Online supplementary file 3. Design classification and how it is related to critical appraisal tools

We have adopted the description of different case-control designs described by Rutjes et al[1], which comprises four categories: single gate classic, single gate with reverse-flow design, two-gate design with normal controls, and two-gate design with alternative diagnosis (Table A). Although two-gate designs are common, they carry inherent challenges, such as the likelihood of producing inflated estimates of diagnostic accuracy. This can happen for sensitivity when individuals are all quite sick or for specificity when all controls are quite healthy[1].

Table A. Types of case-control design according to Rutjes et al[1]

Single gate: Classic	• Both cases and controls are sampled from a single source population. • The index test is performed first, followed by the reference standard. This allows us to identify those with and without a condition • All patients pass through a single gate: a single set of criteria for study admission, typically defined by the clinical presentation (suspicion of disease).
Single gate: Reversed-flow design	• Both cases and controls are sampled from a single source population • The reference standard is performance first, and this allows us to differentiate cases and controls. Then, the index test is carried out • All patients pass through a single gate: a single set of criteria for study admission, typically defined by the clinical presentation (suspicion of disease).
Two-gate design with healthy controls	• Cases and controls are sampled from two distinct source populations • The index test is carried out for each group • Two different sets of inclusion criteria (gates) are used: one for the diseased and another for the non-diseased participants. Cases are typically defined by clinical presentation/suspicion of disease, while controls need to be healthy
Two-gate design with alternative diagnosis controls	• Cases and controls are sampled from two distinct source populations • The index test is carried out for each group • Two different sets of inclusion criteria (gates) are used: one for the diseased and another for an alternative diagnosis. Cases are typically defined by clinical presentation/suspicion of disease, while controls need to meet criteria for another condition, which is often known to produce symptoms and signs similar to those of participants with the target condition

Source: Adapted from Rutjes et al 2005[1]

Poor reporting was a key issue when trying to classify the papers. Nonetheless, only 15 studies were solely classified as a single-gate study (all adopted a reversed-flow design). Over three-quarters of studies investigating healthy controls sourced them from a distinct population (two-gate design). The proportion was much lower for non-malignant/pre-malignant conditions (about a third), but there was much more missing data.

How classification by study design relates to the QUADAS-2 quality assessment checklist for Diagnostic Accuracy Studies[2]:

- *Was a case-control design avoided?* While extracting data and classifying the studies, it was clear that a case-control design was only avoided by one study[3].

- *Were the index test results interpreted without knowledge of the results of the reference standard?* For studies adopting a reversed-flow design, the reference standard was always carried out first so cases could be identified. This was also most often the case for cancer patients in studies adopting a two-gate design. Overall, less than five studies indicated that those interpreting the index test were blinded to reference test results.

- *Were the reference standard results interpreted without knowledge of the results of the index test?* For cancer cases, reference standard results were most often interpreted without knowledge of results of the index test, across all study designs.

- *Did all patients receive the same reference standard?* Cancer cases often received the same reference standard. This was not the case for almost all controls as they seldom received the reference standard (particularly healthy controls in two-gate designs).

References:

[1]Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. Clin Chem. 2005;51(8):1335-1341. doi:10.1373/clinchem.2005.048595

[2]Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-36.

[3]Yanaoka K, Oka M, Mukoubayashi C, Yoshimura N, Enomoto S, Iguchi M, et al. Cancer high-risk subjects identified by serum pepsinogen tests: outcomes after 10-year follow-up in asymptomatic middle-aged males. Cancer Epidemiology, Biomarkers & Prevention. 2008;17(4):838-45.