SARS-CoV-2 immunoglobulin International Standard (I.S.) sample. In December 2020, the WHO released a well-characterized international standard for the purpose of improving comparability of results among different assays in different laboratories and reducing interlaboratory variability of anti-SARS-CoV-2 antibody assays[34]. As described in the user instructions provided by the National Institute for Biological Standards and Controls (NIBSC): "The First WHO International Standard for anti-SARS-CoV-2 immunoglobulin is the freeze-dried equivalent of 0.25 mL of pooled plasma obtained from eleven individuals recovered from SARS-CoV-2 infection. The preparation was evaluated in a WHO International Collaborative study. The intended use of the International Standard is for the calibration and harmonization of serological assays detecting anti-SARS-CoV-2 neutralizing antibodies. The preparation can also be used as an internal reference reagent for the harmonization of binding antibody assays. The preparation has been solvent-detergent treated to minimize the risk of the presence of enveloped viruses"[40].

Calibration approaches and evaluation of concordance. We discuss and compare three different approaches for the calibration of vaccine-induced neutralizing antibody responses measured by the Duke lab and the Monogram lab. The Monogram lab is used as the reference lab in Approaches 2 and 3 because vaccine-induced neutralizing antibody responses from 4 out of 5 of the USG-sponsored COVID-19 phase 3 vaccine trials are being assayed by, or are planned to be assayed by, the Monogram lab for correlates of protection assessments.

In Approach 1, the Duke and Monogram responses are calibrated to a common scale based on the WHO I.S. sample. For measurements from both labs the calibrated responses ID50 and ID80 in international units (IUs) per microliter (ml) (IU50/ml and IU80/ml, respectively) are calculated as the observed ID50 or ID80 titers multiplied by the calibration factor, which is calculated as 1,000 (IU/ml) divided by the geometric mean, median or arithmetic mean ID50 or ID80 titers of the WHO I.S. sample that each lab measured in multiple vials. During the validation step, the calibration factor based on one of the above summary statistics that renders the highest agreement between calibrated responses from the two labs for the same set of samples is selected.

In Approach 2, both the Duke and Monogram responses are calibrated to a common scale using the Monogram lab as the reference based on methods described in Huang et al.[41] Briefly, both readouts are assumed to

share the same underlying true values but carry lab-specific measurement errors, following a bivariate normal distribution. Based on the assay validation experiments from both labs, a 30% coefficient of variation is assumed for the measurement error for ID50 and ID80 titers in both assays. The calibrated values are then calculated as the expected conditional means derived from the joint bivariate distribution as $\mu_1 + \frac{\sigma_{x^{(1)}}^2}{\sigma_{x^{(1)}}^2 + \tau_1^2}\left(\omega^{(1)} - (\mu_1 + \delta_1)\right)$ for the Monogram assay, where μ_1 and $\sigma_{x^{(1)}}^2$ are estimated by the sample mean and variance of the log-transformed ID50 and ID80 titers of the convalescent samples measured by the Monogram assay, τ_1^2 is the estimated measure error of the assay ($= 0.3\wedge 2 = 0.09$), δ_1 is zero because the Monogram assay is considered as the reference, and $\omega^{(1)}$ is the log-transformed pre-calibrated titer of the vaccine sample measured by the Monogram assay. Similarly, the calibrated values for the Duke assay are calculated as $\mu_2 + \frac{\sigma_{x^{(2)}}^2}{\sigma_{x^{(2)}}^2 + \tau_2^2}\left(\omega^{(2)} - (\mu_2 + \delta_2)\right)$, where all parameters are estimated using data from the Duke assay as is done for the Monogram assay, except that δ_2 is estimated by the average of the log-transformed difference between the Duke and Monogram assay readouts of the convalescent samples.

In Approach 3, the Duke responses are calibrated using the relationship between Duke and Monogram readouts of the convalescent samples. Specifically, a linear regression model is used to regress the Monogram readouts on the Duke readouts; the calibrated Monogram scale value of a Duke sample is calculated by plugging in the observed Duke vaccine readout as the predictor in the regression model.

The concordance correlation coefficient (CCC)[42] is used to quantify the level of agreement between the calibrated values of the two labs using samples measured by both labs. The overall CCC is then calculated as the average time-specific CCC using an inverse hyperbolic tangent transformation for stability. For both Approaches 2 and 3, to account for the statistical variability in the calibrated responses, 1000 bootstrap samples of the convalescent paired sample dataset are drawn with replacement to estimate the precision of the estimated CCC.

References

1. Connors, M., Graham, B. S., Lane, H. C. & Fauci, A. S. SARS-CoV-2 Vaccines: Much Accomplished, Much to Learn. *Ann. Int. Med.* **174**, 687–690 (2021).
2. McGill COVID19 Vaccine Tracker Team COVID-19 Vaccine Tracker. https://covid19.trackvaccines.org/ Last updated 5 Mar, 2021. Access date 8 Mar, 2021.
3. Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl. J. Med.* **384**, 403–416 (2021).
4. Voysey, M. *et al.* Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* (2020).
5. Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* (2020).
6. Sadoff, J. *et al.* Safety and efficacy of single-dose Ad26.COV2.S Vaccine against Covid-19. *N Engl. J. Med.* **384**, 2187–2201 (2021).
7. Heath, P. T. *et al.* Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *N Engl. J. Med.* **385**, 1172–1183 (2021).
8. US Food and Drug Administration FDA Approves First COVID-19 Vaccine: *Approval Signifies Key Achievement for Public Health.* FDA News Release. Aug 23, 2021. https://www.fda.gov/news-events/press-announcements/fda-approves-first-covid-19-vaccine. Accessed 25 Aug 2021
9. US Food and Drug Administration COVID-19 Vaccines. COVID-19 Vaccines Authorized for Emergency Use. https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines#news Last updated 11 Mar, 2021. Accessed 17 Mar 2021.
10. World Health Organization Coronavirus Disease (COVID-19): COVID-19 vaccine EUL issued. Available at https://extranet.who.int/pqweb/vaccines/covid-19-vaccines Access date 15 Jul, 2021. (2021).
11. U.S. Department of Health and Human Services, Food and Drug Administration, and Center for Biologics Evaluation and Research. Emergency Use Authorization for Vaccines to Prevent COVID-19: Guidance for Industry. Appendix 2: pp 18–22. Issued on 25 May, 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/emergency-use-authorization-vaccines-prevent-covid-19.
12. World Health Organization ADDENDUM to Considerations for Evaluation of COVID-19 Vaccines for Prequalification or Emergency Use Listing. Considerations for evaluation of modified COVID-19 vaccines: Points to consider for manufacturers of COVID-19 vaccines. version 12 Mar, 2021. Available at https://extranet.who.int/pqweb/sites/default/files/documents/Addendum_Evaluation_Modified_Covid-19%20Vaccine.pdf Access date 15 Jul, 2021. (2021).
13. Dal-Re, R. & Caplan, A. L. Current COVID-19 vaccine trials in high-income countries: are placebo-controlled trials ethical?. *Clin. Microbiol. Infect.* **27**, 1565–1567 (2021).
14. Fleming, T. R., Krause, P. R., Nason, M., Longini, I. M. & Henao-Restrepo, A. M. M. COVID-19 vaccine trials: The use of active controls and non-inferiority studies. *Clin. Trials* **18**, 335–342 (2021).
15. Plotkin, S. A. Correlates of protection induced by vaccination. *Clin. Vaccine Immunol.* **17**, 1055–1065 (2010).
16. Plotkin, S. A. & Gilbert, P. B. Nomenclature for immune correlates of protection after vaccination. *Clin. Infect. Dis.* **54**, 1615–1617 (2012).
17. Plotkin, S. A. & Gilbert, P. B. Chapter three: correlates of protection. In *Plotkin's Vaccines* 7th edn (eds Plotkin, S. A. *et al.*) 35–40 (Elsevier, Inc., Philadelphia, PA, 2018).
18. Addetia, A. *et al.* Neutralizing antibodies correlate with protection from SARS-CoV-2 in humans during a fishery vessel outbreak with a high attack rate. *J. Clin. Microbiol.* **58**, e02107 (2020).
19. Cromer, D. *et al.* Relating in vitro neutralisation level and protection in the CVnCoV (CUREVAC) trial. https://doi.org/10.1101/2021.06.29.21259504. *medRxiv*, 2021.2006.2029.21259504 Accessed 2 Jul 2021.
20. Earle, K. A. *et al.* Evidence for antibody as a protective correlate for COVID-19 vaccines. *Vaccine* **39**, 4423–4428 (2021).
21. Feng, S. *et al.* Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 2032–2040, https://doi.org/10.1038/s41591-021-01540-1 (2021).
22. He, X. *et al.* Low-dose Ad26.COV2.S protection against SARS-CoV-2 challenge in rhesus macaques. *Cell* **184**, 3467–3473. https://doi.org/10.1016/j.cell.2021.05.040 (2021).
23. Khoury, D. S. *et al.* Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat. Med.* **27**, 1205–1211. https://doi.org/10.1038/s41591-021-01377-8 (2021).

24. McMahan, K. *et al.* Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* **590**, 630–634 (2021).
25. Yu, J. *et al.* DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* **369**, 806–811 (2020).
26. Gilbert, P. B. *et al.* Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science*, eab3435. https://doi.org/10.1126/science.abm3425 (2021).
27. Voysey, M. *et al.* Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *Lancet* **397**, 881–891 (2021).
28. US Food and Drug Administration Q2 (R1) Validation of Analytical Procedures: Text and Methodology. March 1995. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/q2-r1-validation-analytical-procedures-text-and-methodology Access date: 5 Aug 2021. Content current as of: 6 Apr 2019.
29. European Medicines Agency ICH Q2 (R1) Validation of analytical procedures: text and methodology. Last update: 1/6/1995. Available at: https://www.ema.europa.eu/documents/scientific-guideline/ich-q-2-r1-validation-analytical-procedures-text-methodology-step-5_en.pdf. Accessed 8 June 2021.
30. Mattiuzzo, G. *et al.* Establishment of the WHO International Standard and Reference Panel for anti-SARS-CoV-2 antibody. 2020, WHO Expert Committee on Biological Standardization. WHO/BS/2020.2403. (2020).
31. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
32. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
33. Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280 (2020).
34. Kristiansen, P. A. *et al.* WHO International Standard for anti-SARS-CoV-2 immunoglobulin. *Lancet* **397**, 1347–1348 (2021).
35. World Health Organization Weekly epidemiological update on COVID-19 - 10 August 2021. Edition 52. 10 Aug, 2021 | Emergency Situational Updates. Available here: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---10-august-2021. Accessed Aug 13 2021.
36. Shen, X. *et al.* SARS-CoV-2 variant B117 is susceptible to neutralizing antibodies elicited by ancestral spike vaccines. *Cell Host Microbe* **29**, 529–539 (2021).
37. Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).
38. Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 - Preliminary Report. *N Engl. J. Med.* **383**, 1920–1931 (2020).
39. Anderson, E. J. *et al.* Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *N Engl. J. Med.* **383**, 2427–2438 (2020).
40. National Institute for Biological Standards and Control (NIBSC) Instructions for use of First WHO International Standard for anti-SARS-CoV-2 Immunoglobulin (Version 3.0, Dated 17/12/2020) NIBSC code: 20/136 https://www.nibsc.org/science_and_research/idd/cfar/covid-19_reagents.aspx Access date Jul 29, 2021.
41. Huang, Y., Huang, Y., Moodie, Z., Li, S. & Self, S. Comparing and combining data across multiple sources via integration of paired-sample data to correct for measurement error. *Stat. Med.* **31**, 3748–3759 (2012).
42. Lin, L. I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268 (1989).

Acknowledgements

Author contributions

Conceptualization: Y.H., O.B., P.B.G., C.J.P., and D.C.M. Formal analysis: J.J.K. and Y.H. Investigation: T.W., S.C., M. S.-K., C.M., A.E., R.P., J.H., K.G., and S.K. Supervision: Y.H., L.C., P.B.G., C.J.P., and D.C.M. Funding acquisition: Y.H., P.B.G., M.J.M., L.C. Writing – Original draft: Y.H., O.B., and L.N.C. Writing – Review and editing: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-03154-6.

Correspondence and requests for materials should be addressed to Y.H. or D.C.M.

Reprints and permissions information is available at www.nature.com/reprints.