## Appendix: (a) Levels of evidence as described by Oxford Centre for Evidence-Based Medicine

| Level | Therapy / Prevention, Etiology / Harm | Differential diagnosis / symptom prevalence study |
|---|---|---|
| 1a | Systematic review (with homogeneity*) of randomized control trials | Systematic review (with homogeneity*) of prospective cohort studies |
| 1b | Individual randomized control trial (with narrow Confidence Interval") | Prospective cohort study with good follow-up**** |
| 1c | All or none§ | All or none case-series |
| 2a | Systematic review (with homogeneity*) of cohort studies | Systematic review (with homogeneity*) of 2b and better studies |
| 2b | Individual cohort study (including low quality randomized control trial; e.g., < 80% follow-up) | Retrospective cohort study, or poor follow-up |
| 2c | "Outcomes" Research; Ecological studies | Ecological studies |
| 3a | Systematic review (with homogeneity*) of case-control studies | Systematic review (with homogeneity*) of 3b and better studies |
| 3b | Individual Case-Control Study | Non-consecutive cohort study, or very limited population |
| 4 | Case-series (and poor quality cohort and case-control studies§§) | Case-series or superseded reference standards |
| 5 | Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" | Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" |

Excerpt from http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009

A minus-sign "-" can be added to denote the level of that fails to provide a conclusive answer because:
***EITHER*** a single result with a wide Confidence Interval
***OR*** a Systematic Review with troublesome heterogeneity.

Such evidence is inconclusive, and therefore can only generate Grade D recommendations.

| | |
|---|---|
| * | By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level. |
| " | Clinical Decision Rule. (These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category.) |
| "¡ | See note above for advice on how to understand, rate and use trials or other studies with wide confidence intervals. |
| § | Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it. |
| §§ | By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders. |
| §§§ | Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples. |
| " " | An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis. |
| "¡"¡ | Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits. |
| " " " | Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor |

| | |
|---|---|
| | reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study. |
| ” ” ” “ | Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and the equally or more expensive. |
| ** | Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'. |
| *** | By poor quality prognostic cohort study we mean one in which sampling was biased in favor of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors. |
| **** | Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (for example 1–6 months acute, 1 – 5 years chronic) |

**Grades of Recommendation**

| | |
|---|---|
| A | consistent level 1 studies |
| B | consistent level 2 or 3 studies ***or*** extrapolations from level 1 studies |
| C | level 4 studies ***or*** extrapolations from level 2 or 3 studies |
| D | level 5 evidence ***or*** troublingly inconsistent or inconclusive studies of any level |

*"Extrapolations" are where data is used in a situation that has potentially clinically important differences than the original study situation.*

## (b) Levels of evidence assigned to published papers reporting on *C. difficile* in Japan

| Reference | Study type | Level of evidence* |
|---|---|---|
| Akahoshi 2016 | Retrospective chart review (*n* = 308, HSCT) | 2b |
| Arimoto 2016 | Systematic review and meta-analysis | 2a |
| Cairns 2017, plus erratum | Bench research (277 isolates) | 5 |
| Collins 2013 | Systematic review | 3a- |
| Daida 2017 | Retrospective case control review (*n* = 189) | 3b |
| Ekma 2012 | Systematic review | 3a- |
| Fujimori 2015 | Short review | 5 |
| Furuichi 2014 | Prospective NI cohort (*n* = 346, pediatric) | 2b |
| Hashimoto 2007 | Retrospective chart review (*n* = 242, LDLT) | 2b |
| Hata 2016 | Phase 3, multicenter, open-label RCT (*n* = 579, colorectal surgery) | 2b |
| Hikone 2015 | Retrospective chart review (*n* = 2193 samples) | 4 |
| Honda 2014 (Curr Opin Infect Dis) | Review‡ (SOT) | 5 |
| Honda 2014 (Anaerobe) | Retrospective chart review (*n* = 22,863) | 2b |
| Hosokawa 2014 | Retrospective cohort (*n* = 201, HSCT) | 2b |
| Igawa 2016 | Bench research | 5 |
| Imase 2008 | RCT (*n* = 19, peptic ulcer) | 2b |

| | | |
|---|---|---|
| Iwamoto 2012 | Prospective observational cohort (*n* = 1226, rheumatology) | 2b |
| Iwashima 2010 | Retrospective cohort (*n* = 610) | 2b |
| Kaneko 2011 | Retrospective cohort (*n* = 137, UC) | 2b |
| Kato 2009 | Prospective cohort (*n* = 22 samples, 17 patients) | 4 |
| Kato 2010 | Prospective cohort (*n* = 160 samples) | 2b |
| Kawada 2011 | Prospective cohort (*n* = 22 samples, 17 patients) | 1b |
| Kikkawa 2007 | Prospective observational | 3b |
| Kiyosuke 2015 | Bioinformatics | 5 |
| Kobayashi 2014 | Retrospective cohort (*n* = 997) | 3b |
| Kobayashi 2017 | Multicenter, retrospective cohort study/chart review (*n* = 160) | 2b |
| Komatsu 2016 | Single-center RCT (*n* = 379, colorectal surgery) | 2b |
| Kunishima 2013 | Bench research (*n* = 157 *C. difficile* isolates) | 5 |
| Kuwata 2015 | Bench research (*n* = 130) | 5 |
| Matsuda 2012 | Bench research (*n* = 83) | 5 |
| Mikamo 2018 | Multicenter, double-blind, Phase 2 RCT (*n* = 93) | 2b |
| Miura 2011 | Bench research (*n* = 26 isolates) | 5 |
| Mizui 2013 | Retrospective cohort (*n* = 29 *C. difficile* diarrhea) | 2b |
| Mori and Aoki 2015 | Retrospective case-control (*n* = 208) | 4 |
| Mori 2015 | Prospective cohort (*n* = 975 samples) | 3b |
| Nomura 2008 | Retrospective cohort (*n* = 8) | 4 |
| Ogami 2013 | Retrospective cohort (*n* = 463) | 2b |
| Oka 2012 | Bench research (73 isolates) | 5 |
| Oshima 2018 | Systematic review and meta-analysis | 3b |
| Roughead 2016 | Retrospective chart review (*n* = 54,957) | 2b |
| Sadahiro 2014 | Prospective RCT (*n* = 294, colon cancer) | 2b |
| Sasabuchi 2016 | Retrospective cohort (*n* = 70,862, peptic ulcer) | 2b |
| Sasahara 2016 | Prospective cohort (*n* = 71) | 3b |
| Sawabe 2007 | Prospective cohort (148 isolates) | 3b |
| Senoh 2014 | Bench research | 5 |
| Senoh 2015 | Retrospective cohort (159 isolates) | 3b |
| Shimizu 2015 | Prospective cohort (334 fecal samples) | 2b |
| Suzuki 2013 | Prospective cohort (*n* = 80) | 2b |
| Takahashi 2014 | Prospective case-control and cohort (*n* = 1025) | 2b |
| Tojo 2014 | Prospective cohort (*n* = 69 fecal samples) | 1b |
| Watanabe 2008 | Retrospective cohort (*n* = 294 fecal | 3b |

| | | |
|---|---|---|
| | samples‡) | |
| Yasunaga 2012 | Retrospective chart review (*n* = 143,652) | 2b |
| Yokohama 2009 | Retrospective chart review (*n* = 252) | 4 |
| Yuhashi 2016 | Retrospective chart review (*n* = 68 samples) | 4 |

HSCT, hematopoetic stem cell transplantation; LDLT, living donor liver transplant; NI, non-interventional; RCT, randomized control trial; SOT, solid organ transplant; UC, ulcerative colitis

*Based on Oxford Centre for Evidence-Based Medicine – Levels of Evidence. Available at www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/. Accessed September 2016; †As part of study methodology; ‡Samples from Kikkawa 2007