Robust Biomarkers: Methodologically Tracking Causal Processes in Alzheimer's Measurement

Vadim Keyser and Louis Sarry[1]

Abstract:

In biomedical measurement, biomarkers are used to achieve reliable prediction of, and useful causal information about patient outcomes while minimizing complexity of measurement, resources, and invasiveness. A biomarker is an assayable metric that discloses the status of a biological process of interest, be it normative, pathophysiological, or in response to intervention. The greatest utility from biomarkers comes from their ability to help clinicians (and researchers) make and evaluate clinical decisions. In this paper we discuss a specific methodological use of clinical biomarkers in pharmacological measurement: Some biomarkers, called 'surrogate markers', are used to substitute for a clinically meaningful endpoint corresponding to events and their penultimate risk factors. We confront the reliability of clinical biomarkers that are used to gather information about clinically meaningful endpoints. Our aim is to present a *systematic methodology for assessing the reliability of multiple surrogate markers (and biomarkers in general)*. To do this we draw upon the robustness analysis literature in the philosophy of science and the empirical use of clinical biomarkers.

After introducing robustness analysis we present two problems with biomarkers in relation to reliability. Next, we propose an *intervention-based robustness methodology* for organizing the reliability of biomarkers in general. We propose three relevant conditions for a robust methodology for biomarkers: (R1) *Intervention-based demonstration of partial independence of modes*: In biomarkers partial independence can be *demonstrated* through exogenous interventions that modify a process some number of "steps" removed from each of the markers. (R2) *Comparison of diverging and converging results across biomarkers*: By systematically comparing partially-independent biomarkers we can track *under what conditions* markers fail to converge in results, and under which conditions they successfully converge. (R3) *Information within the context of theory*: Through a systematic cross-comparison of the markers we can make causal conclusions as well as eliminate competing theories. We apply our robust methodology to currently developing Alzheimer's research to show its usefulness for making causal conclusions.

---

[1]Vadim Keyser
California State University, Fresno
vkeyser@csufresno.edu

Louis Sarry
lsarry@gmail.com

1. Introduction

In biomedical measurement biomarkers are used to achieve reliable prediction of, and useful information about, patient outcomes while minimizing complexity of measurement, resources, and invasiveness. A biomarker is an assayable metric that discloses the status of a biological process of interest, be it pathophysiological or in response to intervention (De Gruttola et al., 2001). The greatest utility from biomarkers comes from their ability to help clinicians (and researchers) make and evaluate clinical decisions. As such, a clinical biomarker is any biomarker deemed universally applicable across the patient population that discloses the state of a biological process suitable to inform clinical decision-making and its subsequent assessment.

Clinical biomarkers (CBs) are stratified in many ways. For example, by: measurement *purpose* (e.g., the extent of disease, nature of response); *what* they measure (e.g., function, biochemistry/histology, behavior, genetics, morphology); clinical use (e.g., predictive, diagnostic, prognostic); situation within a physiological model (e.g., causative, resultant, correlative); how they can be used to examine biology (e.g., interventional analysis); and composition (i.e. singular, composite, serial).[2]

In this paper we analyze the *methodological use* of CBs in pharmacological measurement: some biomarkers are used to substitute for a clinically meaningful endpoint corresponding to events and their penultimate risk factors. For example, the biomarker LDL cholesterol (LDL-C) corresponds to the clinical endpoint heart attack, with penultimate risk factors such as coronary artery stenosis, atherosclerosis with unstable fibrous cap, and severe angina pectoris. A coveted CB is supposed to provide reliable information regarding processes that cannot reasonably be interrogated directly. Such reliable biomarkers are referred to as 'surrogate markers' (SMs).[3]

The story becomes intriguing when multiple surrogate markers are combined for measurement purposes. It seems that a combination of SMs will provide more useful results than just a single SM, to the extent that that those SMs are independent and do not carry the same confounds. But there is neither a scientific nor a philosophical account that can be used to assess the reliability of multiple biomarkers. In this paper we present a *systematic methodology for assessing the reliability of multiple surrogate markers* (and biomarkers in general). To do this we draw upon the robustness analysis literature in the philosophy of science as well as the clinical use of biomarkers.

In Section 2, we present background on robustness analysis to show that the focus within the philosophy of science literature has been on converging results. At first glance, CBs seem like a perfect candidate to be incorporated into a robustness analysis focused on converging results However, in Section 3 we analyze the reliability of CBs that are

[2] See Mayeux (2004) and Aronson (2005) for classification of CBs.
[3] Katz (2004) describes all biomarkers as being "candidate" surrogate markers.

used to gather information about clinically meaningful endpoints, both within single biomarkers and within combined biomarkers. We argue that even in surrogate markers that are validated, readings of a given surrogate marker may diverge from readings of nominally equivalent, or parallel, surrogate markers. For this reason, we focus on a type of robustness analyses that can account for diverging results between modes of measurement. In Section 4, we present an intervention-based systematic methodology for assessing the reliability of surrogate markers. This methodology draws on two key accounts in the robustness literature (Schupbach 2015 and Keyser 2016), both of which focus on elimination and provide uses for diverging results. In Section 5, we present a case study from Alzheimer's research. We argue that current research in Alzheimer's disease is fitting for specifying causal relations and revising theory based on diverging results of CBs.[4]

2. Robustness Analysis

In this discussion, we present an intervention-based systematic methodology for assessing the reliability of surrogate markers. This methodology draws on two key accounts in the robustness literature (Schupbach 2015 and Keyser 2016), both of which focus on elimination and provide uses for diverging results. Before discussing these robustness analyses, it is important to give a bit of background about the different uses of robustness analysis.

2.1 Robustness Analysis: Uses

Philosophers have discussed robustness analysis in the context of models[5] as well as in the context of experiment and evidence.[6] We can specify three robustness analyses that are relevant for setting up biomarker analysis. Woodward (2006) distinguishes 'inferential robustness,' 'measurement robustness,' and 'causal robustness.' The details of these analyses are important because they can be applied to show how we can differentiate modes in biomarker measurement (Section 4.2) and how we can use robustness analysis to specify causal relationships (Sections 4.3 and 4.4).

Woodward uses Aldrich's (1989) account to characterize 'inferential robustness'. This type of robustness is used when there is a fixed body of data (D) and a conclusion (S). This is for cases with a number of additional assumptions (Ai), with no strong scientific preference for any of them. Woodward says: "A number of writers suggest that

[4] The modeling work for this project was completed in 2015 and 2016 when this was still an unfolding empirical puzzle.
[5] See Wimsatt (2007); Levins (1966); Weisberg (2006); and Glymour (1980).
[6] See Horwich (1982; 2011); Hacking (1983); Franklin (1997); Sober (1989); Trout (1998); Culp (1994); Stegenga (2009; 2012).

if for each of these Ai, D supports S, this provides a strong reason for belief in S, even in the absence of information about which of these Ai is correct—S is said in this case to be robust or sturdy or insensitive to alternative assumptions Ai, given D" (2006, 219-220).[7] If S is not robust, then it is believed to be "fragile" and Woodward suggests a suspension of belief (2006, 220).

'Measurement robustness' consists of using different measurement procedures to measure the value of some quantity (Woodward 2006, 234). One of the important features of measurement robustness is that different procedures are used. This increases our confidence that we have measured the quantity "accurately" (234). Woodward says,

> The underlying rationale is usually taken to be something like this: The different measurement procedures are in some relevant sense or senses 'independent' of each other—they involve instruments of different design, operating according to different causal principles, they employ different assumptions to interpret the data they produce and so on. This suggests that while each procedure may be subject to various sources of error, they are unlikely to be subject to exactly the same kinds of error…" (2006, 234)

Both inferential robustness and measurement robustness increase our confidence about an invariant result. However, the mechanism in each analysis differs: inferential robustness data may be produced by a single procedure, even though the assumptions differ. In a case of measurement robustness, there may be shared assumptions whereas measurement procedures differ.[8]

Finally, 'causal robustness' consists of finding a structural relationship that is invariant over changes in intervention. For example, a regression equation is robust if the causal relationship between the independent and dependent variables remains invariant over some class of manipulations on the independent variable (Woodward 2006, 235).

Woodward's account breaks down important features of robustness analyses, namely, theoretical assumptions, procedures, and causal conditions. We can use these features as building blocks to individuate modes.

2.2 Individuating Modes

[7] Woodward adds that inferential robustness "…is taken to show that D supports S or provides a reason for believing S" (2006, 220).

[8] For example, assumptions in kinetic theory of heat can be used to explain the function of two different thermometry procedures.

There are different ways to differentiate modes[9]. Recall Woodward's characterization of independence in assumptions (inferential robustness) vs. independence in physical procedures (measurement robustness). This offers a precise way of characterizing different accounts of 'independence'. Culp (1994) posits that the use of different background theories for evidential processes contributes to the success of robustness analysis. Stegenga (2009) discusses the absence of shared problematic background assumptions to individuate modes. Specifically, converging modes of evidence are more likely to signal truth, if they do not share problematic background assumptions.[10]

The second approach in characterizing 'independence' draws on Woodward's general characterization of different "procedures". These may differ because of underlying physical principles or processes. For example, optical microscopes and electron microscopes are governed by different physical principles. The electron microscope uses a high voltage electron beam to form the image (transmission electron microscope), or it uses detection of low energy "secondary electrons" emitted by the surface of the object as a result of excitation by the "primary electron beam" (scanning electron microscope). In contrast, light microscopes use techniques involving light—such as interference, polarization, phase contrast, direct transmission, and fluorescence.

Physical principles are not sufficient in differentiating modes. For example, both the scanning electron microscopes and transmission electron microscopes use the same physical principles (Stegenga 2012). However, it matters in what way the physical principles are employed. Both microscopes use high-energy beams of electrons but there is a difference between how, in each of the processes, the beams interact with the samples. Similarly, a given set of biomarkers may use the same physiological pathways, but it matters in what way these biomarkers *interact with a given clinical intervention*. The physical principle account can be modified to a *physical process account*.

Woodward's (2006) specification of causal robustness is useful for this. Causal robustness focuses on invariant relationships over interventions or manipulations in conditions. So, if we observe invariance in some EHR result after we have modified recordings of social and behavior choices, we can specify under what conditions invariance occurs. Trends that survive modified conditions are robust and, to use Woodward's (2006) reasoning, indicate causal structure. We can also specify the

[9] Philosophers have addressed the "individuation" (independence) of modes of evidence See Nagel (1939); Horwich (1982); Franklin (1984); Sober (1989); Trout (1993); Culp (1994); Keeley (2002); Staley (2004); Douglas (2004); Wimsatt (2007); and Stegenga (2009; 2012); Lloyd (2010); Schupbach (2015); Keyser (2016).

[10] For example, consider data on some disease derived from electronic health record (EHR) by clustering health and lifestyle variables, and lab data on the development of the same condition from animals exposed to a specific toxin. The EHR data build on the idealized assumption that patients surveyed in the given sample, do not all share a specific hidden confound that would skew the effect. The lab data build on the idealized assumption that the animal physiology used is relevantly similar to human physiology for this type of effect. Because these assumptions are not shared, convergence of results is less likely to come from the same systematic error.

conditions under which the robust result remains invariant—e.g., some result R occurs under admissible changes in a specified range of conditions C.

This section serves as proper background to set up the methodological puzzle of surrogate markers that we present in the subsequent section. At first glance it seems like robustness analysis and the evaluation of independent modes are straightforward in clinical biomarkers. But what we see is that *both robustness analysis and independence analysis require re-conceptualization to make adequate solutions in the complex and divergence-filled realm of biomarkers*.

3. Surrogate Markers

In this section, we describe surrogate markers (SM) and the methodological difficulty of evaluating their reliability as well as the reliability of unvalidated biomarkers. First, we define the specific role of surrogate markers according to the standards in clinical practice. Second, we describe the problematic use of multiple biomarkers: whether biomarkers are validated or unvalidated, the use of multiple markers often yields discordant or diverging results.

3.1 Surrogate Markers and Clinical Standards

Broadly defined, a SM is any assayable metric, which substitutes for a clinically-meaningful endpoint or penultimate pathological state, relating to important patient-centered outcomes—including affective measures and physiological function or survival (De Gruttola et al. 2001). Because an SM achieves its expected value prior to the consummation of the clinical endpoint—i.e. it precedes it in time—surrogate endpoints are of great utility in contexts where the clinical endpoint is grave and best not sampled in a patient or subject population (e.g. kidney failure). Surrogate endpoints are also of great utility in contexts where the clinical endpoint is beyond our ability to easily or reliably measure it—e.g., it would require a measurement of progression over years, prohibitively expensive equipment, or a technical demand beyond current capabilities.

Before describing a problem for SMs and biomarkers in general, it is important to note that there are stringent tests to establish a marker as a 'validated surrogate maker'. The methodological goal is to have surrogate markers that provide reliable information regarding processes that cannot be interrogated directly. We characterize 'reliability' in two senses: predictive and causal-explanatory. 'Predictive reliability' means that there is a high predictive value, resultant in the diagnostic device having both high sensitivity and high specificity. 'Causal-explanatory reliability' means that the use of biomarkers provides key causal information about a given state of a clinical endpoint.

Predictive reliability: There is a high predictive value resultant in both high sensitivity and high specificity. High sensitivity conditions refer to the minimization of false

negatives, whereby the marker always/often achieves an expected value when a process threshold[11] or clinical endpoint is proximate. High specificity refers to the minimization of false positives, whereby the marker never/rarely achieves an expected value when a process threshold or clinical endpoint is not proximate.

Causal-explanatory reliability: The use of biomarkers provides key *causal information* about a given state of a clinical endpoint. The biomarker provides information that is embeddable within available theoretical models. These models can explain the relations between biomarker, clinical endpoint, and the risk factors associated with the endpoint.

In the literature on surrogate markers, most of the focus is placed on satisfying predictive reliability by looking at the close-knit relationship between surrogate (marker or endpoint) and clinical endpoint. Statistically, surrogate endpoints constitute "response variables" for which any manipulation on the surrogate variable will yield an effect that has a relevantly similar statistical significance to the effect (of the same manipulation) on the true clinical endpoint (Prentice 1989). Buyse et al. (2000) suggest that an adequate surrogate must be "tightly correlated" with the true clinical endpoint. In addition to this, any treatment effect on the surrogate must be tightly correlated with the treatment effect on the true clinical endpoint (Buyse et al. 2000). A simple example of this is that elevated concentrations of LDL cholesterol (LDL-C) have been associated with cardiovascular risk (Gofman and Lindgren, 1950; Gofman et al., 1950). With therapeutic interventions—e.g., the use of agents which inhibit the rate-limiting enzyme of the cholesterol biosynthesis pathway (statins)—LDL levels were lowered, which induces a reduction in cardiovascular disease (Downs et al. 1998; LaRosa et al. 2005; Ridker et al. 2008).

However, it has been argued that correlation, even if extensive, is not sufficient for a given metric to be considered a SM (Fleming &DeMets 1996). In order for a surrogate marker to be validated, the correlation must *hold under distinct (causal) interventions*. That is, a given SM and its corresponding endpoint would require causal robustness. For example, a marker specific for kidney failure such as cystatin C would need to be elevated under distinct causes of kidney damage—such as, poison-induced, ischemic, mechanical, genetic, and inflammatory—be they direct or indirect.

By the recommendation of the International Conference on Harmonisation Guideline E9 *Statistical Principles for Clinical Trials,* a surrogate endpoint (and, by extension, an ideal clinical biomarker) should have biological plausibility, prognostic value for the disease outcome, and there should be a strong association between changes in the surrogate and changes in the outcome with therapeutic intervention (Cleophas et al. 2007). Ideally, the change in the surrogate will correspond to changes in the target more

[11] A process threshold is analogous to a point of no return or a penultimate risk factor that relates to some physical/structural/tissue change in the underlying workings of a system. For example, glomerulosclerosis precedes the endpoint of end-stage renal disease/kidney failure.

or less linearly[12]—or according to a very well-characterized curve—e.g. 1-to-1, monotonic, or single-concavity—across their entire quantifiable range, and without respect to the trajectory of change (i.e. no observed hysteresis), and in both natural and interventional contexts. That is, the surrogate must act as *both epidemiological marker and therapeutic responder* (Cohn 2004).

Lassere et al. (2007) and Lassere (2008) summarize three conditions for validated SMs:

(1) serve as harbingers of an irreversible clinical burden of disease, major morbidity or mortality,

(2) have been repeatedly validated in randomized controlled trials using a diversity of interventional targets, and

(3) consistently meet the highest standards of predictive association within and between studies.

For example, despite the fact that low HDL-C cholesterol may be causally relevant to coronary disease, due to the fact that HDL-raising interventions have not shown effectiveness for patient outcomes, HDL is not a classified as a validated surrogate whereas LDL has sufficiently satisfied these criteria (IOM 2010, Ch. 4).[13]

3.2 Biomarkers in Practice: Problems with Uncertainty and Divergence.

While idealized descriptions—e.g., Prentice (1989) and Lassere et al. (2007; 2008)—aim at establishing high standards for validated SMs, in practice, SMs never reach 100% sensitivity or specificity. For example, in a meta-analysis on Alzheimer's disease, dementia was evaluated using the cerebrospinal fluid peptide $A\beta_{42}$ as a biomarker (Mo 2015). In this meta-analysis, 436 out of 1349 participants developed Alzheimer's dementia. According to Ritchie et al. (2014), "Individual study estimates of sensitivity were between 36% and 100% while the specificities were between 29% and 91%" (7). Not only do sensitivity and specificity range between studies for a given surrogate marker, but sometimes it is even difficult to make a proper estimate

[12] Woodward's (2003) account of causation is relevant here. An interesting discussion would be to apply Woodward's operation of manipulating one variable (surrogate marker) to see changes in another (the target variable).

[13] If a surrogate is "reasonably likely" to forecast an outcome, but such a tether is not fully conclusive based on the evidence, a surrogate may be considered *unvalidated* and used for accelerated approval of drugs and medical devices in pressing clinical situations with few alternatives. In accordance with FDA regulations (CFR Title 21 Subpart H), these unvalidated SMs must be subsequently validated (Katz, 2004). Unvalidated surrogates are also used in pre-clinical or pilot trials exploring safety or reasonable likelihood alone. As the spectrum of disease far outstrips our toolkit of validated surrogates, most disease-centered biomedical literature utilizes unvalidated surrogates as sources of evidence. It is important to note that the FDA only lists four validated surrogate markers: systolic blood pressure (SBD), low density lipoprotein cholesterol (LDL) level, forced expiratory volume in 1 second (FEV1), and human immunodeficiency virus (HIV) viral load (http://www.fda.gov/AboutFDA/Innovation/ucm512503.htm).

of sensitivity and specificity, because we do not have information about the true endpoint. That is, the diagnosis of the actual health status of the study participants is itself based on *uncertain measures*. Cure (2014) uses autopsy in order to tighten the range of uncertainty for Alzheimer's measures. While effective for generating specificity and sensitivity, such outcome measures are rarely accessible.

Overall, there is disagreement about cut-offs for sensitivity and specificity in CBs in general. Some researchers continue to posit high standards—e.g., 90% (Brower 2011). Others argue that sensitivity and specificity alone only work for discrete, not continuous variables (Cleophas 2007). Pepe et al. (2016) posits that the thresholds we accept for CBs should depend on the prevalence of the disease endpoint and also on the costs of making a mistake versus the benefits of correct identification. For example, there are unvalidated surrogates, "reasonably likely to predict clinical benefit (for a specific disease setting and class of interventions)"(Fleming and Powers 2012). There are also "correlates that serve as measures of biological activity", such as CD-4, in HIV infected patients and fever in Community Acquired Bacterial Pneumonia (Fleming & Powers 2012, our emphasis).

Divergence between values of validated markers that are presumed to be systematically connected (and assumed to 'line up' in a causal structure), may also occur. When judging the difference in quantitative values between two biomarkers, whose values are expected to be aligned, scientists may refer to theory to embed both biomarker trajectories along a series of interventions and realize that one of the biomarkers is not following its theoretically predicted trajectory. For instance, in the aforementioned example, LDL-C was categorized as a validated SM; but Otvos et al. (2011) show the *inconsistency* of LDL measurement in relation to the clinical endpoint by comparing LDL-C and LDL particle number (LDL-P) measures.

It is worth noting that pragmatic considerations are also important in judgments about converging or diverging biomarkers. Scientists may judge that a given biomarker is "similar enough" or "differs too much" based on considerations that are outside of the theoretical and methodological parameters.

The problem of SM divergence is not directly solvable by referring to already available theoretical or explanatory considerations. Even in the category of validated SMs, causal complexity prevents clean-cut relationships between SMs and endpoints. This is where causal-explanatory reliability, i.e. extracting causal information from surrogates, becomes difficult. As reviewed in Fleming & Powers (2012), a marker may:

> (C1) be a parallel output of an upstream event in the causal pathway, but not tie into the downstream causal pathway of the outcome of interest;
>
> (C2) lie only within one of many parallel causal pathways in multifactorial disease processes;

(C3) be the target of an intervention that has direct or indirect off-target effects on the outcome of interest.[14] (See Figure 1)

In selecting targets, we must be cautious before concluding that the marker is located squarely within the causal pathway for our disease of interest. Even with perfect correlation and proximity, a marker may be part of a parallel process, an inert byproduct, an essential compensation mechanism, or an inessential compensation mechanism; and we can imagine that directly targeting the marker in this case might actually make for worsened outcomes. Hence, (C1)-(C3) reduce the validity of the surrogate marker.

When multiple surrogate markers are used, the methodological situation worsens.[15] Suppose that two surrogate markers diverge in results. Each marker may diverge *for any of the specified reasons (C1)-(C3)*. But in each case we do not know what is *producing the divergence*. Take the example of LDL as a surrogate for cardiovascular disease:

Serum lipoproteins are particles that transport water-insoluble lipids such as triglycerides, cholesterol and cholesterol esters between the organs of energy transformation (liver, kidneys) and the organs of energy expenditure and storage (musculature, adipose). Low-density lipoprotein cholesterol (LDL-C), a measure of the total amount cholesterol contained within all LDL particles, is an accepted surrogate marker for cardiovascular disease risk, despite the fact that about half of hospitalized cardiovascular disease occurring in individuals with normal or "optimal" LDL-C (<100 mg/dL) (Sachdeva et al. 2009). Unsurprisingly, since the carriage capacity of lipoprotein particles is bound by lower and upper limits, the number of LDL particles (LDL-P) correlates strongly with LDL-C and has emerged as a strong predictor of risk.

In a large cohort, LDL-C and LDL-P values *differed* by more than 12 percentile units in half of the participants, with one-quarter of them confined to high LDL-C/low LDL-P and another quarter to low LDL-C/high LDL-P subgroups. Moreover, there seemed to be a difference in predictive value. The low LDL-C/high LDL-P subjects had: a 67% greater risk of a cardiovascular event than the high LDL-C/low LDL-P group; a 25% greater risk than the non-discordant group; and they were found to have a greater prevalence of the metabolic syndrome (MetS) and its associated biomarkers such as elevated triglycerides. Furthermore, in those with LDL-P below the 30th percentile, low LDL-C (also below the 30th percentile) increased cardiovascular event rates by 30%, while in those with LDL-P above the 30th percentile, low LDL-C increased event rate to a similar extent (Otvos et al. 2011).

[14] That is, the intervention may produce other causal interactions that are relevant to the outcome of interest. Additionally, combinations of (C1), (C2), and (C3) are probable in biological systems.
[15] There is disagreement about combining bimarkers to make useful predictions. In Alzheimer's research, Lehman et al. (2014) say that combining adequate markers (e.g., 80% specificity and sensitivity) improves their utility. Palmqvist et al. (2015) argue that combining markers does not improve their predictive utility—although he does not directly address higher ranges of specificity and sensitivity.

In such cases where SMs diverge in results, the causal conclusions are puzzling. The primary methodological instinct is to weed out the SM with low sensitivity and specificity.[16] But in each case we do not know what is producing the divergence. We offer a methodological suggestion: *to use multiple surrogate markers in order to make causal conclusions at the theoretical level*. The solution that we will focus on for the remainder of the discussion is how to methodologically account for multiple surrogate markers with converging and diverging results.

4. Robustness Analysis and Divergence

In this section, we present an intervention-based methodology for assessing the reliability of surrogate markers. We do this by discussing a new method for individuating modes in Section 4.1 and then discussing the importance of converging and diverging results in biomarker readings in Section 4.2. Our intervention-based methodology will refer to two eliminative robustness analyses that use diverging results to make conclusions (Schupbach 2015 and Keyser 2016), both of which focus on elimination and indicate possible inferential uses of diverging results.

4.1 Individuating Modes; (partially) independent pathways vs. physical principles

To use Woodward's (2006) reasoning, it is unlikely that different physical processes would share the same source for error: it is unlikely that different types of biomarkers share the same error source.[17] Standard accounts of robustness analysis rely on distinct physical principles or distinct physical processes underpinning the independence of modes of observation.[18] There are difficulties in specifying what counts as a 'different physical process' (Keyser 2016). Applied to biomarkers, there are many questions about differences in process: Do small adjustments, e.g., in only some of the conditions, constitute changes in physical process? What kinds of differences are important between biomarkers: the source, composition, and/or use? If only some of these conditions differ in two given biomarkers, are the biomarkers relevantly similar?

Keyser (2016) presents a heuristic-based account for differentiating physical modes: *the manipulation indicator*. One way to differentiate biomarkers is to look for indicators of physical independence:

---

[16] As discussed, because sensitivity and specificity carry uncertainty, it would not be a simple case of using the "highest" scoring surrogate marker.

[17] Here, "error" does not refer to measurement error. In other words, we assume that the measurements reflect the actual value of the biomarker. In the case of biomarkers, "error" refers to the interference or confounding of unspecified biological variables in a physiological network.

[18] See Hacking (1983), Barad (2007), Stegenga (2009; 2012), and Keyser (2016), Woodward (2006).

> In the presence of the manipulation indicator, one measurement process can manipulate the results of other processes, and independence is compromised. For example, suppose that in (robustness analysis applied to the measurement process) you have two modes of gathering information through measurement, M1 and M2. To see if M1 and M2 are individuated modes, we can introduce a manipulation to the measurement conditions (e.g., instrumental, preparatory, etc.) behind M1 to see if it produces a change in M2. If it does, then there is successful manipulation, and we conclude that M1 and M2 are not independent. (2016, 17)

This account can be modified to be useful for individuating biomarkers. In the case of Keyser's (2016) measurement individuation, *independence is demonstrated through a manipulation indicator in upstream measurement processes*. It would not be sufficient to demonstrate independence by altering a downstream determinant just prior to or at the point of measurement output.

In using biomarkers as individuation indicators there are some difficulties. The first difficulty is that because we are using biology to probe biology, we are dealing with interconnected systems. This means that full independence of modes may be out of the question. This is where pragmatic considerations become important: modes are independent enough for the diagnostic purpose. Our account of the individuation/independence of biomarkers has a pragmatic component. However, whether partial or full, independence will also be determined using causal models that characterize a given intervention. So, we propose that partial independence is *determined* by using causal models—that characterize a given intervention, including the variables and values that constitute the model—in addition to pragmatic considerations about the purposes of a given intervention.[20]

Demonstrating (partial) independence in biomarkers involves the use of exogenous interventions that do not interact with one of the biomarkers of interest, but rather modify a process some number of "steps" (factors, voxels) removed from each of the markers.[21]

Here, pharmacological intervention is proven to be a fertile device for demonstrating independence in biomarkers. In this case, all we need to show is that, for any two biomarkers, the equation describing the curve of their relationship under natural conditions, is non-identical to the equation describing the curve of their relationship under interventional conditions that are indirect. If this is demonstrated, we have (partial) independence. In simple terms, the intervention does not affect each biomarker in the same way. The following two examples are presented for illustrative purposes:

---

[20] We thank an anonymous reviewer for the suggestion that 'partial independence' involves causal models.

[21] Absent of this, we run into the circularity that two metrics are partially-independent because we observe discordance and we can glean causal information from discordance among said markers because they are partially-independent.

biomarkers of inflammation (CRP and ESR), and biomarkers of cognitive impairment (PET Amyloid and CSF Aβ).

CRP and ESR: The acute-phase response is a systemic defense program initiated by the liver in response to rising levels of interleukin-6 (IL-6), an innate inflammatory messenger produced in response to tissue damage, infectious agents, environmental toxicants and other salient abnormalities in cellular function. During the acute-phase response, levels of several serum proteins change to assist the organism in meeting those needs, which commonly prove most adaptive under such challenges. Therefore, hallmarks of the acute phase response are commonly used as indicators of all-cause inflammation. Two such markers are C-reactive protein (CRP), a protein that approximates the function of a non-specific antibody (Marnell et al. 2005), and the erythrocyte sedimentation rate (ESR) that is essentially a measure of blood viscosity resulting from the abundance of other acute-phase proteins such as fibrinog en and α-globulins.

Despite both markers traditionally being considered interchangeable, given that they derive from the same organ under presumptively identical conditions, CRP and ESR may be discordant/divergent in 10-30% of cases, with singly-elevated ESR being 6-fold more common (Costenbader et al. 2007; Colombet et al. 2010; Feldman et al. 2013; Sbong& Feldman 2015). The observation of discordance in simultaneous measurement of these markers is important in exploring their partial independence.

By introducing a pharmacological intervention we can see how each biomarker is affected. Prednisone is a synthetic glucocorticoid that reduces inflammation by strong activation of the glucocorticoid receptor in various tissues throughout the body. Over 140 days of prednisone administration, CRP will decline monotonically, while ESR will show a backwards-J-shaped non-monotonic decrease over the same time period (McConkey et at 1979). This illustrates that the two biomarkers respond differently to intervention and are therefore partially-independent.

PET Amyloid and CSF Aβ: The drug Semagacestat is an inhibitor of the integral membrane protease complex known as γ-secretase, which is the terminal processing step in the production of the Aβ peptides and other essential protein cleavage products within neurons (and other tissues). As an intervention, Semagacestat is therefore indirect for the biomarkers under consideration.[22] Semagacestat dose-dependently reduces CSF Aβ production in humans (Imbimbo&Giardina 2011). Furthermore, similar compounds, administered to humanized transgenic mice, have been shown to prevent new plaque

[22] Due to a significant worsening of cognitive scores and the emergence of several alarming off-target effects in the semagacestat groups, the full panel biomarker assessments were not completed in humans prior to termination of clinical trials.

formation with no change in existing plaques as visualized with amyloid plaque labels (Garcia-Alloza 2009). Therefore one may conclude that in the presence of these agents, PET-measured amyloid and CSF Aβ change with independent dynamics in contrast to the natural history of Alzheimer's disease.

An intervention-based account of individuation/independence can be useful for understanding robustness analysis. *Once biomarkers are shown to be partially-independent we can begin to ask questions about what we can conclude from converging and diverging results in the presence of intervention*. In the next section we discuss how robustness analysis can be used in biomarkers in order to make conclusions about causal relations.

4.2 Robustness and Elimination: Specifying Causal Relations.

Traditionally, robustness analysis with respect to the modeling process provides a way to illuminate the relationship between the robust consequence and certain aspects of each model in order to conclude that the inessentials of the model (the idealizations) are not producing the robust consequence. Framing robustness analysis this way, we focus on eliminating the inessentials of the model.[23] Two philosophical accounts that focus on elimination in robustness analysis are useful: Schupbach (2015) and Keyser (2016). The former focuses on using robustness analysis to eliminate competing hypotheses. The latter focuses on tracking causal relations based on multimodal comparison.

Schupbach (2015) uses Horwich's (1982) eliminative account of evidential robustness, the emphasis of which is using evidence to eliminate competing hypotheses in relation to some hypothesis (H) (2015, 314). According to Schupbach:

> What really matters for (Robustness Analysis (RA))-diversity is that the means (which may actually be quite similar in most respects) are different in just the sense required to rule out H's salient competitors. Accordingly, when seeking to increase the RA-diversity of our evidence, we search for a new way of detecting (the robust result) R that rules out some of H's still-standing competitors. (2015, 316)

According to Schupbach, new modes of evidence need not be fully independent. Rather, they just have to be "relevantly different" for the purpose of eliminating competing hypotheses (2015, 316). This fits nicely with our use of 'partial independence' in biomarkers.[24] Finding modes that are different enough is important for biomarker

[23] Orzack and Sober's (1993) criticism still looms. Perhaps it is some common core, shared by the individual modes, that is driving the robust result. See Justus (2012) for a summary of the concern: robustness analysis may only reflect shared properties of models rather than anything about the real world system (798).

[24] To illustrate his point, Schupbach uses the example of Perrin's modes of measuring Brownian motion. While Perrin's use of varieties of pollen are not "strongly heterogeneous" because each

analysis. In Alzheimer's research, PET amyloid & CSF Aβ are biomarkers that are individuated enough within the context of intervention. We can slightly change Schupbach's account, so that robustness offers hypothesis *revision*, not just elimination. As we will see in the Alzheimer's research application in Section 5, revision gets at an important process in robustness analysis: the theoretical specification of causal details.

To make the eliminativist picture even more fine-grained, we can focus on the use of multiple modes to figure out what parts of each mode are producing causally relevant results. This is where we can use Woodward's (2006) specification of causal robustness. Causal robustness focuses on invariant relationships over manipulations. Manipulation is relevant to the case of biomarkers that undergo interventions because some relationships remain invariant while others produce diverging results. Trends that survive modified conditions are robust and, to use Woodward's (2006) reasoning, indicate causal structure. We can specify the conditions under which the robust result remains invariant—e.g., some result R occurs under admissible changes in a specified range of conditions C. However, as discussed in Section 3, many biomarkers do not maintain converging results, and results from multiple biomarkers often diverge. So we need a robustness account that provides causal information while effectively interpreting diverging as well as converging results.

Keyser (2016) offers an account of robustness analysis in measurement that consists of the comparison between independent modes of measurement along with the use of theory to make sense of that comparison. He illustrates that theory in the presence of diverging measurement results can pinpoint *what is producing the divergence*. For example, when multiple thermometers converge but others diverge, theory is used to analyze the conditions behind the divergence (Keyser 2016, 10). Keyser proposes that after initial cross-comparison of measurement methods and results, we use theory to home in on specific physical differences in thermometers—e.g., the liquid used in a given thermometer—in order to *explain* how those features produce differences in results (2016, 10-11).[25] This differential comparison and explanation process in the presence of theory is useful for measurement in terms of specifying causal relations within the measurement set-ups, not only in terms of which conditions are *difference-makers* but also in terms of how error is "tracked"—i.e., located (Keyser 2016, 11). This is precisely why this account is useful for biomarkers: Cross-comparison allows for *locating relevant causal relations in biomarker measures*. We can refine and structure Keyser's (2016)

---

experiment uses a type of pollen; the experiments with varieties of pollen are different enough in order to rule out potential confounding hypotheses—such as, Brownian motion is only due to a specific type of pollen (2015, 316).

[25] Keyser (2016) draws on van Fraassen's (2008) discussion of the relation between theory and measurement practice: Theory classifies what is being measured. However, as Keyser points out, this does not have to be a fully developed theory. It can even amount to a theory of how the instruments work. This may be applicable to cases of biomarker measurement where there is no overarching theory.

account in order to provide useful steps for our proposed intervention-based robustness methodology.

The two methods of eliminative robustness can be used in a complementary fashion. *One focuses on eliminating alternative hypotheses (or revising hypotheses). The other can be used to further specify relations between variables within the context of theory*. We can generate three conditions for robustness analysis in biomarkers, which draw on these two methods. These conditions are useful for producing causal information.

R1) *Partial independence of modes*: Demonstrating partial physical independence in biomarkers involves an exogenous intervention that does not directly interact with one of the biomarkers of interest, but rather modifies a process some number of "steps" removed from each of the markers. Partially-independent biomarkers can serve as crosschecks on each other's results.

R2) *Comparison of diverging and converging results across biomarkers*: Partially-independent biomarkers (along with the relevant external causal conditions) are systematically juxtaposed to see: A) where they converge and diverge in results; and B) what conditions are responsible for the convergence and divergence. By systematically comparing partially-independent biomarkers we can track under what conditions they fail to converge in results, and under which conditions they successfully converge. Here, in addition to focusing on the robust result (the result that is invariant over biomarker measures), we *also* focus on the contextual failure of convergence in the cross-comparison of biomarkers (divergence). From both converging and diverging results in cross-comparison we get causal information about how the biomarkers work.

R3) Through (R1) and (R2), theoretical information is generated within the context of theory. By applying theory, cross-comparison provides information about causal relations. That is, through a systematic cross-comparison of the markers we can discover under what conditions certain changes are being produced, including confounding results. This kind of causal specification is useful for theory revision. That is, based on the juxtaposition of results, theories are eliminated and/or revised in order to provide adequate representations of the system being studied.

We apply (R1)-(R3) to analyze how biomarker modes can be useful for making causal conclusions in Alzheimer's research.

5. Robustness Analysis Applied to Alzheimer's Research.

Before we apply our intervention-based robustness analysis, in Section 5.1 we briefly introduce theory and measurement in Alzheimer's disease research with particular attention to the revision of theory in the context of multiple biomarkers. In Section 5.2, we apply our methodology to Alzheimer's research in order to show its adequacy for analyzing currently developing empirical research. Following our robustness methodology conditions (R1)-(R3), we show that when systematically combining different interventions on multiple biomarkers, promising causal information emerges.

5.1 Alzheimer's and Theoretical Revision

The theoretical and biomarker details of Alzheimer's research are important for our applied account of robustness analysis. Through the details, we show *how* theoretical models are *refined in relation to the converging and diverging results in biomarker measures.* A way to usefully apply robustness analysis to Alzheimer's research is to look at how biomarker robustness leads to specifications in theoretical robustness.

Traditionally, Alzheimer's disease was presumptively diagnosed based on clinical observation, such as cognitive testing battery biomarkers like the Mini-Mental State Exam (MMSE), and it was confirmed via autopsy by presence of the neuropathological signs of disease—e.g., amyloid plaques, neurofibrillary tangles, synaptic decay and neuronal loss. Recently, the development of bio-analytics and advanced neuroimaging has allowed these hallmarks of disease to be tracked and visualized in living subjects. Massive initiatives have helped to establish the current use of clinical biomarkers, which themselves have given rise to a multi-decade staging of the disease process (Jack & Holtzman 2013; Young et al. 2014). Notable are the refinements to the prevailing theoretical disease model, the *amyloid cascade hypothesis* (ACH) as a consequence of multiple biomarker measures.

Currently, the biomarkers in greatest use are cerebrospinal fluid (CSF) tau protein (the principal component of neurofibrillary tangles), CSF 42-amino acid amyloid-β ($A\beta_{42}$; hereafter CSF Aβ). regional brain volume as determined by structural MRI, neuronal metabolism measured by glucose analog uptake and positron emission tomography (FDG-PET), and amyloid plaques measured with PET detection of amyloid-binding dyes. CSF Aβ is the protein cleavage product believed to precipitate disease by forming neuron-damaging plaques as it aggregates, and "42 amino acids" refers to the length of this large peptide, measured by the number of basic amino acid building blocks it possesses.[26]

Based on the ACH, it was posited that CSF Aβ would be elevated in Alzheimer's dementia. Thus the discovery that levels were significantly reduced in Alzheimer's was

[26] Additional imaging techniques are being adapted and blood plasma measurements of Aβ are being developed.

initially met with surprise (Hampel et al. 2008). Hardy &Allsop (1991) argued that an imbalance between amyloid production and clearance (potentially caused by a number of factors) led to an accumulation of Aβ which eventually precipitated into plaques, damaged neurons and sequentially led to the downstream pathology (tangles of modified tau, neuronal stress, damage, and ultimately loss) of AD.

Youngkin (1995) advanced that a 42-amino acid (rather than a 40-amino acid) Aβ peptide was likely the culprit by. This peptide was characterized as being "stickier" in the sense of more likely to aggregate, and it was correlated better with the production/elimination imbalance, especially in genetic (early-onset) forms of AD.

Over time many potentially toxic mechanisms of Aβ were characterized, such as triggering inflammation and excitotoxicity (Behl 1997; Canevari et al. 2004; Carillo-Mora et al. 2014)—with many, now hypothesized to result from repeating units of adhered misfolded Aβ peptides referred to as Aβ oligomers. Recently, the discussion of oligomeric forms of Aβ 42 has expanded.[27] Very soon, a gestalt shift in theoretical explanation allowed for a crucial discovery: amyloid plaques within the brain were acting as sinks for the circulating CSF Aβ and, beyond a critical concentration, Aβ would precipitate into these plaques, or, at the very least, newly produced Aβ would be sequestered in plaques to diminish steady-state levels of CSF Aβ. If the brain were imaged, one should observe the retention of amyloid-stains as CSF Aβ dropped.

This specification of the original amyloid cascade hypothesis (call it ACH*) relied on different assumptions about the nature of Aβ and amyloid plaques. In ACH* the assumption is not that soluble monomers (single units) are elevated throughout the time course of pathology; also they are not a primary agent responsible for damage.[28] Even though the importance of Aβ is invariant over both theoretical models, ACH* provided most of the useful explanation. But it too required empirical confirmation; and this was done in relation to multiple *diverging biomarker results*.

ACH* predicted that depressed CSF Aβ and an elevated amyloid plaque PET signal would be fully correlated over their numerical range, in that they were assumed to be dependent biomarkers. That is, PET signals should increase as CSF Aβ levels decrease

---

[27] Currently, these are thought to be the most toxic form, but lesser or interacting toxicity of other forms of Aβ is not discounted. The current discussion revolves around the many forms and sizes of oligomers: some toxic, some not, prefibrillar forms, and fibrillary (fiber-like) forms which can form either diffuse or dense plaques. The general order of formation is asserted to be: peptides, small oligomers, larger oligomers, prefibrillar forms, fibrillary forms, diffuse plaques, dense neuritic plaques. But this formation order is in no way invariant, as there are branches, two-way streets and overlaps. It is important to note that much of this information on ACH and oligomers was hypothesized before the biomarkers were characterized. It can be argued that such theoretical explanations were not fully integrated into the theoretical model of ACH until recently with the help of biomarkers. The most current model of the theory is discussed by Selkoe (2016).

[28] The second reason had been suspected **(**Reviewed in Walsh &Selkoe 2007; Review that monomers may actually be protective: Guifridda et al. 2009).

and vice-versa. Indeed, they were shown to correlate on the order of 85% using threshold values (Landau et al. 2013; Zwan et al. 2014; Vos et al. 2016), but even a *15% discordance appeared to indicate missing causal specifications*. So another theoretical model is now needed. In fact, biomarker discordance, especially in the preclinical and early stages of Alzheimer's coupled with the lack of success of targeted pharmacologic agents has led many to question the accuracy of the ACH/ACH*.

In normal subjects and those with mild cognitive impairment (MCI), PET positivity and CSF Aβ "negativity" (that is, higher values) is the most common form of discordance (Landau et al. 2013).[29] This is despite the fact that early on it was observed that, in cases of discordance, CSF Aβ decline would *precede* PET positivity (Musiek&Holzman 2012; Zwan et al. 2014).

Such discordant/diverging results supported the current development of a new sub-model according to which:

(1) either Aβ is incorporated into soluble oligomers (Lesne 2015), pre-fibrillary forms and/or diffuse fibrillary precipitates *prior to* the appearance of dye-retaining plaques—an issue of laddering (oligomer submodel, OSM, Walsh et al. 2000),

(2) or potential blood-brain barrier dysfunctions increase the dilution of CSF Aβ into peripheral compartments (Erickson & Banks 2013).

The OSM produces useful explanations with new causal components neither included in ACH nor in ACH*. On the basis of OSM, cognitive impairment can precede amyloid biomarker positivity despite later progression to Alzheimer's dementia.[30] Oligomers seem to form to some degree prior to plaques, so we could have high CSF Aβ without much plaquing.[31]

Juxtaposing OSM with ACH and ACH* shows some robustness. Each model—ACH, ACH*, and OSM's agree that Aβ is an important factor in Alzheimer's disease, but they do not posit the same causal relations.

However, the robust explanation, which has to do with the causal role of Aβ in Alzheimer's, doesn't result from using different models to explain a result. Rather, these

---

[29] Such discordance has been observed in 21% of normal individuals, 12% of MCI cases and in 6% of cases with diagnosed Alzheimer's dementia (Mattsson et al. 2015). It is worth noting that the oligomer sub-model, discussed later in the paragraph, can account for both types of discordance mentioned in this study: the larger discordance of the florbetapir^{+} (PET positivity) /CSF Aβ(-) group and the smaller discordance of the florbetapir^{-} (PET negativity)/CSF Aβ(+) group.

[30] This has indeed been demonstrated by Toledo et al. 2014 and Ritchie et al. 2016.

[31] Additionally, OSM can account for the fact that an individual may have cognitive decline even with low CSF Aβ and no plaques because processing enzymes that produce Aβ are also necessary for the production of neurotrophic and neurodifferentiation factors (Willem et al. 2006; Woo et al. 2009; De Strooper et al. 2010). Thus, low activity might lower both CSF Aβ and factors important for optimal neuronal function. This may be seen in neuroinflammation or CNS infection as well (Krut et al. 2013).

models are *refined in relation to the converging and diverging results in biomarker measures*.

One could argue that there are two types of robustness occurring here: the robustness of biomarker measures (biomarker robustness) and the robustness of theoretical models (theoretical robustness). A way to usefully apply robustness analysis to Alzheimer's research is to look at how biomarker robustness leads to specifications in theoretical robustness. As we have just seen, in Alzheimer's research, diverging biomarker readings catalyzed the specification of new theoretical models. Discrepancies in multiple biomarkers in various studies facilitated the move from ACH to ACH* and then to OSM. This can be analyzed as a case where *multiple modes (biomarkers) support theoretical revision*. As mentioned in Section 4 and illustrated in this section, this is precisely why it is important to change Schupbach's account so that robustness offers hypothesis revision, not just elimination. Relevant details of ACH*, for example, were eliminated for OSM, but some previous aspects did not change. As we will see next, revision is an important process in robustness analysis because it is behind the theoretical specification of causal details.

5.2 Intervention-Based Robustness Analysis Applied to Alzheimer's Measurement

Under the amyloid cascade models discussed above (ACH, ACH*, and OSM), reduction of cerebral Aβ is expected to reduce cognitive decline and may afford neurons a chance to repair, either by delaying debilitating dementia or restoring cognitive function. It is predicted that a 25% reduction in Aβ species may prove to be an adequate therapeutic target for delayed disease progression to be observed (Toyn 2015).

Human trials of these agents have heretofore been disappointing largely because agents proved to be poorly tolerated at biochemically active doses. At those same doses disagreement among clinical biomarker findings as well as between biomarkers and established functional endpoints were observed (Toyn 2015). Altogether, this has contributed to several calls for a rejection of the ACH, or at the very least, its role as the cardinal focal point of disease research. Though difficult to interpret resultant to biomarker discordance, fragmentary positive findings have shifted focus toward more upstream interventions as desirable targets in antibody trials (Toyn 2015).

Presently, promising research has emerged: The PRIME trial (Sevigny et al. 2016) is a Phase 1b clinical trial (a dose-escalation trial in a limited number of patients) conducted across 33 locations within the United States testing the efficacy of Aducanumab. Aducanumab specifically targets Aβ fibrils as well a toxic Aβ oligomers including decamers (arc-, disc- or pore-shaped composites consisting of 10 regularly repeated Aβ monomers adhered together) (Ratner 2015). Binding studies reveal that Aducanumab does, in fact, bind fibrils in both diffuse deposits and compact plaques and that chimeric mouse-adapted versions of the drug precipitate Aβ oligomers in vitro by

clumping them together into large complexes.[32] Aducanumab binds CAA deposits (see above) with less prominence and has 10,000-fold target selectivity for these composite forms over Aβ monomers.

Subjects were divided into five arms, to receive monthly IV infusions of either placebo or 1, 3, 6 or 10 mg/kg bodyweight aducanumab over a one-year period. After 12 months, the 10 mg/kg group had PET dye retention reduced 25-30%, almost to the cutoff level for admission into the trial. Dose- and time-dependent reductions in PET positivity were seen across the intervention arms.

Despite being neither designed nor powered to assess efficacy with respect to cognitive outcomes, there was also a dose- and duration-dependent slowing of cognitive decline (but not reversal or improvement) in the Clinical Dementia Rating-Sum of Boxes (CDR-SB) and Mini-Mental State Examination (MMSE) scores. The 10 mg/kg group retained significantly more cognitive function than the placebo group and had effectively unchanged cognitive performance over the 12 months. Placebo-exposed subjects had no more PET-detected amyloid accumulation over the yearlong trial but continued to experience cognitive decline, pointing to one or more unidentified variables, presumably Aβ oligomers. In the 10 mg/kg group, despite continued reductions in PET amyloid positivity from 6 to 12 months, CDR-SB scores did not continue to improve. This indicates that changes in plaque abundance and cognitive function are divorced over their dynamic ranges.[33]

Following our robustness methodology conditions (R1)-(R3), when systematically combining different interventions on multiple biomarkers, promising causal information emerges. We discuss this by directly comparing bapineuzumab and solanezumab interventions, along with aspects of the aducanumab intervention.

Bapineuzumab is a humanized mouse monoclonal antibody that targets the N-terminal (front end) region of fibrillary (the principal configuration within plaques) and monomeric Aβ. In *APOE*-ε4 gene carriers and non-carriers, bapineuzumab reduced Aβ PET positivity and decreased the neurodegeneration biomarker CSF tau, but did not alter levels of CSF Aβ (Toyn 2015). Subjects exposed to active treatment initially showed better retention of MMSE scores in the highest dose group but failed in late-stage clinical trials to improve cognition or activities of daily living versus control (Ratner 2015).

In contrast to bapineuzumab, solanezumab, a humanized mouse monoclonal antibody targeting the central region of Aβ monomers and small oligomers, did not alter

[32] In simpler terms, scientists replaced the "tail", also known as the constant region of the antibody with that from a mouse so that the mouse's immune system wouldn't have an immune response to the human antibody. This shows that it is the specific antigen-binding region from the screen, which interacts with Aβ oligomers.

[33] While awaiting Phase III efficacy trials, which were begun immediately upon consolidation of positive findings, a lingering question (Lee et al, 2006) poses whether plaque destabilization could actually lead to increased exposure of neurons to toxic forms of amyloid as many structural models indicate that plaque-oligomer interconversion could be bidirectional.

levels of plaque assayed by Aβ PET, nor did it alter CSF tau. In addition, it led to an increase in plasma Aβ and a reduction in CSF Aβ (recall, this reduction is a marker for poor cognitive outcome in the *absence* of intervention) (Toyn 2015; Ratner 2015; Selkoe& Hardy 2016).[34]

Putting these results side by side (see Figure 2), we see that there is divergence in relation to multiple biomarkers:

1) levels of plaque assayed by Aβ PET;

2) CSF tau; and

3) CSF Aβ.

Bapineuzumab reduced (1) and (2), but did not alter levels of (3).

Solanezumab did not alter (1) and (2) but led to a reduction in (3).[35]

Methodologically, there are two important points here.

1. There is a contrasting picture between three biomarkers in relation to two interventions (R1 and R2). The divergence and convergence is systematic in that one intervention shows changes in two biomarkers and no change in the final one, and the other intervention shows stability in two biomarkers and a change in the final one. Furthermore, there is a difference in the outcome of cognitive improvement. So, the two different interventions not only produce differences in biomarker measures, but they also produce differences in cognitive results.
2. Within the context of theory, specifically OSM, this systematic divergence and convergence leads to useful causal conclusions (R3). Using robustness analysis, we specify multiple *promising conclusions*, which can be explored in further research:

Conclusion 1) Oligomers are a more promising causal target for preserving cognitive function (C1):

Following the antibody trials, solanezumab as well as aducanumab target oligomers as part of their specificity. Solaneuzumab results suggest improvement in cognition in mild cases based on the post-hoc pooling analysis. Aducanumab seems to improve cognition based on the Phase 1b results relative to control/placebo. Solaneuzumab and aducanumab target oligomers, but without a contrasting intervention

[34] Solanezumab is currently undergoing pooled subgroup re-evaluation from two Phase III trials after small but significant positive findings (34% deceleration of cognitive decline over 18 months versus placebo) were observed in Alzheimer's Disease Assessment Scale (ADAS) cognitive domain scores in those with mild impairment (Toyn 2015; Ratner 2015;Selkoe& Hardy 2016).Results of a follow-up Phase III study are in the offing (http://www.alzforum.org/therapeutics).

[35] As well as an increase in a factor not measured in the bapineuzumab study: plasma Aβ.

that does not target oligomers, we cannot see if oligomer intervention is *necessary* for cognition improvement. Bapineuzumab provides such a contrast. It does not target oligomers and shows no evidence for improving cognition. The *convergence between solaneuzumab and aducanumab compared to the divergence of bapineuzumab can be made sense of in a theoretical context*. In this case OSM makes sense of how the targeting occurs in the case of the different interventions in order to conclude which interventions are and are not targeting oligomers. For example, OSM explains that aducanumab targets Aβ fibrils as well a toxic Aβ oligomers including decamers; solaneuzumab targets the central region of Aβ monomers and small oligomers; and bapineuzumab targets the N-terminal region of fibrillary (the principal configuration within plaques) and monomeric Aβ. From this we understand that solaneuzumab and aducanumab target oligomers and bapineuzumab does not. The converging results of cognition improvement in the oligomer-targeting interventions compared to the diverging results of the non-oligomer-targeting intervention supports causal specification that oligomers modification is necessary for cognitive improvement. That is, without manipulation of oligomers, cognitive improvement will not occur.

Conclusion 2) CSF tau may not be a *complete marker* of neurodegeneration (C2):

It is assumed that cognition reflects neurodegeneration (Terry et al. 1991; Savva et al. 2009). In bapineuzumab, CSF tau declined (presumably less or at lower rate of neurodegeneration) but cognition did not change. In solanezumab, CSF tau did not change (no change in rate of neurodegeneration) but cognition improved relative to control.

Neurodegeneration and cognition are not correlated *in the way that we assume*. For example, there could be extra variables related to rate or duration of damage in upstream or downstream processes that are causally linked to cognition. This option would imply that neurodegeneration and cognitive decline do not change in lockstep due to variable complexity. If tau is a marker for neurodegeneration, and tau and cognition do not correlate, then neurodegeneration and cognition do not correlate. The lack of correlation between tau and cognition is supported by the shared lack of convergence for tau and cognitive markers in bapineuzumab and solanezumab interventions. However, we need to further explore the first conjunct in the antecedent, which is that tau is a maker for neurodegeneration.

Conclusion 3) PET positivity is not a surrogate for oligomers (C3).

Sevigny et al. (2016) refers to PET positivity as a "surrogate" of oligomers. After targeting oligomers, which cannot be measured directly in vivo, as well as plaques, Sevigny et al. (2016) found that PET-detected plaques decreased. This does not point to a clear relationship between PET positivity and oligomers because the latter were not directly measured. One theoretical option is that there is a transition of monomers to

oligomers to plaques that is unidirectional, and reducing plaques will pull oligomers into new plaques (thus reducing them).[36]

However, even without direct oligomer measurements, we can use robustness analysis to question Sevigny's theoretical model. If it is the case that oligomers are causally connected with cognition[37] and solanezumab actually targeted oligomers,[38] then PET positivity cannot be a surrogate of oligomers, as Sevigny claims. The reason for this is that, in the solanezumab intervention, while oligomers were targeted which produced a change in cognition, *PET positivity did not change*. In addition to this, the bapineuzumab intervention resulted in a *PET positivity decline* but cognition did not change. Because of the support from C1, which posits a close connection between oligomers and cognition, we can argue that no cognition change indicates that there is no oligomer change. *Both* interventions converge on PET positivity change not being correlated with cognition change, even though oligomer change is. This implies that PET positivity is not a surrogate for oligomers.[39]

6. Concluding Remarks

We have presented an intervention-based robustness methodology that specifies how to check for partial independence in biomarkers, how divergence as well as convergence is useful for figuring out causal information related to biomarkers and endpoints, and how theory revision works in relation to biomarker analysis. The aim of this discussion was for the purpose of philosophical methodological analysis, but the application of our account results in important consequences for empirical research. When applied to Alzheimer's research, our robustness methodology produces specific conclusions about *promising causal culprits as well as decoupled biomarkers and endpoints (C1)-(C3)*. By transforming robustness analysis into an eliminative account, we have made it useful for dealing with complex and discordant/diverging results in biomarker measurement.

References

[36] This has the corollary that plaques are being pulled out faster than oligomers can form new plaques, and monomers cannot form new oligomers as fast as oligomers go into plaques. This is similar to LeChatelier's principle. This could also be turned into a "bidirectionality" theoretical model.

[37] See C1 for support.

[38] See background on solanezumab earlier in the section for support.

[39] In a stronger form of robustness analysis, we can use the convergence of results to eliminate Sevigny's theoretical model. Alternatively, auxiliary modifications can be made positing a more complex causal relationship between PET positivity, oligomer change, and cognition. But given the analysis in C1 and C3 thus far, we can at least cast doubt on Sevigny's theoretical model.

ALZFORUM. n.d. Networking for a Cure. http://www.alzforum.org/therapeutics.

Aronson, J. K. 2005. Biomarkers and Surrogate Endpoints. *British Journal of Clinical Pharmacology* 59 (5): 491–94. doi:10.1111/j.1365-2125.2005.02435.x.

Barad, Karen. 2007. Meeting the Universe Halfway. In *Meeting the Universe Halfway*, 39–70. Duke University Press. doi:10.1215/9780822388128-002.

Behl, C. 1997. Amyloid β-Protein Toxicity and Oxidative Stress in Alzheimers Disease. *Cell and Tissue Research* 290 (3): 471–80. doi:10.1007/s004410050955.

Brower, Vicki. 2011. Biomarkers: Portents of Malignancy. *Nature* 471 (7339): S19–S20. doi:10.1038/471s19a.

Buyse, M., G. Molenberghs, T. Burzykowski, D. Renard, and H. Geys. 2000. The Validation of Surrogate Endpoints in Meta-Analyses of Randomized Experiments. *Biostatistics* 1 (1): 49–67. doi:10.1093/biostatistics/1.1.49.

Canevari, Laura, Andrey Y. Abramov, and Michael R. Duchen. 2004. Toxicity of Amyloid β Peptide: Tales of Calcium, Mitochondria, and Oxidative Stress. *Neurochemical Research* 29 (3): 637–50. doi:10.1023/b:nere.0000014834.06405.af.

Carrillo-Mora, Paul, Rogelio Luna, and Laura Colín-Barenque. 2014. Amyloid Beta: Multiple Mechanisms of Toxicity and Only Some Protective Effects? *Oxidative Medicine and Cellular Longevity* 2014: 1–15. doi:10.1155/2014/795375.

Cleophas, Ton, Aeilko Zwinderman, and Amel Chaib. 2007. Novel Procedures for Validating Surrogate Endpoints in Clinical Trials. *Current Clinical Pharmacology* 2 (2): 123–28. doi:10.2174/157488407780598126.

Cohn, J. N. 2004. Introduction to Surrogate Markers. *Circulation* 109 (25_suppl_1): IV–20–IV–21. doi:10.1161/01.cir.0000133441.05780.1d.

Colombet, Isabelle, Jacques Pouchot, Vladimir Kronz, Xavier Hanras, Loïc Capron, Pierre Durieux, and Benjamin Wyplosz. 2010. Agreement Between Erythrocyte Sedimentation Rate and c-Reactive Protein in Hospital Practice. *The American Journal of Medicine* 123 (9): 863.e7–863.e13. doi:10.1016/j.amjmed.2010.04.021.

Costenbader, KH, LB Chibnik, and PH Schur. 2007. Discordance Between Erythrocyte Sedimentation Rate and c-Reactive Protein Measurements: Clinical Significance. *Clinical and Experimental Rheumatology* 25 (5): 746–49.

Culp, Sylvia. 1994. Defending Robustness: The Bacterial Mesosome as a Test Case. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994 (1): 46–57. doi:10.1086/psaprocbienmeetp.1994.1.193010.

Cummings, Jeffrey L., Bruno Dubois, José L. Molinuevo, and Philip Scheltens. 2013. International Work Group Criteria for the Diagnosis of Alzheimer Disease. *Medical Clinics of North America* 97 (3): 363–68. doi:10.1016/j.mcna.2013.01.001.

Cure, Sandrine, Keith Abrams, Mark Belger, Michael Happich, and others. 2014. Systematic Literature Review and Meta-Analysis of Diagnostic Test Accuracy in Alzheimer's Disease and Other Dementia Using Autopsy as Standard of Truth. *Journal of Alzheimer's Disease* 42 (1): 169–82.

De Gruttola, Victor G, Pamela Clax, David L DeMets, Gregory J Downing, Susan S Ellenberg, Lawrence Friedman, Mitchell H Gail, Ross Prentice, Janet Wittes, and Scott L Zeger. 2001. Considerations in the Evaluation of Surrogate Endpoints in Clinical Trials. *Controlled Clinical Trials* 22 (5): 485–502. doi:10.1016/s0197-2456(01)00153-2.

De Strooper, Bart, Robert Vassar, and Todd Golde. 2010. The Secretases: Enzymes with Therapeutic Potential in Alzheimer Disease. *Nature Reviews Neurology* 6 (2): 99–107. doi:10.1038/nrneurol.2009.218.

Douglas, Heather. 2004. The Irreducible Complexity of Objectivity. *Synthese* 138 (3): 453–73. doi:10.1023/b:synt.0000016451.18182.91.

Downs, John R., Michael Clearfield, Stephen Weis, Edwin Whitney, Deborah R. Shapiro, Polly A. Beere, Alexandra Langendorfer, et al. 1998. Primary Prevention of Acute Coronary Events with Lovastatin in Men and Women with Average Cholesterol Levels. *JAMA* 279 (20): 1615. doi:10.1001/jama.279.20.1615.

Erickson, Michelle A, and William A Banks. 2013. Blood-Brain Barrier Dysfunction as a Cause and Consequence of Alzheimer's Disease. *Journal of Cerebral Blood Flow & Metabolism* 33 (10): 1500–1513. doi:10.1038/jcbfm.2013.135.

Feldman, Mark, Bilal Aziz, Gha Na Kang, Mildred A. Opondo, Randall K. Belz, and Connie Sellers. 2013. C-Reactive Protein and Erythrocyte Sedimentation Rate Discordance: Frequency and Causes in Adults. *Translational Research* 161 (1): 37–43. doi:10.1016/j.trsl.2012.07.006.

Fleming, Thomas R., and John H. Powers. 2012. Biomarkers and Surrogate Endpoints in Clinical Trials. *Statistics in Medicine* 31 (25): 2973–84. doi:10.1002/sim.5403.

Food and Drug Administration. 2017. FDA Facts: Biomarkers and Surrogate Endpoints. http://www.fda.gov/AboutFDA/Innovation/ucm512503.htm.

———. n.d. FDA Title 21 of US CFR. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=314&showFR=1&subpartNode=21:5.0.1.1.4.8.

Franklin, Allan. 1997. Calibration. *Perspecties on Science* 5: 31–80.

Garcia-Alloza, Monica, Meenakshi Subramanian, Diana Thyssen, Laura A Borrelli, Abdul Fauq, Pritam Das, Todd E Golde, Bradley T Hyman, and Brian J Bacskai. 2009. Existing Plaques and Neuritic Abnormalities in APP:PS1 Mice Are Not Affected by

Administration of the Gamma-Secretase Inhibitor LY-411575. *Molecular Neurodegeneration* 4 (1): 19. doi:10.1186/1750-1326-4-19.

Giuffrida, M. L., F. Caraci, B. Pignataro, S. Cataldo, P. De Bona, V. Bruno, G. Molinaro, et al. 2009. *β*-Amyloid Monomers Are Neuroprotective. *Journal of Neuroscience* 29 (34): 10582–7. doi:10.1523/jneurosci.1736-09.2009.

Glymour, Clark. 1980. *Theory and Evidence*. Princeton, NY: Princeton University Press.

Gofman, J. W., H. B. Jones, F. T. Lindgren, T. P. Lyon, H. A. Elliott, and B. Strisower. 1950. Blood Lipids and Human Atherosclerosis. *Circulation* 2 (2): 161–78. doi:10.1161/01.cir.2.2.161.

Gofman, J. W., F. Lindgren, H. Elliott, W. Mantz, J. Hewitt, B. Strisower, V. Herring, and T. P. Lyon. 1950. The Role of Lipids and Lipoproteins in Atherosclerosis. *Science* 111 (2877): 166–86. doi:10.1126/science.111.2877.166.

Hacking, Ian. 1983. *Representing and Intervening*. Cambridge University Press. doi:10.1017/cbo9780511814563.

Hampel, Harald, Katharina Bürger, Stefan J. Teipel, Arun L.W. Bokde, Henrik Zetterberg, and Kaj Blennow. 2008. Core Candidate Neurochemical and Imaging Biomarkers of Alzheimer's Disease. *Alzheimers & Dementia* 4 (1): 38–48. doi:10.1016/j.jalz.2007.08.006.

Hardy, John, and David Allsop. 1991. Amyloid Deposition as the Central Event in the Aetiology of Alzheimers Disease. *Trends in Pharmacological Sciences* 12 (January): 383–88. doi:10.1016/0165-6147(91)90609-v.

Horwich, Paul. 2011. *Probability and Evidence*. Cambridge University Press. doi:10.1017/cbo9781316494219.

Imbimbo, Bruno P., and Giuseppe A.M. Giardina. 2011. *γ*-Secretase Inhibitors and Modulators for the Treatment of Alzheimers Disease: Disappointments and Hopes. *Current Topics in Medicinal Chemistry* 11 (12): 1555–70. doi:10.2174/156802611795860942.

Institute of Medicine. 2010. Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease. Washington, DC: National Academies Press. doi:10.17226/12869.

Jack, Clifford R., and David M. Holtzman. 2013. Biomarker Modeling of Alzheimer's Disease. *Neuron* 80 (6): 1347–58. doi:10.1016/j.neuron.2013.12.003.

Justus, James. 2012. The Elusive Basis of Inferential Robustness. *Philosophy of Science* 79 (5): 795–807. doi:10.1086/667902.

Katz, Russell. 2004. Biomarkers and Surrogate Markers: An FDA Perspective. *NeuroRX* 1 (2): 189–95. doi:10.1602/neurorx.1.2.189.

Keeley, Brian L. 2002. Making Sense of the Senses. Edited by John Smylie. *Journal of Philosophy* 99 (1): 5–28. doi:10.5840/jphil20029915.

Keyser, Vadim. 2016. A New Theory of Robust Measurement. http://www.apaonline.org/members/group_content_view.asp?group=110424&id=476093.

Krut, Jan Jessen, Henrik Zetterberg, Kaj Blennow, Paola Cinque, Lars Hagberg, Richard W. Price, Marie Studahl, and Magnus Gisslén. 2012. Cerebrospinal Fluid Alzheimers Biomarker Profiles in CNS Infections. *Journal of Neurology* 260 (2): 620–26. doi:10.1007/s00415-012-6688-y.

Landau, Susan M., Ming Lu, Abhinay D. Joshi, Michael Pontecorvo, Mark A. Mintun, John Q. Trojanowski, Leslie M. Shaw, and William J. Jagust and. 2013. Comparing Positron Emission Tomography Imaging and Cerebrospinal Fluid Measurements of β-Amyloid. *Annals of Neurology* 74 (6): 826–36. doi:10.1002/ana.23908.

LaRosa, John C., Scott M. Grundy, David D. Waters, Charles Shear, Philip Barter, Jean-Charles Fruchart, Antonio M. Gotto, et al. 2005. Intensive Lipid Lowering with Atorvastatin in Patients with Stable Coronary Disease. *New England Journal of Medicine* 352 (14): 1425–35. doi:10.1056/nejmoa050461.

Lassere, Marissa N. 2007. The Biomarker-Surrogacy Evaluation Schema: A Review of the Biomarker-Surrogate Literature and a Proposal for a Criterion-Based, Quantitative, Multidimensional Hierarchical Levels of Evidence Schema for Evaluating the Status of Biomarkers as Surrogate Endpoints. *Statistical Methods in Medical Research* 17 (3): 303–40. doi:10.1177/0962280207082719.

Lassere, Marissa N, Kent R Johnson, Maarten Boers, Peter Tugwell, Peter Brooks, Lee Simon, Vibeke Strand, et al. 2007. Definitions and Validation Criteria for Biomarkers and Surrogate Endpoints: Development and Testing of a Quantitative Hierarchical Levels of Evidence Schema. *The Journal of Rheumatology* 34 (3): 607–15.

Lee, Hyoung-gon, Xiongwei Zhu, Akihiko Nunomura, George Perry, and Mark A. Smith. 2006. Amyloid-β Vaccination: Testing the Amyloid Hypothesis? *The American Journal of Pathology* 169 (3): 738–39. doi:10.2353/ajpath.2006.060633.

Lehmann, Sylvain, Julien Dumurgier, Susanna Schraen, David Wallon, Frédéric Blanc, Eloi Magnin, Stéphanie Bombois, et al. 2014. A Diagnostic Scale for Alzheimer's Disease Based on Cerebrospinal Fluid Biomarker Profiles. *Alzheimers Research & Therapy* 6 (3): 38. doi:10.1186/alzrt267.

Lesne, S. 2014. Toxic Oligomer Species of Amyloid-β in Alzheimers Disease, a Timing Issue. *Swiss Medical Weekly*, November. doi:10.4414/smw.2014.14021.

Levins, Richard. 1966. The Strategy of Model Building in Population Biology. *American Scientist* 54 (4): 421–31.

Lloyd, Elisabeth A. 2010. Confirmation and Robustness of Climate Models. *Philosophy of Science* 77 (5): 971–84. doi:10.1086/657427.

Marnell, Lorraine, Carolyn Mold, and Terry W. Du Clos. 2005. C-Reactive Protein: Ligands, Receptors and Role in Inflammation. *Clinical Immunology* 117 (2): 104–11. doi:10.1016/j.clim.2005.08.004.

Mattsson, Niklas, Philip S. Insel, Michael Donohue, Susan Landau, William J. Jagust, Leslie M. Shaw, John Q. Trojanowski, Henrik Zetterberg, Kaj Blennow, and Michael W. Weiner. 2014. Independent Information from Cerebrospinal Fluid Amyloid-β and Florbetapir Imaging in Alzheimers Disease. *Brain* 138 (3): 772–83. doi:10.1093/brain/awu367.

Mayeux, Richard. 2004. Biomarkers: Potential Uses and Limitations. *NeuroRX* 1 (2): 182–88. doi:10.1602/neurorx.1.2.182.

McConkey, B, P Davies, R A Crockson, A P Crockson, M Butler, T J Constable, and R S Amos. 1979. Effects of Gold, Dapsone, and Prednisone on Serum c-Reactive Protein and Haptoglobin and the Erythrocyte Sedimentation Rate in Rheumatoid Arthritis. *Annals of the Rheumatic Diseases* 38 (2): 141–44. doi:10.1136/ard.38.2.141.

Mo, Jin-A, Ju-Hee Lim, Ah-Ram Sul, Min Lee, Young Chul Youn, and Hee-Jin Kim. 2015. Cerebrospinal Fluid β-Amyloid142 Levels in the Differential Diagnosis of Alzheimer's DiseaseSystematic Review and Meta-Analysis. Edited by Rosanna Squitti. *PLOS ONE* 10 (2): e0116802. doi:10.1371/journal.pone.0116802.

Musiek, Erik S., and David M. Holtzman. 2012. Origins of Alzheimer's Disease. *Current Opinion in Neurology* 25 (6): 715–20. doi:10.1097/wco.0b013e32835a30f4.

Orzack, Steven Hecht, and Elliott Sober. 1993. A Critical Assessment of Levinss the Strategy of Model Building in Population Biology (1966). *The Quarterly Review of Biology* 68 (4): 533–46. doi:10.1086/418301.

Otvos, James D., Samia Mora, Irina Shalaurova, Philip Greenland, Rachel H. Mackey, and David C. Goff. 2011. Clinical Implications of Discordance Between Low-Density Lipoprotein Cholesterol and Particle Number. *Journal of Clinical Lipidology* 5 (2): 105–13. doi:10.1016/j.jacl.2011.02.001.

Palmqvist, Sebastian, Henrik Zetterberg, Niklas Mattsson, Per Johansson, Lennart Minthon, Kaj Blennow, Mattias Olsson, Oskar Hansson, and and. 2015. Detailed Comparison of Amyloid PET and CSF Biomarkers for Identifying Early Alzheimer Disease. *Neurology* 85 (14): 1240–9. doi:10.1212/wnl.0000000000001991.

Pepe, M. S., H. Janes, C. I. Li, P. M. Bossuyt, Z. Feng, and J. Hilden. 2016. Early-Phase Studies of Biomarkers: What Target Sensitivity and Specificity Values Might Confer Clinical Utility? *Clinical Chemistry* 62 (5): 737–42. doi:10.1373/clinchem.2015.252163.

Prentice, Ross L. 1989. Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria. *Statistics in Medicine* 8 (4): 431–40. doi:10.1002/sim.4780080407.

Qu, Yongming. 2013. Statistical Evaluation of Surrogate Markers: Validity, Efficiency, and Sensitivity. *Clinical Trials: Journal of the Society for Clinical Trials* 10 (5): 693–95. doi:10.1177/1740774513499652.

Ratner, Mark. 2015. Biogens Early Alzheimers Data Raise Hopes, Some Eyebrows. *Nature Biotechnology* 33 (5): 438–38. doi:10.1038/nbt0515-438.

Reiman, Eric M. 2016. Attack on Amyloid-β Protein. *Nature* 537 (7618): 36–37. doi:10.1038/537036a.

Ridker, Paul M, Eleanor Danielson, Francisco A.H. Fonseca, Jacques Genest, Antonio M. Gotto, John J.P. Kastelein, Wolfgang Koenig, et al. 2008. Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated c-Reactive Protein. *New England Journal of Medicine* 359 (21): 2195–2207. doi:10.1056/nejmoa0807646.

Ritchie, Karen, Isabelle Carrière, Claudine Berr, Hélène Amieva, Jean-François Dartigues, Marie-Laure Ancelin, and Craig W. Ritchie. 2016. The Clinical Picture of Alzheimer's Disease in the Decade Before Diagnosis. *The Journal of Clinical Psychiatry*, February, e305–e311. doi:10.4088/jcp.15m09989.

Sachdeva, Amit, Christopher P. Cannon, Prakash C. Deedwania, Kenneth A. LaBresh, Sidney C. Smith, David Dai, Adrian Hernandez, and Gregg C. Fonarow. 2009. Lipid Levels in Patients Hospitalized with Coronary Artery Disease: An Analysis of 136,905 Hospitalizations in Get with the Guidelines. *American Heart Journal* 157 (1): 111–117.e2. doi:10.1016/j.ahj.2008.08.010.

Savva, George M., Stephen B. Wharton, Paul G. Ince, Gillian Forster, Fiona E. Matthews, and Carol Brayne. 2009. Age, Neuropathology, and Dementia. *New England Journal of Medicine* 360 (22): 2302–9. doi:10.1056/nejmoa0806142.

Sbong, Stephanie, and Mark Feldman. 2014. Frequency and Causes of c-Reactive Protein and Erythrocyte Sedimentation Rate Disagreements in Adults. *International Journal of Rheumatic Diseases* 18 (1): 29–32. doi:10.1111/1756-185x.12537.

Schneider, Lon S., Richard E. Kennedy, and Gary R. Cutter. 2010. Requiring an Amyloid-β1-42 Biomarker for Prodromal Alzheimers Disease or Mild Cognitive Impairment Does Not Lead to More Efficient Clinical Trials. *Alzheimers & Dementia* 6 (5): 367–77. doi:10.1016/j.jalz.2010.07.004.

Schupbach, Jonah N. 2016. Robustness Analysis as Explanatory Reasoning. *The British Journal for the Philosophy of Science*, May, axw008. doi:10.1093/bjps/axw008.

Selkoe, Dennis J, and John Hardy. 2016. The Amyloid Hypothesis of Alzheimers Disease at 25 Years. *EMBO Molecular Medicine* 8 (6): 595–608. doi:10.15252/emmm.201606210.

Sevigny, Jeff, Ping Chiao, Thierry Bussière, Paul H. Weinreb, Leslie Williams, Marcel Maier, Robert Dunstan, et al. 2016. The Antibody Aducanumab Reduces aβ Plaques in Alzheimer's Disease. *Nature* 537 (7618): 50–56. doi:10.1038/nature19323.

Sober, Elliott. 1989. Independent Evidence About a Common Cause. *Philosophy of Science* 56 (2): 275–87. doi:10.1086/289487.

Sperling, Reisa A., Clifford R. Jack, Sandra E. Black, Matthew P. Frosch, Steven M. Greenberg, Bradley T. Hyman, Philip Scheltens, et al. 2011. Amyloid-Related Imaging Abnormalities in Amyloid-Modifying Therapeutic Trials: Recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimers & Dementia* 7 (4): 367–85. doi:10.1016/j.jalz.2011.05.2351.

Staley, Kent W. 2004. Robust Evidence and Secure Evidence Claims. *Philosophy of Science* 71 (4): 467–88. doi:10.1086/423748.

Stegenga, Jacob. 2009. Robustness, Discordance, and Relevance. *Philosophy of Science* 76 (5): 650–61. doi:10.1086/605819.

———. 2012. Rerum Concordia Discors: Robustness and Discordant Multimodal Evidence. In *Characterizing the Robustness of Science*, 207–26. Springer Netherlands. doi:10.1007/978-94-007-2759-5_9.

Terry, Robert D., Eliezer Masliah, David P. Salmon, Nelson Butters, Richard DeTeresa, Robert Hill, Lawrence A. Hansen, and Robert Katzman. 1991. Physical Basis of Cognitive Alterations in Alzheimers Disease: Synapse Loss Is the Major Correlate of Cognitive Impairment. *Annals of Neurology* 30 (4): 572–80. doi:10.1002/ana.410300410.

Toledo, Jon B, Michael W Weiner, David A Wolk, Xiao Da, Kewei Chen, Steven E Arnold, William Jagust, et al. 2014. Neuronal Injury Biomarkers and Prognosis in ADNI Subjects with Normal Cognition. *Acta Neuropathologica Communications* 2 (1). doi:10.1186/2051-5960-2-26.

Toyn, Jeremy. 2015. What Lessons Can Be Learned from Failed Alzheimer's Disease Trials? *Expert Review of Clinical Pharmacology* 8 (3): 267–69. doi:10.1586/17512433.2015.1034690.

Trout, J. D. 1998. *Measuring the Intentional World*. doi:10.1093/0195107667.001.0001.

Van Fraassen, Bas C. 2010. *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.

Vos, Stephanie J.B., Brian A. Gordon, Yi Su, Pieter Jelle Visser, David M. Holtzman, John C. Morris, Anne M. Fagan, and Tammie L.S. Benzinger. 2016. NIA-AA Staging of Preclinical Alzheimer Disease: Discordance and Concordance of CSF and Imaging Biomarkers. *Neurobiology of Aging* 44 (August): 1–8. doi:10.1016/j.neurobiolaging.2016.03.025.

Walsh, Dominic M., and Dennis J. Selkoe. 2007. Aβ Oligomers – a Decade of Discovery. *Journal of Neurochemistry* 101 (5): 1172–84. doi:10.1111/j.1471-4159.2006.04426.x.

Walsh, Dominic M., Bertrand P. Tseng, Russell E. Rydel, Marcia B. Podlisny, and Dennis J. Selkoe. 2000. The Oligomerization of Amyloid β-Protein Begins Intracellularly in Cells Derived from Human Brain. *Biochemistry* 39 (35): 10831–9. doi:10.1021/bi001048s.

Weisberg, Michael. 2006. Robustness Analysis. *Philosophy of Science* 73 (5): 730–42. doi:10.1086/518628.

Willem, M., A. N. Garratt, B. Novak, M. Citron, S. Kaufmann, A. Rittger, B. DeStrooper, P. Saftig, C. Birchmeier, and C. Haass. 2006. Control of Peripheral Nerve Myelination by the -Secretase BACE1. *Science* 314 (5799): 664–66. doi:10.1126/science.1132341.

Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.

Wolz, Robin, Adam J. Schwarz, Katherine R. Gray, Peng Yu, and Derek L.G. Hill and. 2016. Enrichment of Clinical Trials in MCI Due to AD Using Markers of Amyloid and Neurodegeneration. *Neurology* 87 (12): 1235–41. doi:10.1212/wnl.0000000000003126.

Woo, Ha-Na, Jong-Sung Park, A-Ryeong Gwon, Thiruma V. Arumugam, and Dong-Gyu Jo. 2009. Alzheimer's Disease and Notch Signaling. *Biochemical and Biophysical Research Communications* 390 (4): 1093–7. doi:10.1016/j.bbrc.2009.10.093.

Woodward, James. 2004. *Making Things Happen: A Counterfactual Theory of Causal Explanation*. Oxford: Oxford University Press. doi:10.1093/0195155270.003.0005.

Woodward, Jim. 2006. Some Varieties of Robustness. *Journal of Economic Methodology* 13 (2): 219–40. doi:10.1080/13501780600733376.

Young, Alexandra L., Neil P. Oxtoby, Pankaj Daga, David M. Cash, Nick C. Fox, Sebastien Ourselin, Jonathan M. Schott, and Daniel C. Alexander. 2014. A Data-Driven Model of Biomarker Changes in Sporadic Alzheimers Disease. *Brain* 137 (9): 2564–77. doi:10.1093/brain/awu176.

Younkin, Steven G. 1995. Evidence That aβ42 Is the Real Culprit in Alzheimers Disease. *Annals of Neurology* 37 (3): 287–88. doi:10.1002/ana.410370303.

Zwan, Marissa, Argonde van Harten, Rik Ossenkoppele, Femke Bouwman, Charlotte Teunissen, Sofie Adriaanse, Adriaan Lammertsma, Philip Scheltens, Bart N.M. van Berckel, and Wiesje Van der Flier. 2013. Concordance Between CSF Biomarkers and [11c]PIB PET in a Memory Clinic Population. *Alzheimers & Dementia* 9 (4): P830. doi:10.1016/j.jalz.2013.04.476.