eAppendices to: Test-Negative Design and Case-Control studies with population controls during widespread testing of symptomatic persons for the presence of SARS-CoV-2

Jan P Vandenbroucke[1,2,3], Elizabeth B Brickley[1],

Christina M.J.E. Vandenbroucke-Grauls[4], Neil Pearce[1]

(1) Departments of Medical Statistics, Non-communicable Disease Epidemiology and Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

(2) Department of Clinical Epidemiology, Leiden University Medical Center, The Netherlands

(3) Department of Clinical Epidemiology, Aarhus University, Denmark

(4) Department of Medical Microbiology and Infection Prevention, Amsterdam UMC, The Netherlands

Corresponding Author:

Jan P Vandenbroucke, Leiden University Medical Center, Dept. Clinical Epidemiology, PO Box 9600, 2300 RC Leiden, The Netherlands, E-mail: J.P.Vandenbroucke@lumc.nl,

eAppendix A: Assumptions that underly generalizability of estimates of a Test-Negative Designs to a broader population

The TND design has been described in detail in numerous papers. In a recent paper two of the authors of the present paper presented this study design as a variant of 'case-control studies with other disease controls'. [1] These include hospital-based case -control studies in which the cases are identified through a particular health-care facility, and the controls are other patients at the same health-care facility.

Case-control studies with diseased controls have a long history in epidemiology, from the first case-control studies of smoking and lung cancer (where lung cancer patients were compared to other patients of the same wards), to several applications in pharmacoepidemiology,[2] and applications in cancer registries.[3] Such studies are recognized as having produced some major and valid findings, such as the association between smoking and lung cancer.[4] However, they rest upon the assumption that the main risk factor under study does not cause the 'other diseases' which are used as controls.

The main difference between the TND in general and other case-control studies with diseased controls is that in the TND the cases and controls are persons with similar signs and symptoms, but who 'test' differently on a crucial test for a particular disease. As such, the controls (those who test negative) will usually have another disease with similar signs and symptoms to the disease under study.

This means that in the particular context of the application of the TND to risk factors for COVID-19, the standard assumption of the hospital-based case-control study, that the exposure of interest is not a cause of the other disease, is not true for many exposures. The other diseases will mainly be other respiratory infections, and several of their risk factors will be similar to those for COVID-19. As mentioned in the paper, the TND will not detect risk factors that are of equal strength for COVID-19 and the diseases of the controls; however, it will detect differences in magnitude of a particular risk factor (say, crowding), or risk factors that would be totally specific for COVID-19. This situation is similar to the situation of another TND on urinary tract infection with antibiotic resistant bacteria (Extended Spectrum Beta Lactamase producing bacteria) in contrast to sensitive bacteria, with added population controls to both groups.[5] In that study, recent hospitalizations proved to be a risk factor that was specific for acquiring infections with resistant bacteria; other findings could only be interpreted by combining the TND with the population controls (e.g., the effect of maleness which was a strong positive risk factor in the TND analysis, but the inverse was true in the comparison with the population).

The basic question is whether the OR generated by a TND, for a factor that is a risk factor for the disease of interest and not for the control diseases, can in principle be the same as the OR that would have been generated by a population-based case-control design. This question has either been answered with strong doubts by some[6] because the TND does not have a well-defined underlying source population; others have answered it positively, some with a few qualifications;[7-9] some with more qualifications.[10] In essence, we agree with the affirmative answer to that question because the TND permits a positive answer to one 'sentinel question' about the validity of a control group: would this person, who does not have the disease, have presented him/herself for testing in that same facility if s/he would have had the disease under study.[1] The answer for a TND

is obviously positive. As a minimum, the results of the TND are valid for the population of 'the tested'.

An additional subtlety is that in a sense, one might even state that in exceptional circumstances where the inclination to be tested is highly linked to characteristics of the tested person, or to suspicions or habits of doctors who serve a particular segment of the population, the standard case-control study with 'test-positive cases' and random general population controls could produce wrong answers, since the type of persons going for testing at particular medical facilities might be self-selected in a way that the random sample of the underlying population is not. [If they were billionaires with doctors catering for that category of persons, there would be habits that would not differ from other billionaires but might differ quite a lot from the general population]. The TND remedies that situation by its design by choosing controls with the same selection pressures, and will in such circumstances produce a valid answer for the tested.

The next question is whether the OR for the TND (= those who go for testing) is applicable to the general population from which it is drawn, or for other populations, again for risk factors that are not known risk factors for the control diseases. It is possible that the TND may deliver unbiased odds ratios for the source population of the 'tested' (i.e. everyone who would have been tested if they had developed symptoms), but these findings may not be completely 'transportable' to the wider general population or other populations – for example, the effect might be higher among the tested than among the general population. However, there would only be lack of generalizability if there were strong effect modification, and the distribution of effect modifiers was markedly different in other populations.[11]

For example, is it possible that a genetic a risk factor for Covid-19 is found in a TND study, but not in the wider population? This seems unlikely. More importantly, these issues of generalizability to other populations is an issue in many case-control studies, and is not unique to the TND. For example, one might ask whether the original studies on smoking and lung cancer in (white Anglo-Saxon) male British doctors in the 1960s would be equally applicable to lawyers, actors, sewage cleaners, or to woman and non-white non-Anglo-Saxon persons in general. Issues of generalization apply to all case-control study findings, not just the TND.

eAppendix B. Sensitivity analysis: quantification of misclassification bias in TND due to false test-negatives and false test-positives

If one assumes particular values of Sn, Sp in a particular study, then one can estimate the bias that is introduced by misclassification of the disease. This can be used to do a sensitivity analysis where we assume a range of possible values of Sn, and Sp. In theory, one can then do an adjustment. However, given the uncertainties involved, such corrections are usually conducted as additions to the main analysis (which would report the unadjusted associations as the main results but also mention the sensitivity analysis findings), rather than substitutes for it.

This Appendix will explain first the situation of disease misclassification in a TND study that is intended to contrast COVID-19 patients vs. patients with other respiratory disorders (mostly other viral infections); second it will explain how different proportions of COVID-19 vs. other respiratory disorders will influence the role of sensitivity and specificity; and thirdly, that it is possible to calculate back from an observed 2x2 table to the 'true' table, given assumptions on sensitivity and specificity.

One may start from the assumption that the specificity of the RT-PCR is very high, except for mishandling of specimens, mishandling of test information (mixing up of persons), or batch contamination. In contrast, the sensitivity may depend on several factors such as the severity and duration of the disease, the way the sample was obtained, as well as the performance of the test itself (See paper).

To see what might happen in the TND situation, a simple numerical example is useful. All persons with COVID-19 with a number of characteristics (a certain degree of symptoms, and a certain degree of worry, personality, access to health care etc.), will have presented themselves to a particular test facility. If only 70% of them will test positive (sensitivity is 70%), the remaining 30% will be false-negatives, but as they have all presented themselves, they will be classified as test-negatives along with the group of true test-negatives, i.e. persons with other respiratory diseases. If we assume that the number of false-positives is negligible, and suppose that at a particular test-site there are 70 test-positives and 150 test negatives, then the true number of COVID-19 patients will be 70/0.7 = 100. This means that 30 of the 150 test-negatives are actually false-negatives, which is 20%. There will thus be 100 COVID-19 patients out of 220 tested. Thus, all ORs on the observed numbers will be biased towards the null. In principle, we can recalculate the true OR without test misclassification by a standard procedure that surmises that the false-negatives amongst the test-negatives should have the same exposure frequencies as the test-positives (i.e. misclassification is non-differential).[1]

In this example, we have not yet taken into account that a (probably small) number of the test-positives are actually false-positives, which in this numerical example is negligible, but that is not

[1] Suppose that 42 of test-positives were men (60%), and only 50% of test negatives were men, this yields an OR for maleness of 1.5. The real percentage of males amongst the true test-negatives can be calculated as 'X' in the equation: 0.50 = (0.8 x X) + (0.2 x 0.60), which gives an X of 0.475. Then the true odds ratio becomes: (0.60 x 0.525)/(0.475 x 0.40) = 1.66 (rounded off).

always the case, as this depends on the proportion of true COVID-19 patients among the persons tested.

To demonstrates what happens with mutual misclassification, in different scenarios of prevalence of COVID-19 relative to prevalence of other respiratory disorders amongst the persons tested, we can make Table 1, following a reasoning similar to that by Flegal et al.[12] From Table 1 it becomes apparent that if the prevalence of current COVID-19 infection is high relative to the prevalence of other respiratory viral diseases amongst the tested (say, in summer months at the peak of an epidemic in a population without immunity), the main driver of the misclassification is the sensitivity of the test: a large proportion of test-negatives might actually be false negatives. Inversely, in the situation of an almost disappearing epidemic in winter months (say, when a vaccine would be available and it would be winter), the number of test-positives that would actually be false-positives will become quite large, i.e., the specificity becomes more important. In each of these extreme situations, there is misclassification, but it is less severe in the latter instances (sporadic COVID-19).

It is possible to make approximate assumptions about the ratio of the estimated OR to the true OR, based on tables like this, with appropriate estimates of sensitivity and specificity and relative prevalences of non-Covid-19 respiratory diseases amongst the tested.

TABLE 1: Variation of observed OR for different True COVID-19 percentages amongst the tested with the same sensitivity, specificity, and the same exposure frequency in true test-negatives (assumed to be 50%), and same true underlying OR of 2.0.

Fraction TRUE COVID-19	SENS	SPEC	TRUE Odds Ratio (rounded)	**OBSERVED Odds Ratio (rounded)**
0,97	0,70	0,99	2,00	**1,07**
0,90	0,70	0,99	2,00	**1,21**
0,50	0,70	0,99	2,00	**1,69**
0,10	0,70	0,99	2,00	**1,80**
0,03	0,70	0,99	2,00	**1,58**

Finally, there is the possibility to calculate back from an observed 2x2 table to the “true” 2x2 table, given assumptions on sensitivity and specificity. See Chapter 6 on “Outcome misclassification in Lash et al.[13] This “correction”, is based on a single observed 2x2 table, when there has been confounder adjustment, the adjusted odds ratio will differ from the crude odds ratio, and therefore one cannot use the crude 2x2 table for the “correction” – more complex methods are available if one wishes to also correct for confounders while “correcting” for misclassification.[14]

eAppendix C: Reason for not omitting controls that are asymptomatic carriers of SARS-CoV-2

This reasons for not omitting controls that are asymptomatic carriers of SARS-CoV-2 can be illustrated by considering a hypothetical study. Suppose we could identify the population (a subgroup of the general population) which would come for testing if they had symptoms. Ideally, one would then test all of this population and we would estimate the risk of infection in this population, and in various subgroups. In the population (P) there might be a certain number of people (T) who tested positive. Note that the population denominator (P) includes both people who currently have symptomatic infections and people who don't – it is just the total population 'at risk'. The risk of having a symptomatic infection is then T/P. If we compare two subgroups who are exposed or not-exposed to a particular factor, their risks might be T_1/P_1 and T_0/P_0 respectively, and the risk ratio (the ratio of these two proportions) would indicate the relative risk of symptomatic Covid-19 infection in the exposed and non-exposed (e.g. if the RR was 2.0 then the exposed would be twice as likely as the non-exposed to have symptomatic Covid-19 infection).

A case-control study involves studying all of the 'cases' (i.e. T) and a sample of the population which generated them (P). Thus, provided that the controls are a representative sample of the source population P (i.e. everyone who would have come for testing if they had symptoms – which is a reasonable assumption to make since they came with someone who was being tested), then the odds ratio for Covid-19 infection in the case-control study will estimate the risk ratio in the population (P) which generated the cases (T). Of course, a small number of the controls may have asymptomatic Covid-19 infection, and would have tested positive if they had been tested. But this is not a problem – they would have been part of the denominator (P) if a full population survey had been conducted, and they are therefore eligible to be selected as controls. This is analogous to the case-cohort (case-base) design which is a commonly used design for case-control studies .[15, 16]

eAppendix D. Differentiation of risk factors between TND and added case-control studies by triangulation

This is a worked out example from a paper about urinary tract infections with resistant vs. sensitive bacteria.[5] In this example, the TND analysis is a contrast between patients with infections with resistant bacteria *(= test-positives)* vs. infections with sensitive bacteria *(= test-negatives).* For each group, the test-positives and the test-negatives, there is a comparison with the general population.

The TND analysis as well as the comparison between the test-positives and the population is demonstrated on partial data from the original paper which are given below as Appendix D Table 1. The other comparison, between test-negatives and the general population was in the Appendix of the original paper, and is also partially given below at the right hand side of Appendix D Table 2. Male sex is a strong risk factor for infections with antibiotic resistant bacteria in the TND analysis, but completely the reverse happens in the comparison with the general population. This makes sense: in the general population women have much more urinary tract infections than men, but men when they have urinary tract infections are generally older patients (for example, patients with prostatic disease who go in- and out of hospitals and acquire resistant bacteria in hospitals). In the data on the other comparison, between test-negatives and the general population, the OR for being a woman (the reverse of the OR for begin a man) is even more extreme, which is logical because the test-positives are removed from that comparison. If one multiplies the OR for sex of the TND with the population comparison of the test-negatives, one arrives in the direction of the population comparison with the test-positives.

eAppendix D Table 1: Partial table of risk factors for urinary tract infection (UTI) caused by extended-spectrum beta-lactamase-producing (ESBL) *E. coli* compared with non-ESBL *E. coli* UTI controls *(a test-negative design comparison)* and population controls *(case-control comparison of test-positives to general population)*

	OR for ESBL *E. coli* UTI versus non-ESBL *E. coli* UTI (95% CI)		OR for ESBL *E. coli* UTI vs. population controls (95% CI)	
Characteristic	Adjusted for age, sex, and comorbidity	Multivariate adjusted	Adjusted for age, sex, and comorbidity	Multivariate adjusted*
Demographic characteristics				
10 years increase in age	0.90 (0.85 - 0.96)	0.92 (0.86 - 0.99)	1.25 (1.16 - 1.35)	1.16 (1.06 - 1.27)
Male sex	1.55 (1.20 - 2.02)	1.62 (1.22 - 2.14)	0.29 (0.23 - 0.38)	0.48 (0.34 - 0.67)
Citizenship: Northern Europe vs. other countries	0.37 (0.23 - 0.60)	0.40 (0.24 - 0.66)	0.54 (0.33 - 0.87)	0.39 (0.22 - 0.68)
Marital status				
Divorced or widowed vs. married	1.04 (0.79 - 1.37)	--	1.18 (0.87 - 1.59)	--
Never married vs. married	0.82 (0.57 - 1.19)	--	1.36 (0.95 - 1.96)	--
Prior urinary tract infection				
Within 31-180 days before index date	1.27 (1.01 - 1.61)	--	19.1 (14.3 - 25.4)	--
Within 181-365 days before index date	1.19 (0.94 - 1.50)	--	10.6 (7.98 - 14.0)	--

* Adjusted for more factors than shown here; see original publication.[5]

eAppendix D Table 2. Partial table of risk factors for *E. coli* urinary tract infection (UTI) with and without ESBL compared with population controls *(two comparisons of test-positives and test-negatives with general population; first comparison same as in left hand side of Appendix D Table 1)*

Characteristic	OR for ESBL *E. coli* UTI vs. population controls (95% CI)		OR for Non-ESBL *E. coli* UTI vs. population controls (95% CI)	
	Adjusted for age, sex, and comorbidity	Multivariate adjusted*	Adjusted for age, sex, and comorbidity	Multivariate adjusted*
Demographic characteristics				
10 years increase in age	1.25 (1.16 - 1.35)	1.16 (1.06 - 1.27)	1.34 (1.30 - 1.38)	1.24 (1.19 - 1.28)
Male sex	0.29 (0.23 - 0.38)	0.48 (0.34 - 0.67)	0.19 (0.17 - 0.21)	0.23 (0.20 - 0.27)
Citizenship: Northern Europe vs. other countries	0.54 (0.33 - 0.87)	0.39 (0.22 - 0.68)	1.45 (1.10 - 1.93)	1.21 (0.90 - 1.64)
Marital status				
Divorced or widowed vs. married	1.18 (0.87 - 1.59)	--	1.14 (0.98 - 1.32)	--
Never married vs. married	1.36 (0.95 - 1.96)	--	1.59 (1.35 - 1.87)	--
Prior urinary tract infection				
Within 31-180 days before index date	19.1 (14.3 - 25.4)	--	15.0 (12.3 - 18.2)	--
Within 181-365 days before index date	10.6 (7.98 - 14.0)	--	8.90 (7.37 - 10.7)	--

* Adjusted for more factors than shown here; see original publication.[5]

REFERENCES TO eAPPENDICES

1. Vandenbroucke, J.P. and N. Pearce, *Test-Negative Designs: Differences and Commonalities with Other Case-Control Studies with "Other Patient" Controls.* Epidemiology, 2019. **30**(6): p. 838-844.
2. Jick, H. and M.P. Vessey, *Case-control studies in the evaluation of drug-induced illness.* Am J Epidemiol, 1978. **107**(1): p. 1-7.
3. Smith, A.H., N.E. Pearce, and P.W. Callas, *Cancer case-control studies with other cancers as controls.* Int J Epidemiol, 1988. **17**(2): p. 298-306.
4. Doll, R. and A.B. Hill, *Smoking and carcinoma of the lung; preliminary report.* Br Med J, 1950. **2**(4682): p. 739-48.
5. Sogaard, M., et al., *Risk factors for extended-spectrum beta-lactamase-producing Escherichia coli urinary tract infection in the community in Denmark: a case-control study.* Clin Microbiol Infect, 2017. **23**(12): p. 952-960.
6. Westreich, D. and M.G. Hudgens, *Invited Commentary: Beware the Test-Negative Design.* Am J Epidemiol, 2016. **184**(5): p. 354-6.
7. Sullivan, S.G., E.J. Tchetgen Tchetgen, and B.J. Cowling, *Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness.* Am J Epidemiol, 2016. **184**(5): p. 345-53.
8. Jackson, M.L. and J.C. Nelson, *The test-negative design for estimating influenza vaccine effectiveness.* Vaccine, 2013. **31**(17): p. 2165-8.
9. Jackson, M.L., et al., *The impact of selection bias on vaccine effectiveness estimates from test-negative studies.* Vaccine, 2018. **36**(5): p. 751-757.
10. Lewnard, J.A., et al., *Measurement of Vaccine Direct Effects Under the Test-Negative Design.* Am J Epidemiol, 2018. **187**(12): p. 2686-2697.
11. Pearl, J. and E. Bareinboim, *Note on ''Generalizability of Study Results''.* Epidemiology, 2019. **30**(2): p. 186-188.
12. Flegal, K.M., C. Brownie, and J.D. Haas, *The effects of exposure misclassification on estimates of relative risk.* Am J Epidemiol, 1986. **123**(4): p. 736-51.
13. Lash, T.L., M.P. Fox, and A.K. Fink, *Applying Quantitative Bias Analysis to Epidemiologic Data*. 2009: Springer-Verlag New York.
14. Corbin, M., et al., *A comparison of sensitivity-specificity imputation, direct imputation and fully Bayesian analysis to adjust for exposure misclassification when validation data are unavailable.* Int J Epidemiol, 2017. **46**(3): p. 1063-1072.
15. Pearce, N., *What does the odds ratio estimate in a case-control study?* Int J Epidemiol, 1993. **22**(6): p. 1189-92.
16. Vandenbroucke, J.P. and N. Pearce, *Case-control studies: basic concepts.* Int J Epidemiol, 2012. **41**(5): p. 1480-1489.